Published in final edited form as:

Biometrics. 2013 September ; 69(3): 741–747. doi:10.1111/biom.12066.

Mechanistic Analysis of Challenge-Response Experiments

M. S. Shotwell*, K. J. Drake, V. Y. Sidorov, and J. P. Wikswa

Vanderbilt University, Nashville, TN 37232, USA

Summary

We present an application of mechanistic modeling and nonlinear longitudinal regression in the context of biomedical response-to-challenge experiments, a field where these methods are underutilized. In this type of experiment, a system is studied by imposing an experimental challenge, and then observing its response. The combination of mechanistic modeling and nonlinear longitudinal regression has brought new insight, and revealed an unexpected opportunity for optimal design. Specifically, the mechanistic aspect of our approach enables the optimal design of experimental challenge characteristics (e.g., intensity, duration). This article lays some groundwork for this approach. We consider a series of experiments wherein an isolated rabbit heart is challenged with intermittent anoxia. The heart responds to the challenge onset, and recovers when the challenge ends. The mean response is modeled by a system of differential equations that describe a candidate mechanism for cardiac response to anoxia challenge. The cardiac system behaves more variably when challenged than when at rest. Hence, observations arising from this experiment exhibit complex heteroscedasticity and sharp changes in central tendency. We present evidence that an asymptotic statistical inference strategy may fail to adequately account for statistical uncertainty. Two alternative methods are critiqued qualitatively (i.e., for utility in the current context), and quantitatively using an innovative Monte Carlo method. We conclude with a discussion of the exciting opportunities in optimal design of response-to-challenge experiments.

Keywords

Bootstrap; Differential equation; Longitudinal data; Mechanistic model; Optimal experimental design

1. Introduction

Observing the time series of response to designed physiological challenges is a powerful tool for biomedical research and diagnostics. The cardiac stress test is a well-known example. There is an emerging effort to utilize this methodology in studies of isolated tissues, and even single cells (Eklund et al., 2009; Enders et al., 2010; Faley et al., 2008; Snider et al., 2010; Wikswa et al., 2006; LeDuc et al., 2011; Wikswa et al., 1995). This *response-to-challenge* methodology is part of a system-level approach (opposite a

*matt.shotwell@vanderbilt.edu.

Supplementary Material: The Web Appendix, Figures, and Table referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library.

reductionist approach), and has become more prominent in the study of complex metabolic, signaling, and gene regulatory pathways (Garcia et al., 2011). Mechanistic mathematical models (e.g., systems of differential equations) are often used to characterize the mechanisms that underlie a biological system. The system is studied by making empirical observations on some of its parts, and then drawing inferences about the unobserved parts. This type of modeling and inference is common in epidemiology and pharmacology, for example (see Edelstein-Keshet (2005) for a comprehensive introduction), but is largely neglected in other basic sciences.

While the experimental response-to-challenge methodology has advanced, tailored statistical methods have lagged. As a result, the statistical methods employed often make unrealistic assumptions, fail to account for complexities in the data, or fail to adequately characterize (or worse, misleadingly characterize) the evidence about important scientific questions. Indeed, May (2004) highlights the use of conventional statistical methods and assumptions in highly nonlinear dynamical processes as an *abuse of mathematics in biology*. The impetus for the current work was a concern that important biological insights might be missed at the time of statistical analysis; that without sufficient care, statistical analysis might hinder rather than promote scientific discovery.

The response-to-challenge approach was implemented to study the connection between cardiac electrophysiology and metabolism using a series of measurements on isolated rabbit hearts. The cardiac tissue was perfused with either normal media or amino acid enriched media. Cardiac action potential duration (APD) was measured in a time course of intermittent metabolic challenges. In this experiment, two intermittent challenges were imposed by withholding oxygen from perfusion media (anoxia), thereby simulating the effect of acute lung injury, pulmonary embolism, asphyxia, and other respiratory conditions. The primary biological endpoint of the experiment was to draw inferences about the cardioprotective effect of amino acid enrichment under intermittent anoxia challenges.

In separate experiments (data not shown), no cardioprotective effect was evident in the absence of anoxia challenge, or when only one anoxia challenge was applied. We hypothesized that repeated anoxia would further strain the cardiac tissue, and be more likely to reveal a cardioprotective effect for amino acid enrichment.

We employed a mechanistic differential equation model for the change in cardiac APD over time, accounting for initiation and termination of the metabolic challenge events, and to parameterize the effect of amino acid enrichment. This type of mechanistic model serves the immediate goal of drawing statistical inferences about the effect of amino acid enrichment, but is also useful in making predictions about how cardiac tissue behaves under anoxic stress. In particular, we are interested in predicting the consequence of modifying the challenge characteristics (e.g., extending the duration of anoxia) on the mean APD response. This type of prediction is useful, for example, in the design of new response-to-challenge experiments.

Complex variability, including residual autocorrelation and heteroscedasticity, is a prominent feature of the cardiac APD data. As the evidence presented below suggests, the

anoxia challenge affects both mean *and* variance of the response. Because of this complexity and the small-sample nature of these data, there was concern about the use of asymptotic results (under strong parametric assumptions) for statistical inference. Indeed, Sohn and Menke (2002) reported that inference based on asymptotic results had underestimated uncertainty in a related nonlinear regression application, and recommend a bootstrap alternative.

Three distinct estimation and inference strategies were evaluated in the current context: ordinary least squares (OLS) estimation with nonparametric bootstrap inference, maximum likelihood estimation (MLE) with nonparametric bootstrap inference, and maximum likelihood estimation with asymptotic inference. The frequentist coverage of pointwise (i.e., points in time) confidence bands, and the power associated with pointwise hypothesis tests were examined using a novel Monte Carlo method. Surprisingly, coverage was similar for the bootstrap OLS and bootstrap MLE methods. The asymptotic MLE method had lesser coverage in general, and especially less for time points within the anoxia challenge periods.

Our objective in studying alternative estimation and inference strategies was to better inform our analysis of the cardiac APD data. We make no generalized claims about the relative merits of these methods. However, we anticipate that these findings will inform the analysis of future response-to-challenge experiments. In addition, we have not attempted to demonstrate the asymptotic validity of the bootstrap procedures considered herein. While others continue to establish the validity of various bootstrap estimators in nonlinear regression (Hu and Kalbfleisch, 2000; Gonçalves and White, 2004, 2005; Chatterjee and Bose, 2005), we show that a numerical strategy can be successful in evaluating a bootstrap procedure in a narrow context.

Design, implementation, and analysis of response-to-challenge experiments is a developing, multidisciplinary endeavor to study complex biology. This manuscript considers several key statistical issues that are likely to emerge in the analysis of data resulting from challenge-response experiments. We conclude with a discussion of the outstanding statistical problems, including the promising opportunities associated with optimal experimental design.

2. Experimental Data

The experiments we consider were conducted on 14 isolated live rabbit hearts. Seven hearts were perfused with normal media and seven with amino acid enriched media (i.e., with the amino acids glutamine and glutamate). The experiment was designed to assess the cardioprotective effect of amino acid enrichment under anoxic challenge. The durations of cardiac action potentials (APD) were observed at one or two minute intervals for up to 72 minutes. All hearts were observed at baseline for up to twenty minutes. In two subsequent time periods, hearts were challenged with anoxia by withholding oxygen from the perfusion media. The two challenges lasted for six minutes each, and were separated by 20 minutes of recovery. Hence, the time series is characterized by a sequence of five adjacent time periods: baseline, first anoxia, first recovery, second anoxia, and second recovery. Figure 1 illustrates the time course of raw APD measurements on each heart.

3. Statistical Modeling Approach

We considered a nonlinear regression model of the form

$$y_{ijk} = \alpha_j + \eta(t_k, \theta_i) + \varepsilon_{ijk} \quad (1)$$

where y_{ijk} is the measured APD on heart $j \in \{1, 2, \dots, 14\}$ at time $t_k \in \{-20, -19, \dots, 50\}$, under perfusion with media type $i \in \{E, N\}$. The intercept α_j represents the baseline APD for heart j . This varies among hearts due to innate differences and instrument recalibration. The function $\eta(t_k, \theta_i)$ gives the average change in APD from baseline until time t_k , given a fixed parameter θ_i . Finally, ε_{ijk} is a random error with mean zero. The intercepts were considered fixed rather than random to avoid the multiplicity of definitions for “residuals” in models that have random effects (Nobre and Singer, 2007), and more importantly, because the correlations among errors were modeled directly. The disadvantages of this approach are that 1) each α_j must be estimated and 2) no model is available for the purpose of simulating new α values. The former is not serious because there were many measurements on each heart. In the simulations described below, new α values were drawn from the empirical distribution defined by the estimated values $\{\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{14}\}$.

Controlled anoxia challenges demonstrated a strong effect on cardiac APD. Within each anoxia period, the time course of APD was characterized by a small initial upswing, followed by a sharp decline. The recovery periods were characterized by a sharp initial upswing, followed by a period of stability. APD also exhibited a slow decay in time starting at the first exposure to anoxia and persisting throughout the experiment. This sharp, irregular response led us to develop a piecewise differential equation model for the change in mean APD over time. We had briefly considered a restricted cubic splines (Harrell, 2006) model, but found that the cubic splines enforced a high degree of smoothness (Wegman and Wright, 1983) that failed to capture the sharp transitions between the baseline, anoxia, and recovery periods. More importantly, the splines model is not mechanistic, in the sense that predictions about APD under alternative experimental designs (i.e., different types of challenge) are not possible.

We rationalized that the heart would initially compensate for mildly anoxic conditions by contracting longer and with more force, resulting in lengthened cardiac action potentials. However, prolonged anoxia restricts oxidative metabolism and deprives cardiac tissue of the energy needed for contraction, resulting in shortened action potentials. On reintroduction of oxygenated media, oxidative metabolism is restored and APD quickly recovers to near pre-challenge levels. The heart is weakened as a result of anoxia, causing a slow decay in APD over time. We hypothesized that perfusion with amino acid enriched media would decelerate this anoxia-initiated decay in APD. These mechanisms were modeled using a system of piecewise ordinary differential equations that relate action potential duration (D), the proportion of oxygen available for oxidative metabolism (O), and a third proportion (E) that represents the unspecific cellular and metabolic resources that are necessary to sustain a cardiac action potential.

Using $X' = \frac{dX}{dt}$ to denote the derivative of X with respect to the time variable t , consider the following system of differential equations:

$$D' = \begin{cases} 0 & t \leq t_{a0} \\ k_0(E' - O') - k_1 t & t_{a0} < t \end{cases}$$

$$O' = \begin{cases} 0 & t \leq t_{a0} \\ -l_0 O & t_{a0} < t \leq t_{r0} \\ l_1(1 - O) & t_{r0} < t \leq t_{a1} \\ -l_0 O & t_{a1} < t \leq t_{r1} \\ l_1(1 - O) & t_{r1} < t \end{cases} \quad E' = \begin{cases} 0 & t \leq t_{a0} + t_{\text{off}0} \\ -m_0 O & t_{a0} + t_{\text{off}0} < t \leq t_{r0} \\ m_1(1 - O) & t_{r0} < t \leq t_{a1} + t_{\text{off}1} \\ -m_0 O & t_{a1} + t_{\text{off}1} < t \leq t_{r1} \\ m_1(1 - O) & t_{r1} < t \end{cases}, \quad (2)$$

where t_{a0} and t_{a1} are the first and second anoxia start times, respectively, and t_{r0} and t_{r1} are the corresponding anoxia stop times. The expressions for O' and E' are each given in five pieces, corresponding to the five distinct experimental phases. Each anoxia is associated with an “offset”; a the period of time between the start of anoxia, and the point where metabolic resources (i.e., the quantity modeled by E) begin to deplete. The lengths of the “offset” periods are denoted $t_{\text{off}0}$ and $t_{\text{off}1}$ for the transitions from baseline to the first anoxia, and from the first recovery to the second anoxia, respectively. The times t_{a0} , t_{a1} , t_{r0} , and t_{r1} are fixed aspects of the experimental design. The unknown times $t_{\text{off}0}$ and $t_{\text{off}1}$ are treated as model parameters.

The parameter k_1 represents the rate of slow decay in APD. All remaining parameters affect the behavior of APD in response to anoxia. The values of l_0 and m_0 represent the rates of oxygen and bioenergetic resource depletion during the anoxia periods, respectively, and l_1 and m_1 are the corresponding rates of replenishment during the recovery periods. All unknown model parameters were estimated simultaneously using one of the methods described below, under “Estimation and Inference Strategies”.

While D measures APD in milliseconds, O and E are unitless proportions, where 1 represents full availability and 0 represents complete exhaustion. The initial (baseline) values for D , O , and E were fixed to 0, 1, and 1 respectively. This system of equations (2) yields analytical solutions for $D(t)$, $O(t)$, and $E(t)$ (see Web Appendix). For the analysis at hand, only the solution for $D(t)$ is used.

The solution for $D(t)$ given $\theta_i = \{k_0, k_{1i}, l_0, l_1, m_{0i}, m_{1i}, t_{\text{off}0}, t_{\text{off}1}\}$ is denoted $D_i(t_k)$, and substituted for $\eta(t_k, \theta_i)$ in equation 1. Because the degree of anoxia was controlled, and since amino acid enrichment was thought to affect the metabolic process, only the parameters k_{1i} , m_{0i} , and m_{1i} were allowed to vary by perfusion media type (i.e., by index i). Hence, this mechanistic model permits comparisons of normal and amino acid enriched media treatments in terms of their effect on (1) the rates of depletion and replenishment of bioenergetic resources (m_{0i} , and m_{1i}) during and after an anoxic challenge, and (2) the severity of cardiac tissue weakening caused by the initial anoxic stress (k_{1i}).

4. Estimation and Inference Strategies

Accurate assessment of variability is important for inferential purposes, but also for our secondary objective: to predict the variability in cardiac responses to alternative experimental challenges (e.g., more frequent challenges, or for longer durations). The covariance structure in the cardiac APD data is complex. In order to characterize the variability in these data, three different methods for estimation and inference were considered.

In the following discussion, the terms “parametric” and “nonparametric” refer to the methods used to evaluate the sampling distribution of estimators. A method was considered parametric if the errors (i.e., ε_{ijk} , for all i, j , and k) were modeled as having arisen from a parametric probability distribution, *and* variability about sample estimators was evaluated under this assumption. Methods that fail to meet this criterion were considered nonparametric. The terms “parametric” and “nonparametric” are not used in regard to the mean model ($E[y_{ijk}] = \alpha_j + \eta(t_k, \theta_i)$), which is clearly parameterized.

The first method is nonparametric. The model parameters θ_i were estimated using the method of ordinary least squares (OLS). This method was selected initially to help characterize the structure of residual covariability in these data without specifying a parametric error distribution. The adjusted (BCa) percentile method (Efron, 1987; Davison and Hinkley, 1997) was used to construct 95% confidence intervals for each model parameter, and a pointwise 95% confidence band for the difference in mean APD ($D_E(t) - D_N(t)$) associated with amino acid enriched media versus normal media. A pointwise (versus simultaneous) confidence band was used in order to assess the difference in mean APD at the final measurement (52 minutes). Hence, the reader should be cautioned against drawing simultaneous inferences about the difference in mean APD.

In order to preserve the complex covariance structure among resamples, the time series of measurements on individual hearts were considered clustered. Hence, whole clusters were resampled with replacement. This method corresponds to the *cluster bootstrap* of Field and Welsh (2007), a simplification of the *multi-stage bootstrap* for hierarchically structured data (Davison and Hinkley, 1997). Seven clusters were resampled from each of the normal and amino acid enriched media groups. Recently, Crainiceanu et al. (2012) advocate for an extension of this approach in a related application.

Ten thousand resamples were used for all bootstrap procedures. The bootstrap size was selected to ensure that the 2.5% and 97.5% bootstrap percentiles had coefficient of variation (CV) less than 0.01 for each estimator. The method of Maritz and Jarrett (1978) and Efron (1979) was used to estimate the variance of bootstrap percentiles.

It is well known that the bootstrap distributions for certain types of estimators do not have the expected limiting behavior (e.g., see Example 2.15, Davison and Hinkley, 1997). We have not examined the asymptotic properties of the bootstrap methods used herein. However, empirical studies have shown the bootstrap to be reasonable for similar types of inferences (Hinkley and Schechtman, 1987; Duggins, 2010).

Figure 2 illustrates the OLS-fitted data, the estimated difference in mean APD, and the bootstrap-constructed 95% confidence band. Web Figure 5 illustrates the raw APD means plotted against OLS-fitted means, with good agreement. Web Table 1 gives estimates and 95% confidence intervals for each model parameter. In particular, the rate of slow APD decay (k_{1N} vs. k_{1E}) for the amino acid enriched media group was improved by 1.2_{E-2} ms·s⁻² (95% CI: 2.6_{E-3} , 2.2_{E-2}), relative to the normal media group. The rates of depletion and replenishment of bioenergetic resources during anoxia and recovery periods (m_{0N} , m_{1N} , m_{0E} , m_{1E}) were not significantly different for the enriched versus normal media groups. However, bioenergetic resources began to deplete 55 s ($t_{\text{off}0} - t_{\text{off}1}$; 95% CI: 39, 72) earlier following the second onset of anoxia ($t_{\text{off}1}$; 48 s, 95% CI: 31, 84) than at the first ($t_{\text{off}0}$; 102 s, 95% CI: 90, 138). Although parameter estimates were slightly different under the two subsequent statistical strategies, the findings were qualitatively consistent.

Analysis of residuals from the OLS-fitted model revealed significant residual autocorrelation in APD measurements, and increasing residual variance within the anoxia challenge periods (heteroscedasticity). Figure 3 illustrates these findings.

The second estimation and inference strategy is fully parametric. Using the evidence illustrated by Figure 3, ε_{ijk} was modeled as having arisen from the normal distribution with mean zero and covariance such that

$$\begin{aligned} \text{var}(\varepsilon_{ijk}) &= \begin{cases} \sigma^2 & \text{baseline, recovery} \\ \sigma^2 e^{\psi(t_k - t_a)} & \text{anoxia} \end{cases} \\ \text{cor}(\varepsilon_{ijk}, \varepsilon_{ijk'}) &= \rho^{|t_k - t_{k'}|}, \end{aligned} \quad (3)$$

where all other correlations are zero, and t_a is the time anoxia was initiated (i.e., $(t_k - t_a)$ measures the duration of anoxia). This *autoregressive* construction approximates the types of heteroscedasticity and autocorrelation observed in the cardiac APD data. Specifically, the variance in APD measurement is constant for baseline and recovery, but grows exponentially within the anoxia periods. Hence, this parameterizes the notion that perturbed biological systems behave more variably (or measured with less precision) than systems at rest.

Given this distributional information, a maximum likelihood estimate (MLE) was constructed for θ_i , σ^2 , ψ , and ρ . Variability about the MLE was assessed using its asymptotic (i.e., normal) distribution, where the covariance was estimated by the inverse observed Fisher information, evaluated at the MLE. Monte Carlo integration was used to construct a pointwise 95% confidence band for the difference in mean APD.

He and Severini (2010) demonstrated that the maximum likelihood estimator for parametric models with multiple unknown change-points is consistent, and has a multivariate normal limiting distribution with covariance matrix identical to the inverse Fisher information. Their result is valid under mild regularity conditions.

Web Figure 1 illustrates the MLE-fitted data, the estimated difference in mean APD, and the pointwise asymptotic 95% confidence band. Web Figure 2 shows the estimated

autocorrelation and variance for the MLE standardized residuals. Because no significant autocorrelation was found among the standardized residuals, the parametric correlation structure given by expression (3) was considered sufficient. Web Table 1 gives estimates and 95% confidence intervals for each parameter in θ_i , σ^2 , ψ , and ρ .

The third and final approach utilizes a likelihood-based estimator, but characterizes its variability nonparametrically: the MLE was used, but variability about the MLE was assessed using the nonparametric bootstrap, as before. The corresponding estimates and 95% confidence intervals are given in Web Table 1. Web Figure 3 illustrates the MLE-fitted data and estimated difference in mean APD as before, and the pointwise *bootstrap* 95% confidence band.

5. Simulation and Results

Because the processes that generate these experimental data are unknown, it is impossible to assess the utility of competing statistical methods under the “true” data-generating mechanism. However, if a *parametric* process is assumed (as is often done, for analytical and practical reasons), then our assessment may artificially favor a parametric strategy, and *vice versa* when empirical (nonparametric) or semiparametric processes are assumed. We refer to this as “simulation bias”. Indeed, nonparametric methods are often less efficient than parametric methods when the underlying data arise from the corresponding parametric process (for a specific example, see Hollander and Wolfe (1999), pages 104–105).

To counter this potential bias, we constructed a data generating mechanism from an equal mixture (i.e., with mixing probability 0.5) of parametric and nonparametric processes. The former was taken as the MLE-fitted model in the section above. The nonparametric process was identical to the empirical distribution over clustered APD measurements, adjusted so that the mean APD at each time point matched that of the parametric process. Five hundred new data sets were simulated by this mechanism. The size of treatment groups, and frequency and spacing of measurements in time were simulated to match the original cardiac APD experiment.

For each of the estimation and inference alternatives described in the preceding section, the pointwise coverage and power associated with pointwise null hypotheses were evaluated. For each time point, the pointwise coverage was computed as the frequency of 95% confidence intervals that covered the “true”, simulated difference in mean APD. The pointwise null hypothesis $H_0(t) : D_E(t) - D_N(t) = 0$ was rejected when the corresponding 95% confidence interval failed to include zero. Note that this testing strategy does not ensure a fixed type I error rate. Statistical power was approximated by the rejection frequency in simulated experiments.

Figure 4 illustrates the pointwise coverage and power for 1500 simulated experiments. In particular, the asymptotic MLE method is anticonservative about the mean APD estimate, especially within the anoxia periods. As a consequence, the coverage was lesser in these regions, and generally worse than that of the bootstrap methods. However, all methods exhibited less coverage than the nominal level.

Statistical power was generally greater for the asymptotic MLE method versus the bootstrap MLE method. In the most powerful regions (i.e., late in the recovery periods), the difference in estimated power was near 0.10.

Because the simulated difference in mean APD was nonzero at every time point beyond the baseline period, the pointwise type I error rate was not estimated. However, the simulated difference was *near zero* shortly after initiating the first anoxia challenge. Hence, the estimated power is a rough gauge of the type I error rate in this region. Figure 4 indicates that the approximate type I error rate for the asymptotic MLE method was near 0.20. Rates for the bootstrap methods were near 0.10.

It is of particular interest that the bootstrap OLS method did not have significantly better or worse estimated coverage than the bootstrap MLE method, even though the MLE explicitly accounts for the complex covariation observed in these data, and is asymptotically more efficient (i.e., under certain regularity conditions that were not verified in the present context; Casella and Berger (2001), page 472).

6. Discussion

There are significant costs and benefits associated with parametric methods that cannot be quantified in simple statistical measures (e.g., coverage and the rates of type I and II errors). For example, parametric error strategies often require greater human effort and other resources (e.g., model building, diagnostics, additional computer optimization, *etc.*). When the error structure has little inherent scientific value, the additional effort may not be worthwhile. Indeed, our analysis of the cardiac APD data revealed little benefit, in terms of coverage and power, associated with the bootstrap MLE versus bootstrap OLS methods.

In the current context, there is scientific value in characterizing the changes in variability of APD as a consequence of anoxia challenge. In particular, this is important in the design of future response-to-challenge experiments, where the number, duration, and spacing of challenges are design parameters.

To illustrate, consider a hypothetical experiment that aims to evaluate whether average APD continues to decline under anoxic conditions, or stabilizes at some time after the onset of anoxia. Let the design be identical to the cardiac APD experiment considered above, except that each anoxia challenge is prolonged by some amount t_{pro} . An alternative mechanistic model that accommodates the stabilization behavior is then proposed in order to evaluate the competing hypotheses. Both models may be fitted to preliminary data, and used to predict the behavior of APD under the hypothetical experiment.

The value of t_{pro} may then be selected to render the largest discrepancy in predictions between the two competing models (e.g., using the KL-optimality criterion of López-Fidalgo et al. (2007)). Because the variance in APD increases exponentially within the anoxia periods, a model and design criterion that account for this are necessary to fully evaluate the hypothetical experiment. In other words, prolonged anoxia is expected to result in continued APD decline, but the accompanying increase in variability may impose a constraint on the length of anoxia challenges. Indeed, sufficiently long exposure to anoxia

results in fibrillation (and ultimately cardiac arrest), a physiological state where APD measurements are variable in the extreme.

Other types of hypotheses about the cardiac APD mechanism may benefit from the application of more conventional optimal design methods (i.e., those that aim to maximize a function of the Fisher information about one or more model parameters; see Fedorov (2010) for a recent review). The combination of mechanistic mean modeling and parametric error modeling makes this possible. Because the challenge characteristics are the crux of a response-to-challenge experiment, strategies for optimizing its design promise significant and powerful new insights in this field.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by National Institutes of Health grant R01 HL58241-11, through the American Recovery and Reinvestment Act of 2009, U.S. Defense Threat Reduction Agency grant HDTRA-09-1-0013, and the Vanderbilt Institute for Integrative Biosystems Research and Education. We thank Allison Price for her editorial assistance, and two anonymous referees for their thoughtful comments.

References

- Casella, G.; Berger, RL. Statistical Inference. 2nd. Duxbury Press; 2001.
- Chatterjee S, Bose A. Generalized bootstrap for estimating equations. *The Annals of Statistics*. 2005; 33:414–436.
- Crainiceanu CM, Staicu AM, Ray S, Punjabi N. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine*. 2012; 31:3223–3240. [PubMed: 22855258]
- Davison, AC.; Hinkley, DV. Bootstrap Methods and Their Applications. Cambridge University Press; Cambridge: 1997.
- Duggins, JW. Parametric resampling methods for retrospective changepoint analysis PhD thesis. Virginia Polytechnic Institute and State University; 2010.
- Edelstein-Keshet, L. Mathematical Models in Biology. Society for Industrial and Applied Mathematics; 2005.
- Efron B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*. 1979; 7:1–26.
- Efron B. Better bootstrap confidence intervals. *Journal of the American Statistical Association*. 1987; 82:171–185.
- Eklund SE, Thompson RG, Snider RM, Carney CK, Wright DW, Wikswa JP, Cliffl DE. Metabolic discrimination of select list agents by monitoring cellular responses in a multianalyte microphysiometer. *Sensors*. 2009; 9:2117–2133. [PubMed: 22574003]
- Enders JR, Marasco CC, Kole A, Nguyen B, Sevugarajan S, Seale KT, Wikswa JP, McLean JA. Towards monitoring real-time cellular response using an integrated micro-fluidics matrix-assisted laser desorption ionisation/nanoelectrospray ionisation mobility-mass spectrometry platform. *IET Systems Biology*. 2010; 4:416–427. [PubMed: 21073240]
- Faley S, Seale K, Hughey J, Schaffer DK, VanCompernelle S, McKinney B, Baudenbacher F, Unutmaz D, Wikswa JP. Microfluidic platform for real-time signaling analysis of multiple single T cells in parallel. *Lab Chip*. 2008; 8:1700–1712. [PubMed: 18813394]
- Fedorov V. Optimal experimental design. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2010; 2:581–589.

- Field CA, Welsh AH. Bootstrapping clustered data. *Journal of the Royal Statistical Society; Series B*. 2007; 69:369–390.
- Garcia J, Sims KJ, Schwacke JH, Del Poeta M. Biochemical systems analysis of signaling pathways to understand fungal pathogenicity. *Methods in Molecular Biology*. 2011; 734:173–200. [PubMed: 21468990]
- Gonçalves S, White H. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*. 2004; 119:199–219.
- Gonçalves S, White H. Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association*. 2005; 100:970–979.
- Harrell, FE, Jr. *Regression Modeling Strategies*. Springer-Verlag New York, Inc; Secaucus, NJ, USA: 2006.
- He H, Severini TA. Asymptotic properties of maximum likelihood estimators in models with multiple change points. *Bernoulli*. 2010; 16:759–779.
- Hinkley D, Schechtman E. Conditional bootstrap methods in the mean-shift model. *Biometrika*. 1987; 74:85–93.
- Hollander, M.; Wolfe, D. *Nonparametric statistical methods*. Wiley series in probability and statistics: Texts and references section Wiley; 1999.
- Hu F, Kalbfleisch JD. The estimating function bootstrap. *Canadian Journal of Statistics*. 2000; 28:449–481.
- LeDuc PR, Messner WC, Wikswo JP. How do control-based approaches enter into biology? *Annual Review of Biomedical Engineering*. 2011; 13:369–396.
- López-Fidalgo J, Tomassi C, Trandafir PC. An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society, Series B*. 2007; 69:231–242.
- Maritz JS, Jarrett RG. A note on estimating the variance of the sample median. *Journal of the American Statistical Association*. 1978; 73:194–196.
- May RM. Uses and abuses of mathematics in biology. *Science*. 2004; 303:790–793. [PubMed: 14764866]
- Nobre J, Singer J. Residual analysis for linear mixed models. *Biometrical Journal*. 2007; 49:863–875. [PubMed: 17638292]
- Snider RM, McKenzie JR, Kraft L, Kozlov E, Wikswo JP, Cliffl DE. The effects of cholera toxin on cellular energy metabolism. *Toxins*. 2010; 2:632–648. [PubMed: 22069603]
- Sohn RA, Menke W. Application of maximum likelihood and bootstrap methods to nonlinear curve-fit problems in geochemistry. *Geochemistry Geophysics Geosystems*. 2002; 3:1525–2027.
- Wegman EJ, Wright IW. Splines in statistics. *Journal of the American Statistical Association*. 1983; 78:351–365.
- Wikswo JP, Lin SF, Abbas RA. Virtual electrodes in cardiac tissue: a common mechanism for anodal and cathodal stimulation. *Biophysical Journal*. 1995; 69:2195–2210. [PubMed: 8599628]
- Wikswo JP, Prokop A, Baudenbacher F, Cliffl D, Csukas B, Velkovsky M. Engineering challenges of BioNEMS: the integration of microfluidics, micro-and nanodevices, models and external control for systems biology. *IEEE Proceedings –Nanobiotechnology*. 2006; 153:81–101. [PubMed: 16948492]

Figure 1.

Raw action potential durations for 14 isolated rabbit hearts. Each heart was observed for up to 72 minutes. The vertical gray bars represent the periods of anoxia challenge.

OLS + Bootstrap Strategy

Figure 2.

Ordinary least squares model fit. The top panel shows raw APD measurements (dotted lines) adjusted vertically by subtracting the corresponding $\hat{a}_j$. Estimated mean APD curves (solid lines) are overlaid. The lower panel shows the estimated difference in mean APD for enriched versus normal media groups, with pointwise **bootstrap** 95% confidence band. In either panel, vertical gray bars delineate the periods of anoxia challenge.

Figure 3.

Residual analysis of the OLS model fit. Upper panel: estimated residual autocorrelation among time series measurements, in lags up to 52 minutes. Lower panel: estimated residual variance at each time where measurements were collected. The zero-lag time in the upper panel was aligned with time zero in the lower panel to illustrate the effect of repeat anoxia challenge on the estimated autocorrelation. However, note that autocorrelation estimates are increasingly uncertain for higher lag times. The autocorrelation for lags near 26 minutes are artifacts of the experimental design (i.e., repeat anoxia). No significant autocorrelation was observed among the MLE standardized residuals (see Web Figure 2).

Figure 4.

Estimated pointwise power and coverage for each of three strategies. The upper panel displays the pointwise 95% confidence bands for each method using the original cardiac APD data. The middle panel illustrates the estimated power associated with tests of the pointwise null hypothesis $H_0 : D_E(t) - D_N(t) = 0$. The gray horizontal lines represent typical values of interest (i.e., 0.80 and 0.05). The lower panel shows the estimated coverage for 95% confidence bands. The gray horizontal line represents the nominal coverage (0.95).