Adv Physiol Educ 28: 2–14, 2004;
10.1152/advan.00042.2003.

Statistical analyses of repeated measures in physiological research: a tutorial

Michael Kristensen[1] **and Thomas Hansen**[2]
[1]*August Krogh Institute, University of Copenhagen, and* [2]*Department of Epidemiology and Biostatistics, National Institute of Public Health, DK-2100 Copenhagen Ø, Denmark*

Received 22 September 2003; accepted in final form 29 November 2003

Kristensen, Michael, and Thomas Hansen. Statistical analyses of repeated measures in physiological research: a tutorial. *Adv Physiol Educ* 28: 2–14, 2004; 10.1152/advan.00042.2003.—Experimental designs involving repeated measurements on experimental units are widely used in physiological research. Often, relatively many consecutive observations on each experimental unit are involved and the data may be quite nonlinear. Yet evidently, one of the most commonly used statistical methods for dealing with such data sets in physiological research is the repeated-measurements ANOVA model. The problem herewith is that it is not well suited for data sets with many consecutive measurements; it does not deal with nonlinear features of the data, and the interpretability of the model may be low. The use of inappropriate statistical models increases the likelihood of drawing wrong conclusions. The aim of this article is to illustrate, for a reasonably typical repeated-measurements data set, how fundamental assumptions of the repeated-measurements ANOVA model are inappropriate and how researchers may benefit from adopting different modeling approaches using a variety of different kinds of models. We emphasize intuitive ideas rather than mathematical rigor. We illustrate how such models represent alternatives that *1*) can have much higher interpretability, *2*) are more likely to meet underlying assumptions, *3*) provide better fitted models, and *4*) are readily implemented in widely distributed software products.

experimental design; longitudinal data; analysis of variance; nonlinear mixed effects models

EXPERIMENTS INVOLVING REPEATED MEASUREMENTS, i.e., several consecutive measurements, on experimental units (laboratory animals or human subjects) subjected to different treatments are commonly encountered in physiological experiments (e.g., Refs. 3, 7–9, 12, 13, 15–17, 19, 20, 23–26, 30–32, 35, 37). Descriptions of the experimental protocols and various figures showing the data suggest that many studies in the archives of, for example, the journals published by the American Physiological Society and the references cited above are published wherein somewhat inappropriate statistical properties of the data are assumed (but see, e.g., Refs. 4, 11, 21, 22, 29, 33). Statistical analyses based on untenable assumptions may produce "correct" conclusions but may at worst produce meaningless and false results. However that may be, conclusions based on inappropriate statistical models should not be relied on and should therefore be avoided (1, 2).

The aims of this paper are *1*) to explain and illustrate some of the principal properties of repeated-measurements data that need to be considered during data analysis, *2*) to illustrate the primary strengths and shortcomings of the apparently so widely used repeated-measurements ANOVA, and *3*) to give examples of other types of analysis that may not only remedy shortcomings of the repeated-measurements ANOVA but that are also readily implemented in widely distributed software packages like SAS and S-PLUS/R. In our summarizing discussion, we give examples of situations where the repeated-measurements ANOVA may very well be a valid choice of statistical model.

To address the nonstatistician audience, the discussion is kept at an intuitive rather than a mathematical level. To make the discussion more readable, we use an experiment designed to test the effects of pinacidil on muscle fatigue as an example to explain the different statistical issues that should be considered during data analyses.

PROPERTIES OF THE DATA

The data sets of all the above-cited studies and the example given below share certain properties. First, the experiments involve two or more "treatment groups" [typically a placebo/control and some treatment(s)]. Second, a number of experimental units are subjected to either one or all of the different treatments. Third, the experimental units are measured consecutively. To concretize these properties, consider the following experiment (M. Kristensen, unpublished data).

Experimental Setup

Male Wistar rats weighing 80 ± 5 g were anesthetized with mebumal delivered intraperitoneally at a dose of 5 mg/100 g body wt and then killed by cervical dislocation. Before the experiment, the animals were kept at 20°C with day/night lengths of, respectively, 10 and 14 h. Animals were fed ad libitum. The handling of animals was in accordance with Danish Animal Welfare Regulations.

Address for reprint requests and other correspondence: T. Hansen, Dept. of Epidemiology and Biostatistics, National Institute of Public Health, Svanemøllevej 25, DK-2100 Copenhagen Ø, Denmark (E-mail: THa@niph.dk).

1043-4046/04 $5.00 Copyright © 2004 The American Physiological Society

Both soleus muscles were excised from each animal immediately and were randomly selected to be placed into a Krebs-Ringer solution (in mM: 122 NaCl, 25 $NaHCO_3$, 2.8 KCl, 1.2 KH_2PO_4, 1.2 $MgSO_4$, 1.3 $CaCl_2$, 5.0 D-glucose) with or without pinacidil. The Krebs-Ringer solution was equilibrated at room temperature, before and throughout the experiment, with a mixture of 5% CO_2-95% O_2 (pH 7.4).

After 1 h of incubation in either placebo or 100 μM pinacidil-Krebs-Ringer solution, the muscles were adjusted to produce the same passive force on a force transducer and then stimulated once. After another 5 min of incubation, the muscles were stimulated to fatigue [1-s-long trains (33-Hz pulse) with 2-s brake between each train continuing for 7 min]. Force data were recorded with an A/D converter (Duo-18, version 1.1). Data were recorded every 30 s, giving rise to 15 different time points. The reduction in force was measured relative to the first time point at $t = 0$ by transforming the raw force measurements x_{tij} according to

$$y_{tij} = 100 \times x_{tij}/x_{0ij} \qquad (1)$$

We index the observations as follows: time is indexed by $t = 0, 1, \ldots, T$ (here $T = 14$); treatment group is indexed by $i = 1, 2, \ldots, I$ (here $I = 2$) and finally experimental units (animals) are indexed by $j = 1, 2, \ldots, n$ (here $n = 7$). The data appear in Table 1 and in Fig. 1.

Controlling Variability Among Experimental Units

Often, experimental units vary in different but more or less uninteresting ways. Laboratory animals may differ subtly in size and other characteristics. Such differences are likely to be manifested in our measurements. Usually, among experimental units variability will be considered "noise" that needs to be controlled statistically rather than something that one is specifically interested in.

Another aspect of the variability among experimental units is if different units react differently to the treatments. Figure 1 provides an example of this problem. *Animal VI* "flips" the treatment effect. With this animal, the placebo curve is below the pinacidil curve, whereas the other six animals show the opposite pattern of reaction. We will refer to this phenomenon as an "interaction" between experimental units and treatment. Such interactions pose considerably more trouble than simple variability among experimental units because it means that any treatment effect cannot generally be assessed. Even if the average pinacidil curve had been considerably below the average placebo curve, one cannot conclude that pinacidil speeds up fatigue when some animals flip the treatment responses. The reason is that this conclusion can be assessed only for those animals that do not flip the curves. Obviously, the problematic flippers give no weight to the conclusion that pinacidil speeds up fatigue.

The different ways that these two kinds of variability are handled depend on one's choice of statistical model and will be discussed later.

Repeated Measurements: Controlling Variability Within Experimental Units

Whereas observations from different experimental units may very well be independent–assuming proper experimental design–different observations from the same animal are probably not independent. In fact, observations from the same experimental unit are likely to resemble one another compared with observations from different experimental units. We say that observations from the same experimental unit are correlated (6, 14, 34). In our example, there are two levels of within-experimental unit correlations.

First, the two different soleus muscles from each animal sujected to the placebo and the pinacidil treatments respectively may be correlated. Because the muscles in each pair originate from the same animal they are likely to resemble each other more than they resemble randomly selected muscles from other animals. Furthermore, as the experiment progresses, correlations between the two muscles within each pair are likely to ease, because treatment effects will start to override biologically (e.g., genetically) based correlations.

Second, more important, however, is the correlation among consecutive measurements of the same muscle: measurements taken 30 s apart are highly correlated, and even measurements further apart are likely to be correlated. To illustrate this, we correlate all observations taken at time point $t = 1, 2, \ldots, 13$ to all subsequent observations $t' = 2, 3, \ldots, 14$ (we ignore the

Table 1. *Entire data set of relative forces*

Animal	*I*		*II*		*III*		*IV*		*V*		*VI*		*VII*	
Time	Placebo First and Pinacidil Second													
0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	79.8	42.7	81.2	79.0	66.7	64.2	64.3	47.6	81.5	88.1	78.9	83.7	96.3	82.5
2	68.2	29.9	70.8	66.4	56.3	48.8	48.4	30.6	69.8	77.9	65.6	74.3	86.9	73.0
3	59.6	19.7	63.1	50.4	52.4	39.8	37.7	21.0	62.5	63.9	56.3	69.3	76.6	61.5
4	53.5	12.8	58.1	42.9	46.4	30.9	31.6	17.3	56.5	46.7	47.7	59.9	71.5	51.5
5	49.5	10.7	53.1	37.8	40.1	25.6	29.5	16.5	54.0	38.9	40.2	53.7	60.7	46.0
6	45.5	9.4	50.8	34.5	34.5	22.0	27.0	15.3	50.0	30.7	37.5	49.0	56.1	41.0
7	41.4	8.5	45.4	31.9	30.2	21.1	23.8	14.5	45.2	25.8	32.0	44.4	50.5	35.5
8	36.4	8.1	40.8	31.1	26.2	20.3	20.5	12.9	41.5	22.1	29.3	40.1	45.3	31.0
9	32.3	7.7	37.7	29.8	23.0	18.7	17.6	12.9	37.1	19.7	27.0	36.6	39.3	28.5
10	29.8	7.3	34.2	28.2	20.6	17.5	15.6	11.7	29.8	18.0	21.9	33.5	33.6	25.0
11	26.8	7.7	29.6	28.2	19.0	16.7	13.9	10.9	27.4	16.4	19.9	31.1	29.0	23.0
12	24.2	7.7	27.7	26.1	18.3	15.4	12.7	10.9	26.6	15.6	18.0	28.8	26.2	21.5
13	22.2	7.7	26.5	26.1	18.3	15.0	11.5	12.5	25.4	15.2	17.6	27.2	23.8	20.0
14	20.2	7.7	26.2	24.4	18.3	15.0	11.1	10.9	23.4	14.8	17.2	25.3	21.0	19.0

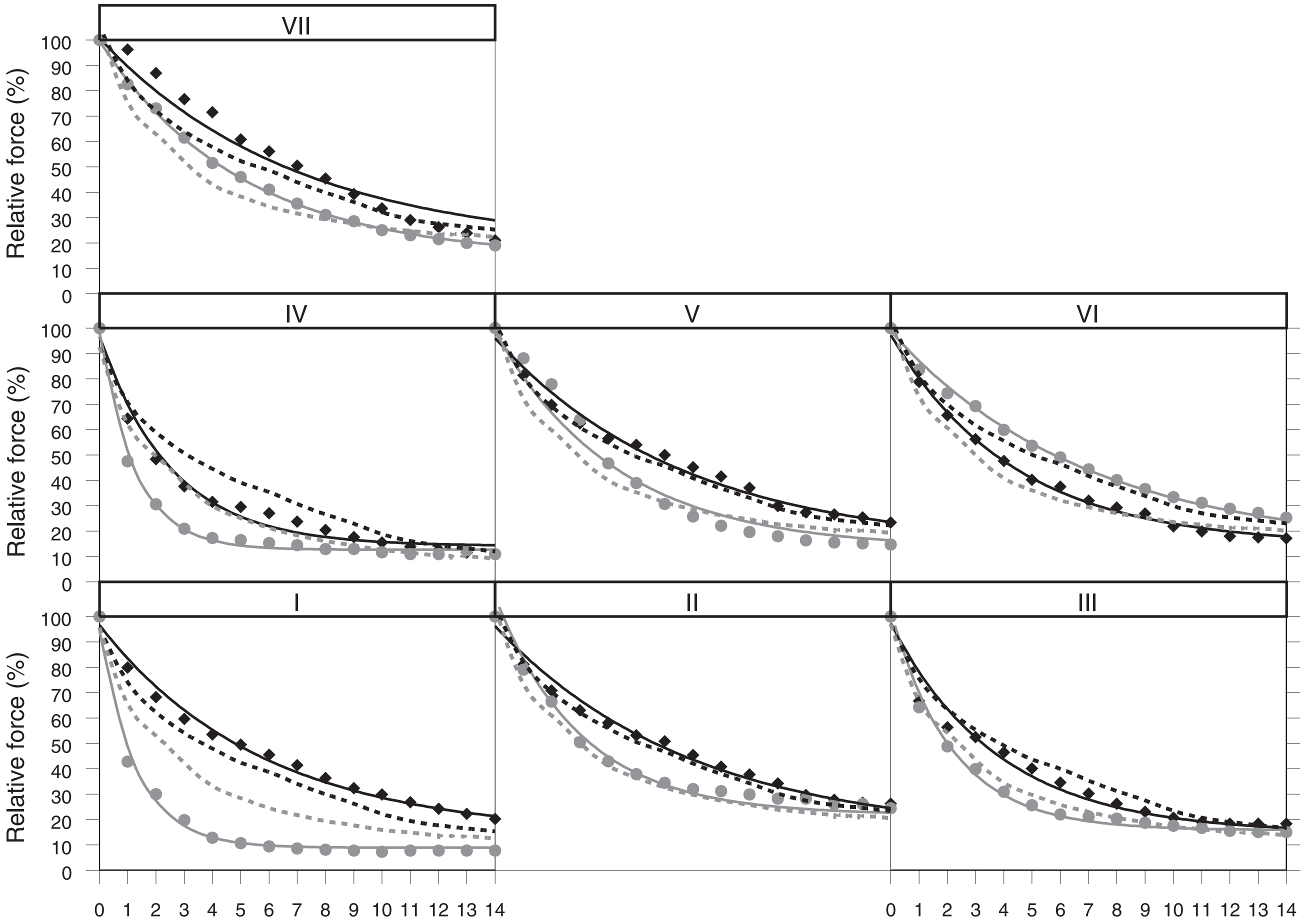

Fig. 1. Panels represent each of the 7 experimental units (animals). ◆/black represents the placebo group; ●/gray the pinacidil groups. Solid lines are fitted from the nonlinear mixed-effects model (M4) presented later in the article; dotted lines are fitted from the M1 model, presented later in the article. Abscissa represents time $t = 0, 1, 2, \ldots, 14$.

first time point). This corresponds to correlating all rows in Table 1 to all subsequent rows and then proceeding until one correlates time points 13 and 14. The result is $(14 \times 13)/2 = 91$ correlation coefficients. Now, 13 of these correlations are between observations that are one time step apart $[(t,t') = (1,2), (2,3), \ldots, (13,14)]$, 12 are between observations that are two time steps apart $[(t,t') = (1,3), (2,4), \ldots, (12,14)]$, and so on. The 91 correlation coefficients are depicted in Fig. 2, grouped into categories according to the length of the time distance between observations.

Figure 2 shows that *1*) the correlation coefficients are generally high ($r \gtrsim 0.75$), and *2*) there is an almost monotonous decrease in the correlation as the distance between consecutive measurements increases.

Treatment Groups

In most experiments, interest centers on systematic differences among two or more treatment groups. In our pinacidil experiment, we are interested in whether the onset of fatigue is more rapid in the muscles that received the pinacidil treatment compared with muscles receiving the placebo treatment. How to characterize such differences depends on the particular statistical model one chooses. This choice, however, must accommodate the discussed different properties of the data at hand.

STATISTICAL MODELS

In this section, we will discuss various types of statistical analyses designed to handle types of data sets like the one considered in our example. Additionally, we will focus on ways of adapting one's data to be fitted by relatively simple statistical models. If one can accommodate one's data to such simple models, e.g., by log-transforming the response or by considering only some part(s) of the entire data set, this is clearly worthwhile doing.

We will discuss several types of models, each with its virtues and shortcomings. The models we will consider are

- M1: repeated-measurements ANOVA on y_{tij}
- M2: linear mixed-effects model on $\log(y_{tij})$ and using fewer of the consecutive measurements
- M3: nonlinear model for each set of observations from each muscle, coupled with a two-way ANOVA on parameter estimates
- M4: nonlinear mixed-effects model

For further discussion see Refs. 6, 14, 25, 34, and 36. All the models considered are relatively straightforward to implement in SAS and S-PLUS/R.

We by no means wish to imply that the four models constitute an exhaustive list of possible applicable models. Rather,

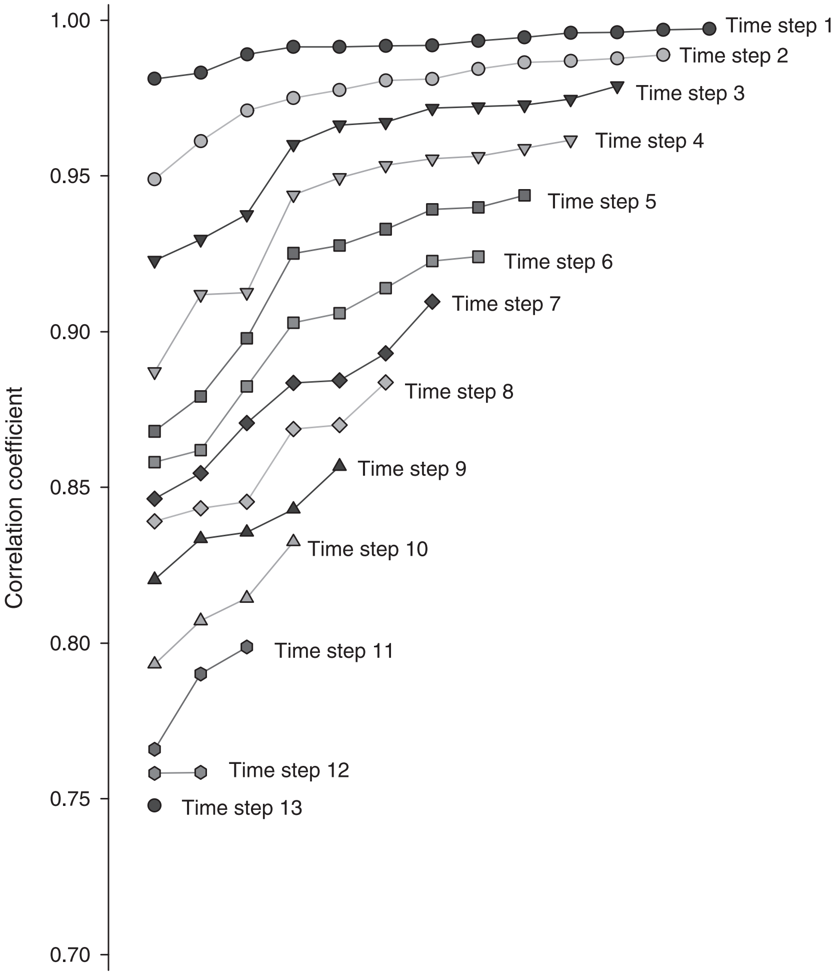

Fig. 2. Development in correlation among observations taken at different time points (minus the first $t = 0$ time point). Different curves represent correlations between observations spaced 1, 2, . . ., 13 time points apart, respectively.

the models are chosen to illustrate various aspects of the data set at hand and different ways of approaching the problem of analyzing the data.

The first two models we consider are ANOVA-like models insofar as they treat time as a *qualitative* variable without any particular order among the 15 levels (time points). On the other hand, the two latter models regard time as a *quantitative* (and continuous) variable.

Notation

To be able freely to discuss various statistical features of the considered models, some notation must be agreed on.

A *statistical model* is a functional relationship between the response y_{tij} and certain *predictors* or *explanatory variables*. In our case, predictors are time, treatment group, animal ID, and possible combinations thereof.

During the statistical analysis we estimate *effects* of the predictors. An effect is a scalar with which we want to characterize the effects of, e.g., treatment groups or time on the response. One can think of an effect as, e.g., a regression coefficient in a linear regression, the mean difference between two treatments, or the variance of the response among subjects.

Having estimated the effects, we also say that *the model has been estimated*. Using the estimated model, we can *predict* the outcome, which we denote using a ★ superscripted to the variable in question, e.g., $y^{\star}_{tij}$. For example, if one fits a linear regression model $q(t) = a + (b \times t)$ to some data pairs (t,q) and estimates $a^{\star} = 2$ and $b^{\star} = -0.5$, the predicted outcome at $t = 7$ is $q^{\star} = -1.5$.

A model's fit to the data can be measured by the *residuals*, which are here defined as the difference between the observed data and the predicted values $r^{\star}_{tij} = y_{tij} - y^{\star}_{tij}$. Thus for each observation there is one corresponding residual.

M1: Repeated-Measurements ANOVA Model

The repeated-measurements ANOVA model may be thought of as designed to assess treatment differences while controlling between-subject variability when each of these is measured "a few" consecutive times. The model is, as such, simple to interpret and does–apparently–take into account the various aspects of the repeated-measurements data set (6, 14, 34). Furthermore, it is readily implemented in many software packages, and this is presumably the reason why many researchers adopt the model. The interesting question is, however, whether it performs sufficiently well.

In our discussion here of the repeated-measurements ANOVA model, we focus on four of its shortcomings that are readily checked and that are likely to be encountered in connection with its use in physiological research. Thus we do not discuss all assumptions of the model, insofar as many of these are unlikely to be violated to any detrimental degree. The four aspects of the model are *1*) interpretability, *2*) model fit, *3*) within-experimental-unit correlations, and *4*) variance homogeneity of the residuals among the 15 different time points.

Analysis and results. Table 2 shows the results of applying the repeated-measurements ANOVA model to the data in Table 1. The effect of time is obvious, indicating that the relative strength force changes (decreases) over time. Furthermore, there is a significant time × treatment interaction. This corresponds to the conclusions that can be reached by a set of time-by-time *t*-tests: there are differences between treatment groups at some time points but not at others. The treatment effect is almost significant ($P = 0.0839$), but, due to the significant time × treatment interaction, we conclude that the pinacidil treatment has some effect, although it varies over time. It is important to stress that, under the repeated-measurements model, there is an effect of the pinacidil treatment notwithstanding that it varies over time.

Interpretability. The interpretation of a treatment effect depends on the presence/absence of whether the treatment factor enters into interaction terms (like time × treatment or animal × treatment).

ABSENCE OF INTERACTION TERMS. The interpretation of a treatment effect is that, averaged across the 15 time points and the seven animals, the mean pinacidil strength force differs from the mean placebo strength force. It means that we expect a constant difference between the two treatments. Given the expected physiological effects of pinacidil (see, e.g., Ref. 17) we may predict that *1*) as time progresses, the relative strength force becomes similar in the two treatment groups; and *2*) it is the speed with which the relative strength force decreases that is believed to be affected by pinacidil (see Fig. 1). These expectations have no bearing on any constant difference between the two treatments. Therefore, the interpretation of a treatment effect is completely detached from the biological reality of the experiment.

PRESENCE OF INTERACTION TERMS. In this scenario, the interpretation of a treatment effect becomes close to biologically meaningless. For example, if there is an interaction between time and treatment, any inference about the effect of treatments must be qualified or conditioned by particular time points. In our case, it implies that the speed-up effect that pinacidil has on fatigue is present at *time points 2–5*, say, but not at other time points. Similarly, interactions between animals and treatment imply that discussion of treatment effects must be qualified by stating which animals we are discussing (all but the flipper, *animal VI*, say).

Naturally, the expected effects of pinacidil are corroborated by the significant time × treatment interaction.

Model fit. The second problem with the model is its lack of fit. Any reasonable model must be able to predict the observations to some reasonable degree of accuracy. When the observations to the predicted values are compared (Fig. 1), it is obvious that the fit is very poor. Especially, the four animals, *animals I*, *IV*, *VI*, and *VII*, fit the model very poorly. Put simply, this indicates that the model is incorrect.

Within-experimental-unit correlations. A third drawback of the repeated-measurements ANOVA model regards the way it handles the within-experimental unit correlations (Fig. 2). When applying the repeated-measurements ANOVA model, one assumes that the correlations among the residuals from the same muscle remain constant across different time points. One way to check this assumption is described in the discussion of the M2-model. Another way is to consider plots like Fig. 2. Although this figure shows the correlations among observations rather than their corresponding residuals, the conspicuous systematic changes (decrease) in the correlation coefficients–as the time distance between consecutive measurements increases–strongly suggest that assuming a constant correlation across time points is untenable. However, plots like Fig. 2 do not indicate "how wrong" this assumption is.

Variance homogeneity. The fourth and last of the problems that we will discuss in connection with the repeated-measurements ANOVA model is the assumption of variance homogeneity. When using the models discussed here, one assumes that the variances of the residuals at each of the 15 different time points are more or less constant. To assess this graphically, one can plot the residuals from each time point as box plots (Fig. 3). The distance between the end points of each "whisker" indicates variability in the residuals at each time point. Clearly, this variability is not constant. There are 15 variances (one for each time point), and the range thereof is 14.6–154.0 (*time points 14* and *2*, respectively). On the basis of Fig. 3 alone, the assumption of a constant variance through time is clearly unrealistic.

M2: Linear Mixed-Effects Model on log(y_{tij})

The discussion above provides a clear example of some shortcomings of the repeated-measurements ANOVA model. However, several of the encountered problems can easily be remedied by *1*) log-transforming the response, *2*) considering fewer time points (e.g., every 2nd observation), and *3*) allowing a somewhat more elaborate correlation pattern among the consecutive measurements of the same muscle than just one constant correlation.

Briefly, the reasons to expect that these three actions may remedy the problems encountered are as follows. If the onset of fatigue follows an exponential decay–and Fig. 1 might suggest just that–log-transforming the data would linearize the response curves. And linear response curves are much easier to fit. Furthermore, log-transforming the data reduces overall

Table 2. *Analysis of the repeated-measurements ANOVA model (M1) using SAS proc mixed*

Source	*df*	Type II SS	Mean Square	*F* Value	*P* Value
Treatment*	1	3,651.85	3,651.85	4.29	0.0838
Animal*	6	12,612.42	2,102.07	2.47	0.1479
Animal × Treatment	6	5,107.82	851.30	27.31	0.0001
Time	14	108,317.51	7,736.96	248.18	0.0001
Time × Treatment	14	1,071.34	76.52	2.45	0.0036
Error	168	5,237.47	31.18		

*The factors Treatment and Animal are tested using the MS Animal × Treatment as an error term. The rest of the terms are tested against the Error MS.

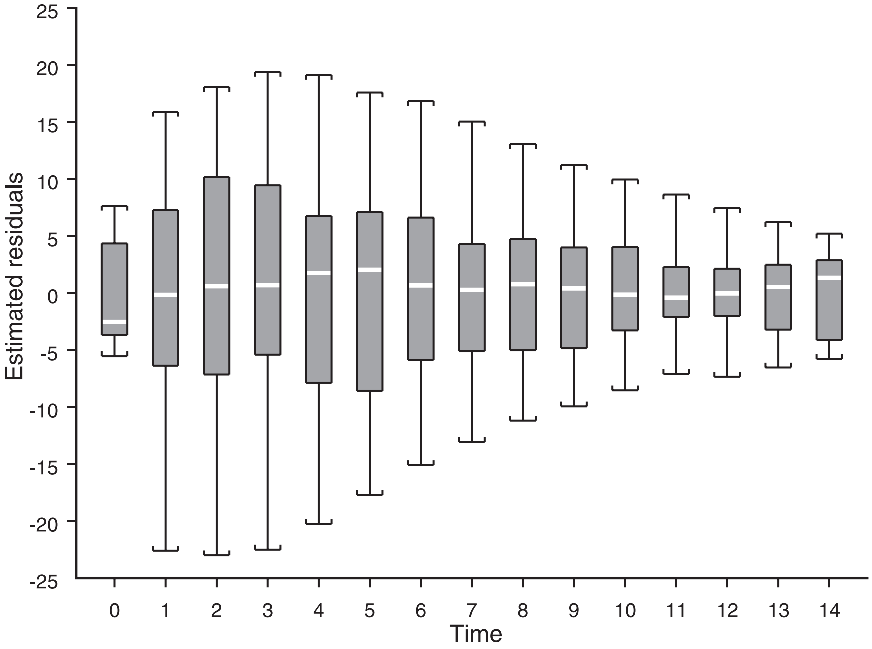

Fig. 3. Box plots of the estimated residuals from the split-plot model at the 15 different time points. The 2.5% (97.5%), 25% (75%), and 50% (median) percentiles are indicated as whiskers, bars, and white lines inside bars. Averages of the residuals at each time point are all zero. Distances between end points of whiskers illustrate variability at the different time points.

variance in the data and may thus help to remedy the problems of variance heterogeneity (Fig. 3). Finally, with the use of untransformed data, the treatment effect corresponds to comparing the difference between two means $d = y(\text{pinacidil}) - y(\text{placebo})$ [where the y(pinacidil) and y(placebo) represents the overall means within each of the two treatments] to zero. If d differs from zero, the two treatments differ. We argued above that d has little biological relevance. However, if we analyze $\log(y_{tij})$ rather than y_{tij}, the treatment effect becomes a difference on a log scale, and back-transformed to the original scale this difference becomes a ratio $d' = y(\text{pinacidil})/y(\text{placebo})$. And on the original scale, assuming a constant *factor* between the two treatments is probably not so far-fetched. The reason for analyzing fewer time points is simply that, in this way, the model becomes simpler insofar as fewer between-time points correlations must be handled. The last action is in line therewith. The more complex the between-time points correlations can be modeled, the more reasonable the model becomes.

Analysis and results. Our first objective is to choose how we wish to model the way that correlations among observations at consecutive time points depend on the time distance among the observations. One common way to do this is to fit a model with only the factors of direct interest, including a term for each of the different time points (a so-called saturated model) and obtain the residuals from this analysis (6, 34). In our case, this model is an ordinary two-way ANOVA, with treatment, time, and their interaction included. Having calculated the residuals thus obtained, we correlate all observation pairs that are one, two, and up to 13 time steps apart (we ignore the first $t = 0$ time point). Figure 4 shows the results of these calculations

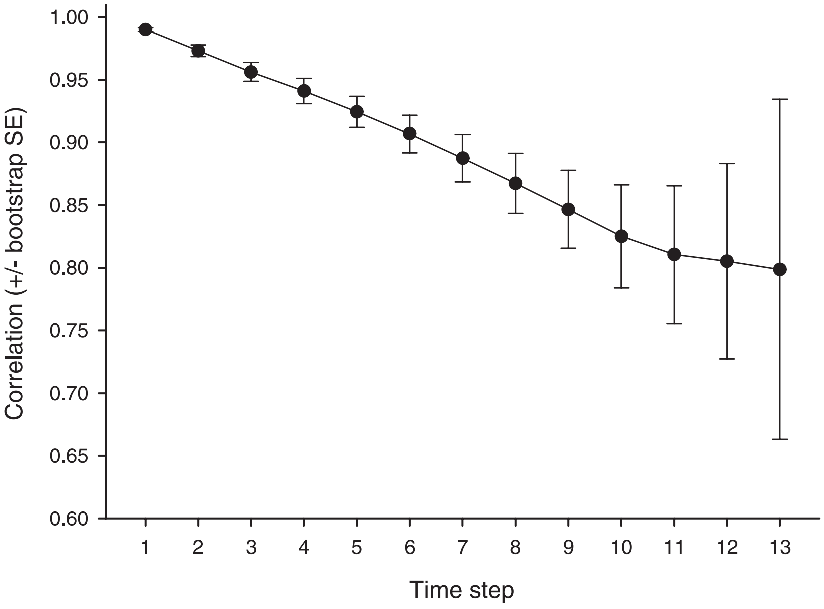

Fig. 4. Decrease in correlation among residuals from different time points (minus the first $t = 0$ time point) obtained from the saturated model (see text for explanation) analyzing $\log(y_{tij})$. SE bars are estimated by bootstrapping each of the 13 correlation coefficients 1,000 times.

based on a two-way ANOVA of log(y_{tij}), using observations taken at both even and uneven time points.

Considering the accuracy with which each correlation coefficient [denoted ρ(*h*), where *h* is the time step size] is estimated, it seems reasonable to expect that the correlation decreases according to

$$\rho(h) \approx k \times r^h \qquad (2)$$

where *k* is a constant, and $r \approx 1$. The reason is that Fig. 4 shows a close-to-linear decrease in ρ(*h*) with increasing *h*, and if $r \approx 1$, the power function (*Eq. 2*) is close to linear for moderate *h*. Thus performing a linear regression of $\log(\rho(h)) = \log(k) + h \times \log(r)$ and viewing log(*r*) as the unknown regression coefficient yields $r^\star = 0.98$ and $k^\star = 1.01$. In the statistical jargon used in the analysis of repeated measurements, a correlation structure like *Eq. 2* is termed a first-order autoregressive function, and it is readily implemented in software packages like SAS and S-PLUS/R (14, 25, 34).

Table 3 shows the result of the analysis of the pinacidil data with the three modifications discussed [log-transformation, dropping (every even) time points, and using the first-order autoregressive correlation structure]. The *Q*-tests are so-called likelihood ratio tests and are approximately χ^2-distributed with the degrees of freedom shown in the *df* column. The pinacidil treatment is, in this analysis, significant ($P = 0.0194$), with a log-average difference between the placebo and the pinacidil treatments of $d'^\star = 0.33$. This means that muscles in the placebo treatment are on average exp(0.33) = 1.39 *times* (or 39%) more powerful than muscles in the pinacidil treatment measured as the relative force. Naturally, the interaction between treatment and time still makes interpreting the treatment effect difficult, as previously discussed. Thus the model predicts that, as time progresses from $t = 1$ over $t = 7$ to $t = 13$, the ratio between the placebo and the pinacidil averages changes from 1.15 over 1.62 to 1.23. Hence the need to qualify the treatment effects by stating the time point at which one considers the difference/ratio between the two treatments.

Figure 5 shows the fit of this model to the log-transformed data. The fit is not great. Although the response curves have been considerably linearized, the fit still does not capture nonlinear features of the log-transformed data; we are still left with what seems to be a wrong model.

Apart from the poor fit between model and data, the three actions taken actually perform quite well. *1*) The correlation between adjacent time points ($r^\star$ from *Eq. 2* is 0.94, which corresponds very well to Fig. 4. *2*) Using Bartlett's test (28) to test variance homogeneity among the seven different time points yields that the variances have indeed been homogenized (data not shown). *3*) The variability among animals is no longer significant (data not shown), and the troublesome interaction between animals and treatment groups is confounded with the correlation between adjacent time points (6, 14, 34). *4*) Finally, the log-transformed data have a more meaningful biological interpretation, albeit still not perfect.

Table 3. *Analysis of the linear mixed-effects model on log(y_{tij}) using a first-order autoregressive correlation structure (M2) in SAS proc mixed*

Source	*df*	*Q*-Test	*P* Value
Treatment	1	5.47	0.0194
Time	6	628.91	0.0001
Time × Treatment	6	28.17	0.0001

Only observations taken at odd-numbered time points (1, 3, . . . , 13) are included.

NONLINEAR MODELS

One basic premise of the two applied models is that they are linear and that the response can be predicted by linear combinations of treatment and time effects. Many of the articles cited here present repeated-measurements data that are quite nonlinear (e.g., Refs. 7, 16, 17, 19, 20, 26, 30, 31, 35), so it seems natural to try to model the data by using nonlinear models. Above, it was argued that, if the force decay was exponential, log-transforming the response would yield a linear model. However, log(y_{tij}) is still not linear.

On the basis of knowledge of the mechanics of the experiment, some features of the data might be predicted a priori and help to build a reasonable nonlinear model. Looking at the figures showing the original data, one can argue that, because the experiment is terminated after 8 min, the decay in force is not allowed to reach zero. Rather, one could view the response curves as exponential plus some constant. Moreover, the model should encompass the fact that all observations equal 100% relative force at the first time point (the *y*-axis intercept). How precisely the onset of fatigue develops through time is probably not easy to predict, but assuming an exponential decay is at least flexible.

One function that accommodates these features is

$$y(t) = \alpha_1 \times \exp(\alpha_2 \times t) + \alpha_3 \qquad (3)$$

where the three parameters have the following interpretations: α_2 characterizes the decay in muscle force as time progresses. Obviously, $\alpha_2 < 0$. The more negative α_2 becomes, the faster the decay; α_3 is the asymptotic relative strength force as time progresses. Finally, the sum $\alpha_1 + \alpha_3$ is the *y*-axis intercept, i.e., the relative force at the first time point (should be close to 100). Additionally, from the equation $\alpha_1 \times \exp(\alpha_2 \times t_{½}) + \alpha_3 = (\alpha_1 + \alpha_3)/2$, we see that the half-time to the observed maximal fatigue is

$$t_{1/2} = \log\,([\alpha_1 - \alpha_3]/[2\alpha_1])/\alpha_2 \qquad (4)$$

There are several possible ways by which one can come from the mechanistic function (*Eq. 3*) to a statistical analysis. Here and in the following paragraph, we will describe two quite different methods.

When we consider the interpretation of the α-parameters and the expected physiological effects of pinacidil (speeding up fatigue), we could suggest the following (one-sided) hypothesis

$$\alpha_2(\text{placebo}) > \alpha_2(\text{pinacidil}) \qquad (5)$$

Whether the α_1 and α_3 should differ between treatments is less clear.

Hence, one strategy could be to let one or all three α-parameters depend on treatment group

$$y_{tij} = \begin{array}{ll} \alpha_{11} \times \exp(\alpha_{21} \times t) + \alpha_{31} & \text{for placebo} \\ \alpha_{12} \times \exp(\alpha_{22} \times t) + \alpha_{32} & \text{for pinacidil} \end{array}$$

and then test whether $\alpha_{11} = \alpha_{12}$, $\alpha_{21} = \alpha_{22}$, and $\alpha_{31} = \alpha_{32}$.

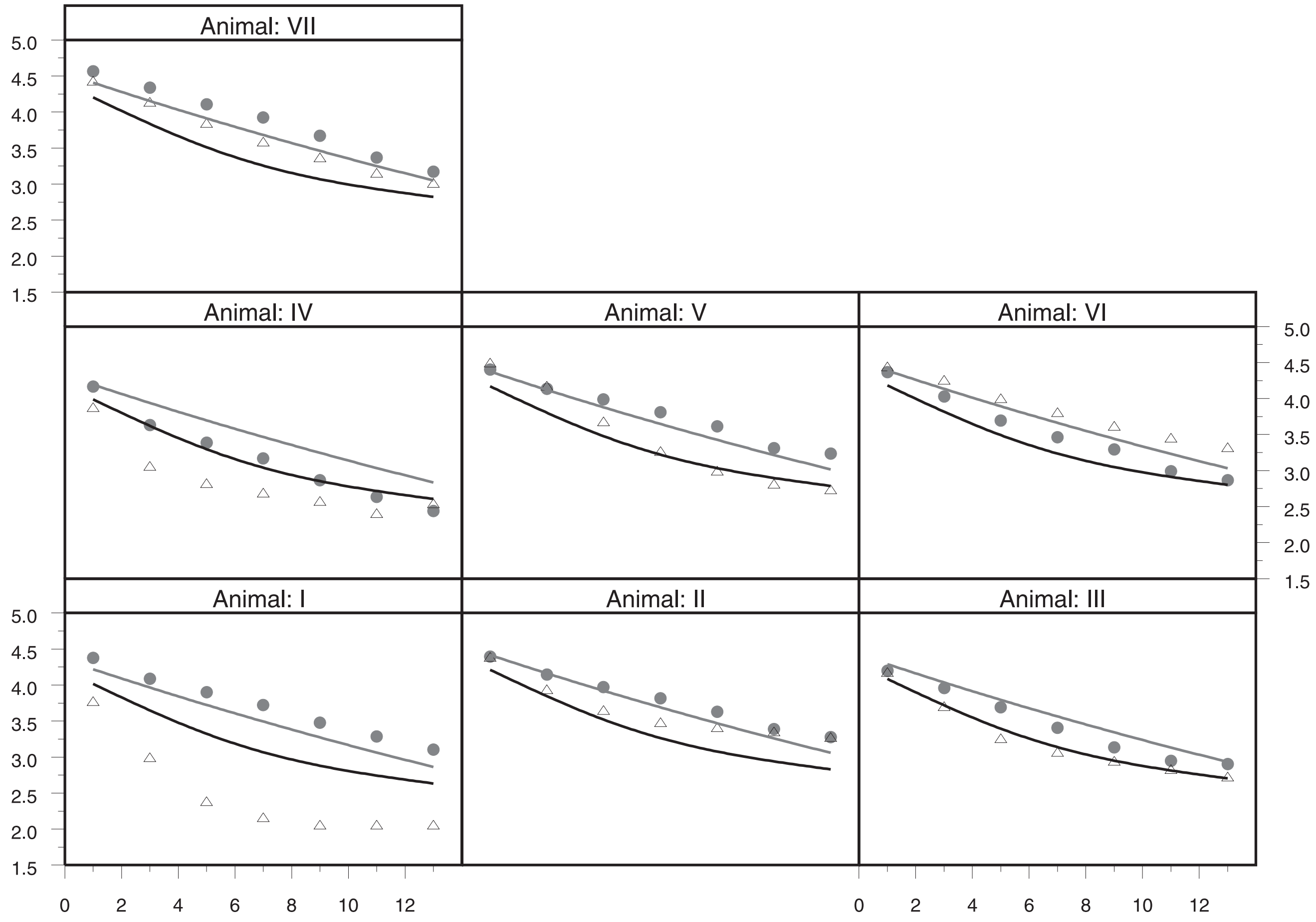

Fig. 5. Panels represent each of the 7 experimental units (animals). Symbols represent the observed log-transformed data. ●, placebo group; △, pinacidil group. Drawn lines, the fitted model to the log-transformed data and are based on a first-order autoregressive correlation structure and dropping every 2nd (even) time points. Gray symbols/lines, placebo group; black symbols/lines, pinacidil group.

Looking at Fig. 1, it is obvious that this model does not hold. The speed of onset of fatigue among animals varies clearly. For example, *animal I* responds very strongly to pinacidil (the onset of fatigue is rapid), whereas *animal VI* flips the treatments. Apparently, there is also a substantial variation in the asymptotic relative force among animals. The points raised here suggest that there is an individual animal-level variability in the α-parameters regardless of any treatment effects thereon.

M3: Paired t-Test or Two-Way ANOVA of α-Parameter Estimates

Because the α-parameters have reasonably obvious interpretations, one strategy to deal with the problem of animal-level variability in the responses could be to fit *Eq. 3* to each individual muscle. For a particular α-parameter (α_2, say), this would result in 14 different estimates of α_2 grouped into two treatments and seven animals. Because of the distinct interpretation, we could then proceed without loss of meaning, analyzing the 14 α_2 estimates by, e.g., paired *t*-tests or a two-way ANOVA (had there been more than two treatment groups) (6, 14).

Table 4 shows the results of performing this analysis. Using the paired *t*-test (thus controlling among animal variability), we find that, when the two treatment groups are compared, only the α_2-parameter differs marginally significantly, implying that the onset of fatigue is marginally faster in the pinacidil group compared with the placebo group. This conclusion is quite strong insofar as it is in compliance with pinacidil's expected physiological effects.

Note that this type of analysis makes little sense unless the parameters of the model can be interpreted in some relevant

Table 4. *Pairwise t-tests of parameter estimates*

	α_1		α_2		α_3	
Animal	Placebo	Pinacidil	Placebo	Pinacidil	Placebo	Pinacidil
I	79.89	90.18	−0.18	−0.81	15.77	8.33
II	77.16	75.46	−0.16	−0.34	18.47	25.23
III	76.89	81.47	−0.26	−0.45	16.41	16.40
IV	81.71	86.36	−0.39	−0.83	14.33	12.83
V	80.72	92.87	−0.15	−0.22	14.51	9.51
VI	84.05	81.42	−0.23	−0.16	13.78	16.49
VII	92.87	86.32	−0.13	−0.19	11.02	13.06
Mean	81.90	84.87	−0.22	−0.43	14.90	14.55
t-Test		−1.14		2.34		0.19
P Value		0.2995		0.0577		0.8548

The α-parameters are estimated for each muscle separately using SAS proc nlin. This method is referred to in the text as the M3 model.

biological way (6, 14). Moreover, the fit is naturally reasonably good insofar as each muscle has been fitted to *Eq. 3* (Fig. 6).

It is important to stress that the assumptions underlying the analysis above are quite simple. *1*) The model (*Eq. 3*) must be reasonably correct. *2*) The pairwise differences between group-specific parameter estimates are assumed to be normally distributed. This latter assumption is probably as good as any assumption, insofar as it is virtually impossible to check given only seven pairwise observations. Moreover, it is easy to overcome problems of nonnormality by using either nonparametric tests (e.g., the Mann-Whitney *U*-test) or, e.g., bootstrap statistics (5, 28). We have estimated a nonparametric bootstrap confidence interval for the difference between the group-specific α_2-parameters ($\alpha^{\star}_{21} - \alpha^{\star}_{22}$) and the resulting 95% confidence interval, based on 1,000 resamples, to be [0.06; 0.39], significantly above zero, corroborating the result from the *t*-test.

M4: Nonlinear Mixed-Effects Model

The last type of analysis considered may be thought of as a combination of the M1 and M3 models. The idea is simply that the nonlinear model (*Eq. 3*) apparently fits the data well when the α-parameters are allowed to vary among animals or muscles. So, if *Eq. 3* could be modified to incorporate treatment effects (which we are interested in) while allowing subject-specific variability in the α-parameters, the analysis could be conducted in one overall analysis.

To incorporate such variability, *Eq. 3* may be extended to encompass individual muscle-specific variation (6, 14, 25)

$$y_{tij} = \alpha_{1i} \times \exp([\alpha_{2i} + u2_{ij}] \times t) + [\alpha_3 + u3_{ij}] \qquad (6)$$

where subscripts are similar to those used in *Eq. 1*. The three α-parameters are thus allowed to vary between treatment groups. The two *u*-parameters characterize $u2_{ij}$: individual variation in the decay of relative strength force, and $u3_{ij}$: individual variation in asymptotic fatigue. The reason not to include a random variable associated with the α_1-parameter is that the sum $\alpha_{1i} + \alpha_{3i}$ is the *y*-axis intercept and that two random variables associated therewith would be highly correlated. To avoid this, the α_{1i}-parameter does not have an associated random variable.

A rigorous interpretation of this latter model (*Eq. 6*) is not as straightforward as one might prefer. The reason is that the expected value of this model differs from the right-hand-side expression of the model without random effects (*Eq. 3*) (6, 14, 25). However, the qualitative interpretation of the three α-pa-

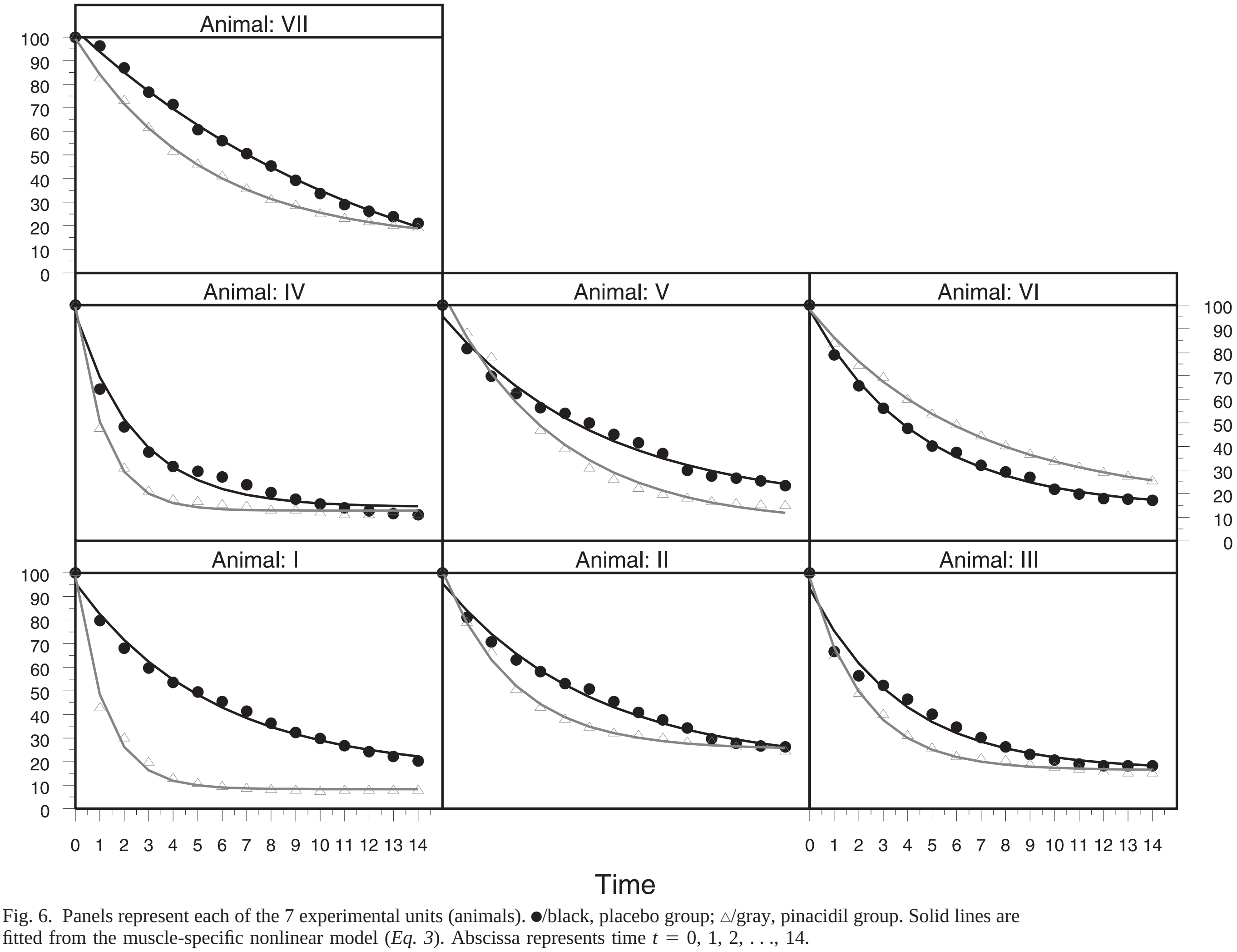

Fig. 6. Panels represent each of the 7 experimental units (animals). ●/black, placebo group; △/gray, pinacidil group. Solid lines are fitted from the muscle-specific nonlinear model (*Eq. 3*). Abscissa represents time $t = 0, 1, 2, \ldots, 14$.

rameters remains unchanged. One possible way of interpreting the α-parameters is to consider the effects of changing their values. Consider, for example, a lowering of α_{2i} from −0.20 to −0.40. In the model without any random muscle-specific components (*Eq. 3*), this lowering in α_{2i} halves $t_{1/2}$. However, in the model with random muscle-specific components (*Eq. 6*), we can state only that $t_{1/2}$ decreases. Alternatively, we can state that there is a positive correlation between α_{2i} and $t_{1/2}$ in the model with random animal-specific effects.

The parameters $u2_{ij}$ and $u3_{ij}$ are assumed to be normally distributed "errors" associated with the α_{2i}- and α_{3i}-parameters following a two-dimensional normal distribution. As such, they are characterized by the variance of the two u_{ij}-parameters (denoted σ_2^2 and σ_3^2, respectively) and the covariance between them (denoted $\sigma_{2,3}$; see below for explanation). Qualitatively, the idea of the two u_{ij}-parameters is that the different experimental units are not expected to react completely uniformly to the treatments and, hence, cannot be expected to follow one and the same functional relationship. Thus the two u_{ij}-parameters allow the model to harbor individual random variation owing to differences among muscles. Although the qualitative idea of the two u_{ij}-parameters remains, the interpretation of the two u_{ij}-parameters is not as straightforward as indicated above. The reason is that we, at least as a start, do not assume that the two u_{ij}-parameters are independent of one another; rather, the covariance–or equally, the correlation–characterizes the dependence between the two u_{ij}-parameters. To explain the meaning of this: if, for example, animals that, for whatever reason, lose relative force comparatively "fast" [meaning that $u2_{ij}$ is "small" (negative) in order for the exponential part of the model to decrease "fast"], if these animals, for whatever unknown reason, maintain a comparatively "high" asymptotic relative force [meaning that $u3_{ij}$ is "big" (positive)], the covariance is negative, and vice versa.

Table 5 shows the results of the nonlinear mixed-effects model. The presence of the variances and covariance of the two u_{ij}-parameters is considered as noise and is, as such, just complicating the analysis/interpretation, although they help to fit the model to the data. Therefore, before proceeding to test the three α-parameters (treatment effects) we first want, if possible, to test whether the two variances σ_2^2 and σ_3^2 and the covariance $\sigma_{2,3}$ contribute significantly to the fit of the model. Table 5, *top* and *middle*, shows the results of these tests. Fortunately, the covariance between the two u_{ij}-parameters is not significant and is hence removed from the model ($\sigma_{2,3}^{\star} = 0.245$, $P = 0.1681$). It comes as little surprise that the two variances associated with the two u_{ij}-parameters are very significant indeed ($\sigma_2^{2\star} = 0.035$, $P < 0.0001$ and $\sigma_3^{2\star} = 10.395$, $P < 0.0001$).

Table 5. *Analysis of the nonlinear mixed-effects model (M4)*

Parameter	Estimate	*df*	*Q*-Test	*P* Value	Placebo	Pinacidil
Tested against a model with unstructured variance-covariance matrix						
$\sigma_{2,3}$	0.245	1	1.9	0.1681		
Tested against a model with diagonal variance-covariance matrix						
σ_2^2	0.035	1	189.3	0.0001		
σ_3^2	10.395	1	24.4	0.0001		
Tested against a model with diagonal variance-covariance matrix						
α_1	2.858	1	2.8	0.5849	82.661	85.518
α_2	−0.204	1	4.1	0.0429	−0.216	−0.419
α_3	−0.465	1	0.0	1.0000	14.704	14.239

The *top* part of the table presents the analysis of the covariance; the *middle* part presents the analysis of the variances, and finally the *bottom* part presents the analysis of the fixed effects in the model. The Estimate column presents estimated covariance, variances and the differences in the fixed effects between the pinacidil and the placebo groups, respectively. The actual α-estimates for the two treatment groups appear in the *bottom right* corner of the table. The *Q*-tests are likelihood ratio tests, which are approximately χ^2-distributed with degree of freedoms as shown in the *df* column. All analyses are carried out using SAS proc nlmixed.

Last, we test the three α-parameters, i.e., whether the *y*-axis intercept, speed of onset of fatigue, and asymptotic fatigue differ between the two treatments. Only the α_{2i}-parameters differ between treatments (Table 5, *bottom*: $P = 0.0429$). The loss in relative force is thus faster in the pinacidil group ($\alpha_{22} = -0.42$, SE = 0.176) compared with the placebo group ($\alpha_{21} = -0.22$, SE = 0.073). This difference between the two groups may be quantified by stating that the $t_{1/2}$ in the placebo group is approximately twice that of the pinacidil group.

To complete our analysis, we assess whether the fitted model actually fits the data well (Fig. 1). The complete lines comprise the fitted model predictions, and for all animals the fitted model is very close to the actual data. This even regards the two most abnormal animals (*animals I* and *VI*). Finally, the interpretability of the model is high compared with the repeated-measurements ANOVA model. Especially, the two fixed parameters α_2 and α_3 have clear and logical interpretations.

CONCLUSIONS AND RECOMMENDATIONS

We have reviewed four different kinds of models analyzing the same repeated-measurements data. Each model has its pros and cons and is suitable in certain situations. Furthermore, we have illustrated and focused on features of the statistical analysis that need to be considered explicitly during the analysis of repeated-measurements data. In this final section, we will focus on *1*) checking of model assumptions, *2*) checking of model fit, *3*) the model's interpretability, *4*) other possible models, and finally *5*) software implementation. Throughout, we assume the discussion to center on repeated-measurements data. We will additionally assume that different experimental units are independent and that responses either comply with assumptions regarding normality of residuals or have been transformed to comply therewith.

Model Assumptions

It is important to bear in mind that any statistical analysis depends on the fulfillment of certain assumptions. Some assumptions are strong, in the sense that "much is needed" for their fulfillment, whereas others are comparatively weak. Assumptions that should routinely be checked include variance homogeneity of the residuals and correlation structure of the consecutive measurements on the same experimental units.

Plots like Figs. 2–4 are well suited in this respect. They can give hints to appropriate correlation structures and pointers to transformations that may help to remedy variance heterogeneity.

Model Fit

When performing a parametric statistical analysis, one assumes an underlying model, a functional relationship between response and explanatory variables. If the model does not fit the data (i.e., predicted values are "far" from observed values) the model is incorrect. Whether the lack of fit is sufficiently large to be detrimental to the analysis or acceptable is to a large degree subjective. However, obvious "wrong models" consistently predict values wrongly. For example, the repeated-measurements ANOVA model almost consistently over- or underpredicts values. The observed values are not randomly scattered around the predicted curves; they deviate consistently.

When one is assessing model fit, the most useful tool is plots of response and predicted values against one or more explanatory variables (time and/or treatments) preferably for each experimental unit (25).

Interpretability

A somewhat overlooked part of the data analysis is the interpretation of the statistical models (25). The repeated-measurements ANOVA model is well suited (its interpretation makes biological sense) when the responses in different treatments are reasonably parallel over time. This means that differences between treatments are reasonably constant. However, if responses are not parallel (and cannot be transformed to parallelism), or if treatment interacts with time, say, the interpretation of the repeated-measurements ANOVA model becomes difficult. For example, when Keller et al. (13, their Fig. 1) find a significant treatment effect (difference between a "low glycogen trail" and "control trail") on the plasma concentration of interleukin 6 during a 3-h exercise experiment, this probably makes little sense insofar as there probably exists time × treatment effects. This is, however, acknowledged by the authors, who use post hoc *t*-tests to assess pairwise differences at different time points. Nevertheless, the use of post hoc *t*-tests (e.g., Refs. 14 and 36) corresponds to reporting time × treatment effects, thus making the reporting of overall treatment effects appear rather confusing.

However this may be, compared with "mechanistic" (nonlinear) models with easy-to-interpret parameters, the interpretation of ANOVA-like models is often less biologically obvious.

Other Possible Models

The four models discussed by no means comprise an exhaustive list of possible models that can be used to analyze our pinacidil experiment. Our opinion is that models M1 and M2 fall short because of the inherently nonlinear nature of the data (and of the log-transformed data). It is beyond the scope of this study to consider in detail other types of models, but we will consider one type, random intercept polynomial regressions.

Nonlinearity can be incorporated into statistical models by other means than the M3 and M4 models. One could, for example, consider polynomial regressions of y_{tij} on time and higher orders thereof and let the corresponding regression coefficients depend on treatment group. One such possible model could be

$$y_{tij} = (\alpha_{i0} + b_{i0}) + [(\alpha_{i1} + b_{i1}) \times \text{time}^1] + [(\alpha_{i2} + b_{i2}) \times \text{time}^2] + \cdots + [(\alpha_{is} + \text{b}_{is}) \times \text{time}^s]$$

where the α-parameters are regression coefficients and the b_{jk}'s are individual, subject-specific variations therein. This would correspond to fitting an *s*-degree polynomial to each subject and then testing whether the regression coefficients (the α's) differ between treatments (6, 14, 34). This method is relatively straightforward and easy to implement in different software packages. However, the interpretation of this polynomial fit is not so straightforward. Let us, for the purpose of argument, assume that $s = 3$, so we are fitting a third-degree polynomial function to the data. Let us furthermore assume that the regression coefficients for time^2 differ between the two treatments. If $\alpha_{i2} > \alpha_{i'2}$, we *cannot* conclude that the onset of fatigue is faster, say, in the *i*th treatment compared with the *i*′th treatment. The reason is that *shapes* of the predicted curves depend on the "interaction" of the three regression coefficients (α_{i1}, α_{i2}, and α_{i3}).

Software

All of the discussed models can be implemented in SAS and S-PLUS/R. In the APPENDIX, we list the SAS code used to analyze models M1-M4. Moreover, models M1 and M3 can be implemented in GUI software like SigmaStat.

Recommendations

Given that the repeated-measurements ANOVA model is widely used, the question as to the circumstances under which the model is appropriate arises. Assuming that *1*) one is interested in properties of the data that may adequately be coined by, e.g., treatment or time point averages and *2*) variances are reasonably homogeneous among different time points and treatments (obtained by, e.g., transforming the data), the repeated-measurements ANOVA model may very well be appropriate if the number of repeated measurements is "small" (<5, say) and the distance between adjacent time points "large" and equidistant (6, 14, 34, 36).

Several factors make the use of nonlinear (mixed-effects) models, like the two presented in this paper, attractive alternatives to simpler ANOVA-like methods. First, occasionally the functional form of the model can be predicted on the basis of knowledge of the mechanisms of study systems. One primary advantage thereof is that model parameters can have stringent interpretations enabling much clearer conclusions. Second, the flexibility of nonlinear models makes them almost tailor-made to fit repeated-measurement data from physiological experiments. Although a model not fitting the data is synonymous with the model being wrong, one cannot conclude otherwise that a model that does fit the data is correct in a mechanistic sense. Yet often the interpretation of model parameters suffices. Third, treating time as a continuous variable rather than as a factor, such as in the repeated-measurements ANOVA model, seems much more appropriate. Fourth and finally, the fact that nonlinear models are relatively easy to implement in several widely distributed software packages makes the nonlinear models an attractive alternative to the ANOVA-like analyses.

APPENDIX: SAS CODE

The Data

The data of the four analyses must be on the following format:

```
Treatment  Time  Animal  Animal_ID     Y       logY
Placebo     0      I         5      100.000  4.60517
Placebo     1      I         5       79.798  4.37950
Placebo     2      I         5       68.182  4.22218
Placebo     3      I         5       59.596  4.08759
Placebo      4    I      5     53.535  3.98034
Placebo      5    I      5     49.495  3.90187
Placebo      6    I      5     45.455  3.81671
Placebo      7    I      5     41.414  3.72362
Placebo      8    I      5     36.364  3.59357
Placebo      9    I      5     32.323  3.47579
Placebo     10    I      5     29.798  3.39444
Placebo     11    I      5     26.768  3.28719
Placebo     12    I      5     24.242  3.18810
Placebo     13    I      5     22.222  3.10109
Placebo     14    I      5     20.202  3.00578
Pinacidil    0    I     17    100.000  4.60517
Pinacidil    1    I     17     42.735  3.75502
Pinacidil    2    I     17     29.915  3.39834
Pinacidil    3    I     17     19.658  2.97849
Pinacidil    4    I     17     12.821  2.55105
Pinacidil    5    I     17     10.684  2.36872
Pinacidil    6    I     17      9.402  2.24089
Pinacidil    7    I     17      8.547  2.14558
Pinacidil    8    I     17      8.120  2.09429
Pinacidil    9    I     17      7.692  2.04022
Pinacidil   10    I     17      7.265  1.98306
Pinacidil   11    I     17      7.692  2.04022
Pinacidil   12    I     17      7.692  2.04022
Pinacidil   13    I     17      7.692  2.04022
Pinacidil   14    I     17      7.692  2.04022
```

and then repeated six times for the six remaining animals.

Repeated-Measurements ANOVA

```
1: proc mixed data=Pinacidil method=MIVQUE0;
2: class Animal Treatment Time;
3: model Y = Treatment Time Treatment*Time;
4: random Animal Animal*Treatment;
5: repeated/Type=CS Subject=Animal*Treatment;
6: estimate ‘Placebo − Pinacidil’ Treatment
   −1.0 +1.0;
7: run; quit;
```

1. `proc mixed` invokes SAS's mixed-model procedure. The `data=Pinacidil` tells the program to use the pinacidil data set. The `method=MIVQUE0` makes sure that `proc mixed` estimates the model using ANOVA methods.

2. The `class` statement tells `proc mixed` that the variables `Animal Treatment Time` are to be considered as categorial variables.

3. The model statement tells SAS to analyze the `Y` variable with `Treatment Time Treatment*Time` as fixed effects.

4. The `random` statement tells SAS that `Animal Animal*Treatment` are to be considered as random effects.

5. The `repeated` statement specifies the constant correlation structure among observations (`Type=CS`) from the same muscle (`Subject=Animal*Treatment`).

6. The `estimate` statement calculates the mean difference between the two treatment groups.

7. The `run; quit;` terminates the procedure.

Linear Mixed-Effects Model on log(y_{tij})

```
 1: data subset;
 2: set Pinacidil;
 3: if Time in (0,2,4,6,8,10,12,14) then delete;
 4: proc mixed data=subset method=ML;
 5: class Animal Treatment Time;
 6: model logY = Treatment Time Treatment*Time;
 7: random Animal;
 8: repeated/Type=AR(1) Subject=Animal*Treatment;
 9: estimate ‘Placebo − Pinacidil’ Treatment
    −1.0 +1.0;
10: run; quit;
```

1–3. The `data` step defines a (new) data set (named `subset`) in which observations at even time points are deleted.

4–9. There are two important differences here compared with the previous analysis. First, the `method=ML` defines that `proc mixed` uses the maximum likelihood estimation method to estimate the model. Second, the `Type=AR(1)` defines the first-order autoregressive correlation structure. Naturally, the model `logY = ...` corresponds to analyzing log(y_{tij}) rather than y_{tij}.

Paired t-Test or Two-Way ANOVA of α-Parameter Estimates

```
 1: proc nlin data=Pinacidil method=marquardt;
 2: by Animal Treatment;
 3: parms alpha1=80 alpha2=−0.25 alpha3=15;
 4: model Y = alpha1*exp(alpha2*Time) + alpha3;
 5: output out=Parameter PARMS=alpha1 alpha2
    alpha3;
 6: data Parameter;
 7: set Parameter;
 8: if time=0;
 9: /*Animal Treatment  alpha1   alpha2   alpha3
10: I        Placebo    79.8898 −0.17978 15.7739
11: I        Pinacidil  90.1777 −0.80868  8.3330
12: II       Placebo    77.1571 −0.16329 18.4717
13: II       Pinacidil  75.4576 −0.34351 25.2337
14: III      Placebo    76.8947 −0.26500 16.4085
15: III      Pinacidil  81.4740 −0.44850 16.3985
16: IV       Placebo    81.7068 −0.39405 14.3307
17: IV       Pinacidil  86.3574 −0.83024 12.8309
18: V        Placebo    80.7214 −0.15237 14.5055
19: V        Pinacidil  92.8730 −0.22106  9.5056
20: VI       Placebo    84.0525 −0.22540 13.7762
21: VI       Pinacidil  81.4218 −0.15598 16.4902
22: VII      Placebo    92.8730 −0.13062 11.0187
23: VII      Pinacidil  86.3162 −0.19384 13.0602*/
24:
25: proc mixed data=Parameter method=ML;
26: class Animal Treatment;
27: model alpha2 = Treatment;
28: random Animal;
29: estimate ‘Placebo − Pinacidil’ Treatment
    −1.0 +1.0;
30: run; quit;
```

1. `proc nlin data=Pinacidil method=marquardt` invokes the `nlin` procedure on the pinacidil data set using the marquardt iterative method to estimate parameters in the model.

2. The `by Animal Treatment` ensures that the nonlinear model (*Eq. 3*) is fitted for each combination of animal and treatment (musclewise).

3. Starting values for the marquardt iterative method.

4. Corresponds to *Eq. 3*.

5. Defines an output data set named `Parameter` containing (among other things) the three α-parameters named `alpha1, alpha2, and alpha3.`

6–8. Reshapes the `Parameter` data set.

9–23. The `Parameter` data set.

25–30. Performs a two-way ANOVA on the `Parameter` data set using maximum likelihood to estimate the model. The results of this analysis are not reported in the text and are provided here simply as an example.

Nonlinear Mixed-Effects Model

In the SAS code below the variable `Treat` equals 1 when the treatment group is `Pinacidil` and 0 otherwise (treatment group is `Placebo`).

```
1: proc nlmixed data=Pinacidil;
2: parms mu1=82 mu2=-0.28 mu3=17 a1=-4 a2=-0.08
   a3=3 s_u2=0.05 s_u3=21 corr2_3=0 s=5;
```

```
3: pred = (mu1 + a1*Treat)*exp((mu2 +
   a2*Treat + u2)*Time)+mu3+a3*Treat+u3;
4: model Y ~ normal(pred,s);
5: random u2 u3 ~ normal([0,0,[s_u2,corr2_3,s_u3)
   subject=Animal_ID;
6: PREDICT pred out=pred;
7: run; quit;
```

1. Invokes the `nlmixed` procedure on the `Pinacidil` data set.
2. Defines starting values for the parameters to the iterative method used to estimate the model.
3. Corresponds to *Eq. 6*. Note that the treatment effects on the three α-parameters has been defined via dummy variables. Thus the three α-parameters for the placebo group equals `mu1`, `mu2`, and `mu3`, respectively, whereas the three α-parameters in the pinacidil group equals `mu1+a1`, `mu2+a2`, and `mu3+a3`, respectively.
4. Defines that Y = y_{tij} is normally distributed with mean pred and variance s (see `parms` statement).
5. Defines the distribution of the two random effects `u1` and `u2` (two-dimensional normal). The `subject=Animal_ID` defines that the hierarchial level of the two random effects is `Animal_ID`, which is a variable taking different numerical values for each individual muscle.
6. Defines an output data set (`pred`) to include (among other things) the predicted values from the model.

ACKNOWLEDGMENTS

Three anonymous reviewers provided helpful comments on earlier drafts of the manuscript.

REFERENCES

1. **Benos DJ.** Ethics, revisited. *Adv Physiol Educ* 25: 189–190, 2001.
2. **Benos DJ, Kirk KL, and Hall JE.** How to review a paper. *Adv Physiol Educ* 27: 47–52, 2003.
3. **Billat VL, Richard R, Binsse VM, Koralsztein JP, and Haouzi P.** The $\dot{V}O_2$ slow component for severe exercise depends on type of exercise and is not correlated with time to fatigue. *J Appl Physiol* 85: 2118–2124, 1998.
4. **Brooks EM, Morgan AL, Pierzga JM, Wladkowski SL, O'Gorman JT, Derr JA, and Kenney WL.** Chronic hormone replacement therapy alters thermoregulatory and vasomotor function in postmenopausal women. *J Appl Physiol* 83: 477–484, 1997.
5. **Davison C and Hinkley DV.** *Bootstrap Methods and Their Application*. Cambridge, UK: Cambridge Univ. Press, 1997.
6. **Diggle P, Heagerty P, Liang K-Y, and Zeger S.** *Analysis of Longitudinal Data*. Oxford, UK: Oxford Univ. Press, 2002.
7. **Gong B, Miki T, Seino S, and Renaud J-M.** A K_{ATP} channel deficiency affects resting tension, not contractile force, during fatigue in skeletal muscle. *Am J Physiol Cell Physiol* 279: C1351–C1358, 2000.
8. **González E and Delbono O.** Age-dependent fatigue in single intact fast and slow fibers from mouse EDL and soleus skeletal muscles. *Mech Aging Devel* 122: 1019–1032, 2001.
9. **Heunks LM, Bast A, van Herwaarden CL, Haenen GR, and Dekhuijzen PN.** Effects of emphysema and training on glutathione oxidation in the hamster diaphragm. *J Appl Physiol* 88: 2054–2061, 2000.
10. **Hirschfield W, Moody MR, O'Brien WE, Gregg AR, Bryan RM Jr, and Reid MB.** Nitric oxide release and contractile properties of skeletal muscles from mice deficient in type III NOS. *Am J Physiol Regul Integr Comp Physiol* 278: R95–R100, 2000.
11. **Horton TJ, Miller EK, Glueck D, and Tench K.** No effect of menstrual cycle phase on glucose kinetics and fuel oxidation during moderate-intensity exercise. *Am J Physiol Endocrinol Metab* 282: E752–E762, 2002.
12. **Juel C, Pilegaard H, Nielsen JJ, and Bangsbo J.** Interstitial K^+ in human skeletal muscle during and after dynamic graded exercise determined by microdialysis. *Am J Physiol Regul Integr Comp Physiol* 278: R400–R406, 2000.
13. **Keller C, Steensberg A, Pilegaard H, Osada T, Saltin B, Pedersen BK, and Neufer PD.** Transcriptional activation of the IL-6 gene in human contracting skeletal muscle: influence of muscle glycogen content. *FASEB J* 15: 2748–2750, 2001.
14. **Littell RC, Miliken Stroup WW GA, and Wolfinger RD.** *SAS System For Mixed Models. Cary*, NC: SAS Institute, 1996.
15. **Lunde PK, Verburg E, Eriksen M, and Sejersted OM.** Contractile properties of in situperfused skeletal muscles from rats with congestive heart failure. *J Physiol* 540: 571–580, 2002.
16. **Matar W, Lunde JA, Jasmin BJ, and Renaud J-M.** Denervation enhances the physiological effects of the K_{ATP} channel during fatigue in EDL and soleus muscle. *Am J Physiol Regul Integr Comp Physiol* 281: R56–R65, 2001.
17. **Matar W, Nosek TM, Wong D, and Renaud J-M.** Pinacidil suppresses contractility and preserves energy but glibenclamide has no effect during muscle fatigue. *Am J Physiol Cell Physiol* 278: C404–C416, 2000.
18. **McGuire M, Cantillon D, and Bradford A.** Effects of almitrine on diaphragm contractile properties in young and old rats. *Respiration* 69: 75–80, 2002.
19. **Nagaraj RY, Nosek CM, Brotto MAP, Nishi M, Takeshima H, Nosek TM, and Ma J.** Increased susceptibility to fatigue of slow- and fast-twitch muscles from mice lacking the MG29 gene. *Physiol Genomics* 4: 43–49, 2000.
20. **Nethery D, DiMarco A, Stofan D, and Supinski G.** Sepsis increases contraction-related generation of reactive oxygen species in the diaphragm. *J Appl Physiol* 87: 1279–1286, 1998.
21. **Nevill AM, Holder RL, Baxter-Jones A, Round JM, and Jones DA.** Modeling developmental changes in strength and aerobic power in children. *J Appl Physiol* 84: 963–970, 1998.
22. **Olsen H, Vernersson E, and Lanne T.** Cardiovascular response to acute hypovolemia in relation to age. Implications for orthostasis and hemorrhage. *Am J Physiol Heart Circ Physiol* 278: H222–H232, 2000.
23. **Pilegaard H, Keller C, Steensberg A, Helge JW, Pedersen BK, Saltin B, and Neufer PD.** Influence of pre-exercise muscle glycogen content on exercise-induced transcriptional regulation of metabolic genes. *J Physiol* 541: 261–271, 2002.
24. **Pilegaard H, Ordway GA, Saltin B, and Neufer D.** Transcriptional regulation of gene expression in human skeletal muscle during recovery from exercise. *Am J Physiol Endocrinol Metab* 279: E806–E814, 2000.
25. **Pinheiró JC and Bates DM.** *Mixed-Effects Models in S and SPLUS*. Heidelberg: Springer Verlag, 2000.
26. **Plant DR, Gregorevic P, Williams DA, and Lynch GS.** Redox modulation of maximum force production of fast- and slow-twitch skeletal muscles of rats and mice. *J Appl Physiol* 90: 832–838, 2001.
27. **Russ DW, Elliott MA, Vandenborne K, Walter GA, and Binder-Macleod SA.** Metabolic cost of isometric force generation and maintenance in human skeletal muscle. *Am J Physiol Endocrinol Metab* 282: E448–E457, 2002.
28. **Sokal RR and Rohlf FJ.** *Biometry* (3th ed.). New York: WH Freeman, 1996.
29. **Sprikkelman AB, Van Eykern LA, Lourens MS, Heymans HS, and Van Aalderen WM.** Respiratory muscle activity in the assessment of bronchial responsiveness in asthmatic children. *J Appl Physiol* 84: 897–901, 1998.
30. **Steensberg A, van Hall G, Keller C, Osada T, Schjerling P, Pedersen BK, Saltin B, and Febbraio MA.** Muscle glycogen content and glucose uptake during exercise in humans: influence of prior exercise and dietary manipulation. *J Physiol* 541: 273–281, 2002.
31. **Supinski GS, Nethery D, Stofan D, Hirschfield W, and DiMarco A.** Diaphragmatic lipid peroxidation in chronically loaded rats. *J Appl Physiol* 86: 651–658, 1999.
32. **Supinski GS, Stofan D, Ciufo R, and DiMarco A.** *N*-acetylcysteine administration alters the response to inspiratory loading in oxygen-supplemented rats. *J Appl Physiol* 82: 1119–1125, 1997.
33. **Torelli GF, Meguid MM, Moldawer LL, Edwards CK III, Kim HJ, Carter JL, Laviano A, and Rossi Fanelli F.** Use of recombinant human soluble TNF receptor in anorectic tumor-bearing rats. *Am J Physiol Regul Integr Comp Physiol* 277: R850–R855, 1999.
34. **Verbeke G and Molenberghs G.** *Linear Mixed Models for Longitudinal Data*. Heidelberg: Springer Verlag, 2000.
35. **Verbrug E, Schiøtz Thorud HM, Eriksen M, Vøllestad NK, and Sejersted OM.** Muscle contractile properties during intermittent nontetanic stimulation in rat skeletal muscle. *Am J Physiol Regul Integr Comp Physiol* 281: R1952–R1995, 2001.
36. **Yandell BS.** *Practical Data Analysis for Designed Experiments*. New York: Chapman and Hall, 1997.
37. **Yensen C, Matar W, and Renaud J-M.** K^+-induced twitch potential is not due to longer action potential. *Am J Physiol Cell Physiol* 283: C169–C177, 2002.