American Journal of Primatology

RESEARCH ARTICLE

Who Is Who Matters—The Effects of Pseudoreplication in Stable Isotope Analyses

ROGER MUNDRY* AND VICKY M. OELZE
Max Planck Institute for evolutionary Anthropology, Leipzig, Germany

Stable isotope analysis in free-ranging primates is a promising new avenue in reconstructing feeding niches and temporal dietary variation. Particularly, the large sample sizes obtained from non-invasively collected hair and fecal samples from nests of great apes offer great potential. However, analyzing repeated observations of the same individuals without controlling for potential differences among them means to "pseudoreplicate" and can lead to a greatly inflated probability of erroneous significance. We here test the effects of pseudoreplication in stable isotope data of great ape hair by means of simulations. We show that pseudoreplication can severely affect the probability of erroneous significance as well as non-significance. We suggest several strategies to avoid pseudoreplication in primate isotope ecology. First, if applicable, information on individual identity should be included in statistical analyses. Second, if samples derive from unhabituated animals, sampling at far apart locations or territories should avoid resampling of the same animal. In great apes, sampling of independent nests within nest groups can ensure that each sample derives from a different individual. Third, we encourage the combination of genetic surveys with sampling for isotope analyses to ensure the (genetic) identification of individuals. Am. J. Primatol.

Key words: type I error rate; type II error rate; primate; hair; mixed models

INTRODUCTION

The investigation of stable isotope ratios in primate ecology has gained increased popularity in recent years [Blumenthal et al., 2012; Crowley et al., 2014; Fahy et al., 2013; Loudon et al., 2014; Oelze et al., 2011, 2014]. The first studies in this field have used measurements in bulk hair samples from unhabituated populations to make broad estimates of feeding behavior, including the consumption of C_4-plants in savanna chimpanzees and the feeding niche of arboreal primates [Schoeninger et al., 1997, 1999; Sponheimer et al., 2006]. More recently, the application to habituated primate populations could relate the isotopic response measured in primate tissue samples (hair or dung) to observational data on focal communities to infer about dietary sex differences, the frequencies of meat consumption and inter-annual dietary variation. So far, these studies have been limited to habituated groups of great apes and to mouse lemurs [Blumenthal et al., 2012; Crowley et al., 2014; Fahy et al., 2013; Oelze et al., 2011].

Only very recently stable isotope analysis has been extended to semi- and unhabituated great apes to reconstruct feeding niche and temporal variation in feeding behavior. Oelze et al. [2014] investigated the hair isotope ratios in sympatric gorillas and chimpanzees from Gabon and found isotopic evidence for feeding niche separation that was varying between seasons. Such investigations in unhabituated and, thus, elusive primates are very appealing because they have the potential to infer, for instance, feeding niches, niche partitioning, hunting and meat consumption, insectivory, dietary sex differentiation and the relevance of specific plant foods such as C_4-plants, legumes, terrestrial herbaceous vegetation or gum [Blumenthal et al., 2012; Crowley et al., 2014; Fahy et al., 2013; Oelze et al., 2014; Schoeninger et al., 1999], as well as human-primate conflicts such as crop raiding [Loudon et al., 2014]. Particularly hair is a convenient sample matrix, as it is not prone to degradation or contamination but records and retains an isotopic signature related to the diet over long periods of time [Cerling et al., 2009; Oelze et al., 2011; Schwertl et al., 2005]. For example

*Correspondence to: Roger Mundry, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: roger_mundry@eva.mpg.de

Received 31 January 2015; revised 7 October 2015; revision accepted 26 October 2015

DOI: 10.1002/ajp.22499
Published online XX Month Year in Wiley Online Library (wileyonlinelibrary.com).

in apes, one hair sample can shed light on the feeding behavior of a single individual over a period of 6–10 months [Oelze et al., 2014]. At the same time, sampling of hair can be conducted strictly non-invasively or minimal invasively [Oelze, This volume]. Particularly in great apes, hair samples can be obtained in sufficient amounts from sleeping nests in the absence of the study subjects. As great apes commonly build a new nest on a daily basis and nest-reuse is rare [Fruth & Hohmann, 1996], each hair sample collected in a nest represents a single individual and its recent isotopic profile. However, for hair (or dung) samples taken from free-ranging, unhabituated individuals the identity of the sampled individual will be unknown. As a consequence, a corresponding study might suffer from "pseudoreplication" [Hurlbert, 1984].

Pseudoreplication refers to treating data points in a statistical analysis as if they were independent although, in fact, they are not [Hurlbert, 1984]. In this study, we focus on scenarios with repeated observations of the same individuals and treating them in the statistical analysis as if they were independent (i.e., from different individuals). Most studies dealing with the consequences of pseudoreplication addressed the case of having repeated observations of the same individuals and ignoring them in an analysis of the effects of a predictor which varies between individuals (e.g., species or sex), and these publications found that pseudoreplication can lead to a greatly inflated type I error rate (i.e., increased probability of erroneous significance [e.g., Hurlbert, 1984; Machlis et al., 1985; Mundry & Sommer, 2007]). Pseudoreplication is now widely recognized as an issue in research fields as diverse as ecology, medical research and linguistics (to mention just a few). On the other hand, still today some ignorance toward pseudoreplication issues can be found in the published literature, and it is still quite common in certain research fields (see Waller et al. [2013] for the example of animal communication research).

Basically, there are two ways of accounting for pseudoreplication. The first is to average the measures per individual (or whatever the source of non-independence is), an option only applicable when the key predictor varies between individuals and when the sample size with regard to the number of individuals is sufficiently large. The other approach is to control for subject identity in the statistical analysis. This is only possible in habituated communities of primates, for which individuals can be identified during direct observations or nest to nest follows [see Oelze, This volume]. Classical statistical tools controlling for individual identity are procedures for "paired" or "related" data (e.g., Wilcoxon matched pairs-, Friedman-, McNemar-, Cochran's Q- and paired samples *t*-test, as well as repeated measures ANOVA [e.g., Quinn & Keough, 2002; Siegel & Castellan, 1988; Zar, 1999]). Recently, Mixed Models [e.g., Baayen, 2008; Bolker et al., 2008; Gelman & Hill, 2007] have rapidly gained popularity in primate ecology research and other fields. Mixed Models greatly outperform the aforementioned classical tools because of their far superior flexibility with regard to the data designs that can be analyzed and the assumptions about the data (or precisely, "residuals," i.e., deviations of the observations from the model).

In stable isotope studies of wild animal populations, pseudoreplication is probably a common risk because researchers might frequently not know from which particular individuals the samples they collect originate. Also, due to the logistical constraints of working in remote regions difficult to access, sampling will usually take place in a small area, increasing the risk of sampling from the same individuals repeatedly. In this paper we treat the consequences of pseudoreplication with a particular focus on stable isotope analyses in free-ranging primates, particularly great apes. More specifically, we deal with the effects of pseudoreplication on the results for predictors which vary between individuals (e.g., species, sex, or age) as well as on the level of predictors which vary within individuals (e.g., temporal variation in food availability or, more generally, "seasonality"). We address these issues by simulating pseudoreplicated data and then analyzing them while ignoring pseudoreplication as well as using methods accounting for the non-independence of the data and then comparing the results. Since Mixed Models are an essential tool for analyzing pseudoreplicated data we also give a very brief introduction to Mixed Models. We conclude with recommendations for future study designs.

METHODS

Since no live animals served as subjects in this study (i.e., all simulated data sets were based on already published stable isotope data [Fahy et al., 2013; Oelze et al., 2011]). The current study did not require review by an institutional animal care and use committee (IACUC) or its equivalent. The research protocols of the studies from which the data originated met the legal requirements of the countries in which the studies were conducted, and adhered to the American Society of Primatologists Principles for the Ethical Treatment of Nonhuman Primates.

We used simulations to assess the impact of pseudoreplication on type I and type II error rate. Making a type I error means to incorrectly reject a true null-hypothesis, whereas making a type II error means failing to detect an effect although it is present. The advantage of using simulations is that one knows exactly which effects do and do not exist, and, hence, can evaluate a model's reliability by comparing its results with what one knows about the data. For instance, one can simulate data with no

differences between sexes and then determine the probability for a model to reveal significant sex differences despite their absence, and how this probability is affected by pseudoreplication.

All data were generated in R (version 3.x; R Core Team, 2014; http://www.r-project.org/). Throughout, we simulated data consisting of "hair section samples" (hereafter "hairs") sampled from "individuals". We systematically varied the number of hairs sections sampled per individual, the amount of variability in the response (stable carbon and nitrogen isotope ratios) that is due to differences between individuals and between hair sections sampled from the same individual, and the residual variance (i.e., random fluctuations and measurement error in isotope ratios). We analyzed the simulated data using pseudoreplicated analyses (ignoring individual identity) and also conducted analyses appropriately accounting for individual identity and compared the results. We focused our assessments of model reliability on a between subjects factor ("sex") and a within-subjects factor ("season"). In simulation 1 we generated data with no impacts of sex or season in order to assess how type I error rates for a between- and a within-subjects predictor are influenced by pseudoreplication. In simulation 2 we focused on the impact of pseudoreplication on the type I error rate obtained for the within-subjects effect of season, this time simulating temporal autocorrelation of isotope ratios within hairs (simulation 2a) and within individuals (simulation 2b) but no overall effect of season. In simulation 3, we focused on type II error rate, this time simulating effects of season that act on different individuals and hairs in the same way. In the final simulation we assessed how our results could be affected by imbalanced sampling (i.e., varying numbers of hairs sampled per individual).

Simulated data were based on published stable isotope data from habituated (and thus identified to individual level) great apes (see ethics statement above). These included data from free-ranging bonobos from Salonga National Park [Oelze et al., 2011] and western chimpanzees from Ivory Coast [Fahy et al., 2013]. In these datasets dietary differences between the sexes were investigated and temporal variation in the isotope signatures was described. From these datasets we extracted the following data settings: We set the average isotope ratios of individuals (after controlling for differences between hair sections) to be normally distributed with a standard deviation of sd_{ind}, the average isotope ratios of hairs to be normally distributed (after controlling for differences between individuals) with a standard deviation of sd_{hair}, and the residuals to be normally distributed (after controlling for differences between individuals and hairs) with a standard deviation of sd_{error}. In the most simple version of the simulation the response was then generated by adding three random numbers drawn from three normal distributions, that is,

$$\begin{aligned}\text{response} &= \text{random normal}(sd_{ind}) \\ &+ \text{random normal}(sd_{hair}) \\ &+ \text{random normal}(sd_{error}) \end{aligned} \quad (1)$$

A response generated according to Eq. (1) does not vary between sexes nor does it show any seasonal variation and hence no effects of these two predictor variables should be detected by a corresponding statistical model. More precisely, across a number of simulated data sets, the proportion of models revealing significance for, e.g., sex should roughly equal 0.05, the expected type I error rate. If pseudoreplication were no issue, the type I error rate should be unaffected by whether the analysis accounts for pseudoreplication or not, but if pseudoreplication is an issue, an analysis not accounting for it should reveal an elevated type I (or type II) error rate.

Simulation 1: Type I Error Rates

In the first simulation we aimed to address the impact of pseudoreplication at the level of the individual on type I error rate. We therefore simulated data sets in which individuals provided repeated hairs. We systematically varied the number of hairs per individual from 2 to 8 (increment 2) which is realistic according to previous studies on great apes. We also systematically varied sd_{ind} from 0 (i.e., no differences between individuals) to 1 (with an increment of 0.2). We also systematically varied the values of sd_{hair} and sd_{error} from 0.2 to 1 (with an increment of 0.2). We generated the response according to Eq. (1). The particular values for sd_{ind}, sd_{hair}, and sd_{error} were chosen because a mixed model testing for differences between two species (bonobos and chimpanzees) and sexes indicated them to be roughly in this order of magnitude. The number of simulated individuals was set to 50.

After the response was generated, we simulated two predictors to be tested, namely sex and season. The predictor of sex we generated by randomly allocating each individual to either sex. For simulating the predictor season we first assigned each hair a sampling date by randomly drawing numbers from a uniform distribution with a minimum of 1 and a maximum of 365. Note that this means that hairs of the same individual were sampled independently with regard to date. The individual sections of the hairs we then assumed to each correspond to 1 month of growth [Oelze, This volume], and each section was assigned a date according to its midpoint. The number of sections (i.e., months) per hair we randomly sampled from an empirical distribution encompassing values between 1 and 9 (Appendix Fig. S1). We then tested each of the simulated data sets using a model

ignoring individual identity and one that does appropriately account for it (see below for details). If pseudoreplication is no issue, we would expect the proportion of significant *P*-values obtained from these tests to be 0.05, regardless of the particular combination of simulation settings and whether the model controls for pseudoreplication or not.

Simulations 2a and b: Type I Error Rates in the Presence of Autocorrelation

The first simulation addressed type I error rates in the complete absence of any effects of the within- (season) and between-subjects (sex) predictor. However, it seems plausible that consecutive measures taken from the same hair show some trend or autocorrelation. Thus, we conducted the second simulation to assess whether such a temporal non-independence (within hairs, simulation 2a; and within individuals, simulation 2b) could affect type I error rates for the effect of season. The key aspect of simulation 2a was that we added temporal autocorrelation within hairs. Such temporal autocorrelation could, for instance, happen when individuals go through periods of physiological stress and when the period of growth is unknown for the individual hairs or when the physiology of hair growth affects hair isotope ratios. We simulated such temporal autocorrelation by adding the sine of the date corresponding to the midpoint of the respective hair section (as a circular variable). So the response generating function for this simulation was

$$\begin{aligned}\text{response} &= \text{random normal}(\text{sd}_{\text{ind}}) \\ &+ \text{random normal}(\text{sd}_{\text{hair}}) \\ &+ \text{d} \times \text{c}_{\text{ac}} \times \text{sine}(\text{section midpoint}) \\ &+ \text{random normal}(\text{sd}_{\text{error}}) \end{aligned} \quad (2a)$$

where c_{ac} is an "autocorrelation" parameter that we systematically varied from 0.2 to 1 (increment 0.2) in the simulation and d is either -1 or 1 (randomly chosen for each hair) having the effect that the resulting curve begins with an increase or decrease, respectively. It is important to note that the sine function was determined relative to the length of the hair but not related at all to an absolute periodicity being parallel across hairs. So an overall effect of seasonality was not created (since each hair was later assigned a sample date by randomly drawing an integer from the interval between 1 and 365).

While each hair showed a pattern of temporal autocorrelation in simulation 2a that was by definition independent to that of other hair samples, even from the same individual, in simulation 2b we set hair samples of the same individual to show the same annual periodicity in the sense that each individual went through its own annual periodicity. Specifically, each hair was assigned a date when it was sampled and different hairs of the same individual then followed the same periodic pattern with a period duration of 1 year. Hence, periodicities of different individuals showed the same period duration of 1 year but were not "parallel" in the sense that peaks and valleys did not fall on the same period of a year. Specifically, the response generating function for simulation 2b was

$$\begin{aligned}\text{response} &= \text{random normal}(\text{sd}_{\text{ind}}) \\ &+ \text{random normal}(\text{sd}_{\text{hair}}) \\ &+ \text{d} \times \text{c}_{\text{ac}} \times \text{sine}(\text{individual phase} \\ &+ \text{hair start day} + \text{hair section midpoint}) \\ &+ \text{random normal}(\text{sd}_{\text{error}}) \end{aligned} \quad (2b)$$

where individual phase and hair start day where both integer numbers randomly chosen from the interval from 1 to 365, and individual phase was the same for all hairs of a given individual, and hair start day was the same for all sections of a given hair. The value of d was again either -1 or 1 (randomly chosen), but this time the same for all hairs of a given individual.

Simulations 2a and b additionally differed from simulation 1, as we varied the number of individuals from 10 to 50 (increment: 10), the number of hairs per individual where 1, 2, 4, 6, or 8, and the value of c_{ac} (Equation 2a and 2b) we varied from 0.2 to 1 (increment: 0.2). We set the other three parameters (sd_{hair}, sd_{ind}, and sd_{error}) to a fixed value of 0.5 each. The models implemented were the same as in simulation 1, with the exception that here we did not include a fixed effect of sex. As in simulation 1, if pseudoreplication would have no effect we would expect the proportion of significant *P*-values obtained from these tests to be 0.05 (regardless of the particular combination of simulation settings and whether the model controls for pseudoreplication or not).

Simulation 3: Type II Error Rates When Testing Seasonality

Simulation 3 addressed whether pseudoreplication at the level of individual can affect the probability of detecting an effect of a predictor that actually has an effect and varies within subjects. The response generating function was almost identical to that used for simulation 2, with the exception that in simulation 3 we implemented a "real" seasonal effect, namely

$$\begin{aligned}\text{response} = &\text{random normal}(\text{sd}_{\text{ind}}) \\ &+ \text{random normal}(\text{sd}_{\text{hair}}) \\ &+ \text{c}_{\text{season}} \times \text{sine}(\text{hair start day} \\ &+ \text{section midpoint}) \\ &+ \text{random normal}(\text{sd}_{\text{error}}) \end{aligned} \quad (3)$$

where hair start day was drawn from a uniform random distribution with a minimum of 1 and a

maximum of 365. Note that this simulation reveals a response actually varying seasonally in a parallel fashion across individuals. The key question in this simulation was whether or to what extend the model's ability to detect such a seasonal variation depends on the relative magnitude of seasonal variation, the magnitude of differences between individuals, the number of hairs per individual and the lengths of the individual hair samples. Hence, we systematically varied the value of c_{season} from 0.2 to 1 (increment: 0.2) and the value of sd_{ind} from 0 to 1 (increment: 0.2). Furthermore, we systematically varied the number of hairs per individual from 2 to 8 (increment: 2) and the lengths of the individual hairs from 1 to 10 (unit: sections; increment: 1). The values of sd_{hair} and sd_{error} we set to fixed values of 0.5.

Other than in simulations 1 and 2 we this time expected that season would reveal significance for a considerable proportion of models, and, if pseudoreplication on the level of individual were no issue, that these proportions would not be affected by whether it is controlled for or not and also not by the parameters we systematically varied (c_{season}, sd_{ind}, number of hairs per individual, lengths of the individual hairs). The models fitted for data from this simulation were identical to those models fitted for simulation 2 (see below for details).

Simulation 4: Unbalanced Sample Sizes

Until now our simulations assumed equal numbers of hairs per individual as we focused on the effects of pseudoreplication. However, realistic data sets from free-ranging primates would most likely vary in the number of samples obtained from each individual. In order to assess to what extend such unbalanced contributions of individuals affect type I and type II error rates we repeated simulations 1 and 2b from above but with unbalanced contributions of individuals. Sample sizes per individual were determined by randomly sampling from a population of 50 individuals, but varying the total number of samples drawn from 20 to 100 (increment: 10). Throughout these simulations we set each of the values of sd_{ind}, sd_{hair}, and sd_{error} to a fixed value of 0.5 and in the simulation replicating simulation 2b also c_{ac}. Apart from that we used the same methods as described for the respective simulations above. In these simulations a few models did not converge which we excluded from the evaluation of the results.

Implementation

We simulated the data in R (version 3.x; R Core Team, 2014; http://www.r-project.org/) using the functions "runif" and "sample." In each simulation we generated 1,000 data sets per combination of values of the respective parameters varied (e.g., sd_{ind}, sd_{hair}, and sd_{error} in simulation 1).

How simulation parameters translate into stable isotope ratios can be directly inferred from the values simulated (e.g., sd_{ind}, sd_{error}, etc.). For instance, the simulated difference between sexes in simulation 1 and the simulated effects of season in simulation 1 and 2 were all zero, meaning that isotope values did not differ at all between sexes or vary seasonally. The simulated changes in isotope ratios from the low to the high season in simulation 3 equaled twice the value of c_{season} (i.e., ranged from 0.4 to 2). Obviously, these numbers can only be interpreted in relation to the magnitude of variation between individuals (sd_{ind}), hairs of the same individual (sd_{hair}) and sections of the same hair (sd_{error}) or also the magnitude of the effect of autocorrelation within hairs or individuals (sd_{ind}; simulation 2a and b, respectively). Using simulation 2b as an example in which sd_{hair}, sd_{error}, and sd_{ind} were all set to a fixed value of 0.5, a value of c_{ac} being 0.2 means that the within individuals magnitude of change in stable isotope ratios per half a year that is due to autocorrelation is roughly corresponding to the standard deviations of the means per individual, means per hair within individuals and individual sections within hair.

Statistical Analysis

After a data set was generated using the above simulations we analyzed it by running a Linear Mixed Effects Model with sex (simulation 1 and 4) and season (sine and cosine of date converted into a circular variable) as fixed effects and hair ID as a random intercept as well as random slopes [Barr et al. 2013; Schielzeth & Forstmeier, 2009] of season within hair ID. As an overall test of the combined effects of the two test predictors season and sex (simulations 1; [Forstmeier & Schielzeth, 2011; Mundry, 2014]) we compared this model with a null model comprising only the random effects (intercept and slopes) using a likelihood ratio test [Dobson, 2002]. Furthermore, we derived individual *P*-values for season and sex by comparing the full model with two respective reduced models lacking either the fixed effect of sex or the two fixed effects representing season [Barr et al., 2013]. Since such a model does not account for potential differences between individuals it represents a pseudoreplicated analysis.

If pseudoreplication had no effect we would expect the proportion of significant full-null model comparisons and full-reduced model comparisons obtained from these tests to be 0.05 in simulations 1, 2, and 4 (since in these we simulated the response disregarding any effects of sex (simulations 1 and 4) or season), regardless of the number of hairs per individual or the different values of sd_{ind}, sd_{hair}, c_{ac} (only simulation 2), and sd_{error}. To determine whether an appropriate analysis would lead to a

correct type I or type II error rate we ran a corresponding additional set of models into which we also included the random effect of subject ID as well as random slopes of season within subject. Throughout the analyses, we did not include the correlations among random slopes and intercepts in our models, in order to reduce computation time and because neglecting them does not obviously affect type I error rate [Barr et al., 2013]. Hence, in R annotation, the models implemented were

$$\begin{aligned}&\text{isotope ratio} \sim \text{sex} + \sin(\text{season}) + \cos(\text{season})\\&+ (1|\text{hair}) + (0 + \sin(\text{season})|\text{hair})\\&+ (0 + \cos(\text{season})|\text{hair})\end{aligned}$$

in case of the pseudoreplicated analyses and

$$\begin{aligned}&\text{isotope ratio} \sim \text{sex} + \sin(\text{season}) + \cos(\text{season})\\&+ (1|\text{hair}) + (0 + \sin(\text{season})|\text{hair})\\&+ (0 + \cos(\text{season})|\text{hair}) + (1|\text{individual})\\&+ (0 + \sin(\text{season})|\text{individual})\\&+ (0 + \cos(\text{season})|\text{individual})\end{aligned}$$

in case of the not pseudoreplicated analyses (note that sex was not included in models fitted in the course of simulations 2 and 3).

We analyzed the data in R (version 3.x; R Core Team, 2014; http://www.r-project.org/). Simulations were based on the functions "runif" and "sample" and data were analyzed using the function "lmer" of the package lme4 [Bates D, Maechler M, Bolker B, Walker S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7, http://CRAN.R-project.org/package=lme4]. The models were fitted with Gaussian error structure and identity link. In all models we tested for the effect of season by including the sine and cosine of Julian date (divided by 365 and then multiplied by 2 and π) as fixed effects into the model. For the analysis of the results of simulation 3 we used a Wilcoxon signed-ranks matched pairs test (Siegel & Castellan 1988). We considered two-tailed *P*-values and those smaller than or equal to 0.05 as significant.

Mixed Models

Linear Mixed Effects Models (thereafter "Mixed Model") are an extension of the general (e.g., regression, ANOVA and ANCOVA; [e.g., Aiken & West, 1991; Cohen & Cohen, 1983]) and Generalized Linear Model (e.g., logistic and Poisson regression; [e.g., McCullagh & Nelder, 1989]) that allow for the analysis of data comprising a mixture of predictors having fixed and random effects [e.g., Baayen, 2008; Bolker et al., 2008; Gelman & Hill, 2007]. A fixed effects predictor can be a factor (i.e., categorical] or covariate (i.e., quantitative) whereby covariates are always assumed to be fixed effects predictors. A fixed effects factor is one of which all of its possible levels (i.e., particular cases) are represented in the study, whereas a random effects factor is one of which only a few of its potential levels are represented in the data. Typical fixed effects factors are "species," "sex," or "experimental condition" whereas typical random effects factors are "individual" or "social group." For a fixed effects predictor a Mixed Model estimates how much the response changes when the predictor changes by one unit, and for a random effects predictor it estimates the variation in the response being due to differences between its levels ("random intercepts") and also how much the impact of fixed effects predictors on the response varies among its levels ("random slopes" [Barr et al., 2013; Schielzeth & Forstmeier, 2009]). A more thorough introduction to Mixed Models is beyond the scope of this paper, but is detailed in, for example, Bolker et al. [2008], Baayen [2008], and Gelman and Hill [2007].

RESULTS

Simulation 1: Type I Error Rates

As expected, we found an increased type I error rate for the between subjects fixed effect (i.e., "sex"), and this elevation of type I error rate increased with an increasing number of replicates (i.e., hairs) per individual and also with an increasing magnitude of differences between individuals (Fig. 1). The increase of the type I error rate was slightly weakened when the magnitude of the residual variance increased, but never to such an extent that the effect of pseudoreplication disappeared. In fact, the only scenario in which the type I error rate was at the nominal level of 5% was when there were no isotopic differences between individuals at all. Considering the full-null model comparison we found that the average type I error rate for both predictors combined was slightly smaller (average type I error rate across all parameter combinations: 0.142) as compared to the type I error rate of the reduced model only comprising the factor sex (0.188), but the overall pattern was very similar (appendix, Fig. S2). Regarding the within-subjects effect (i.e., "season") the type I error rate was 0.045 (pooled across all parameter combinations) and, hence, very close to the nominal level of 0.05. When we controlled for repeated observations of the same individual by including the respective random effect into the model (and also random slopes of season within individual), type I error

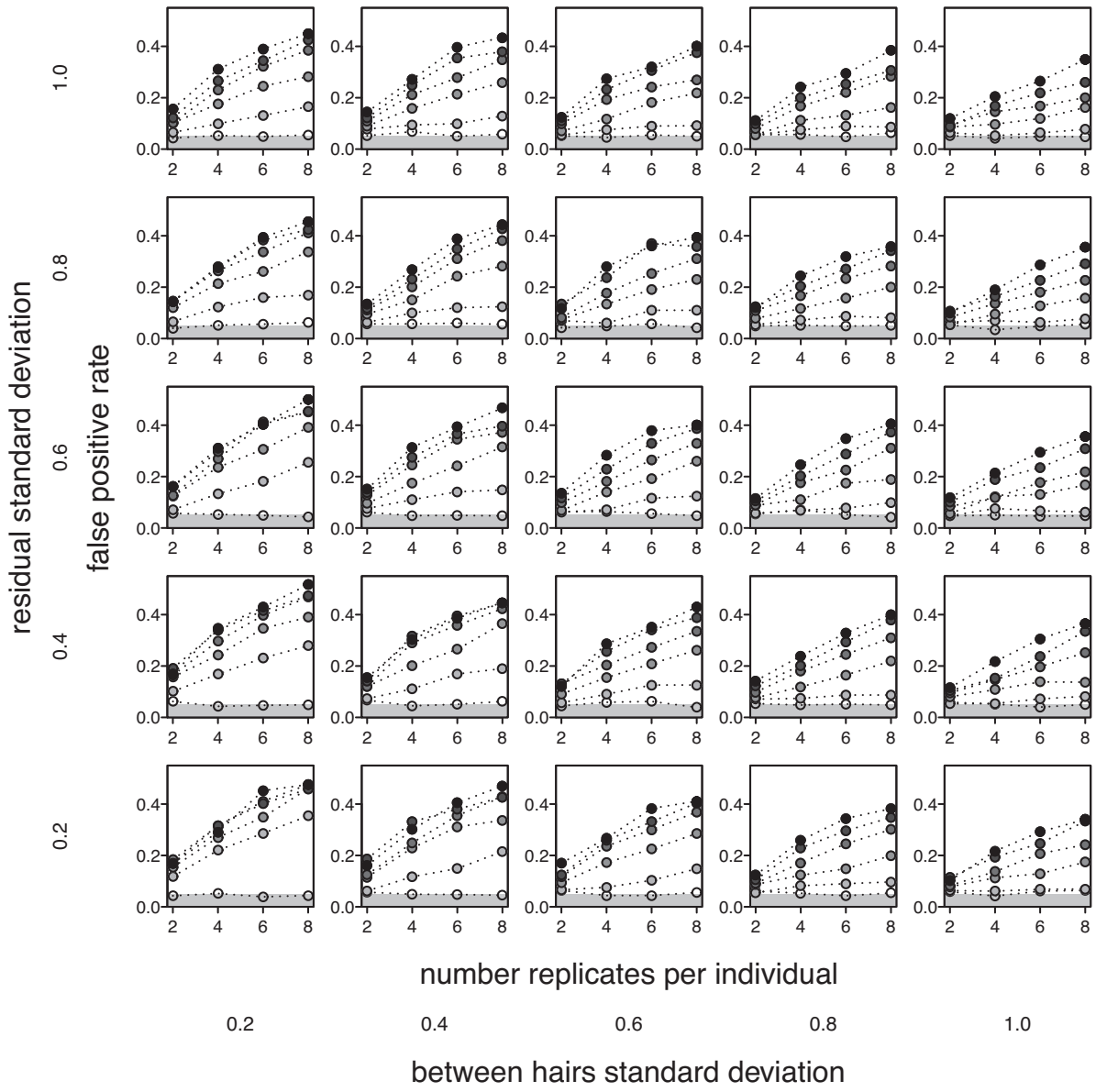

Fig. 1. Rate of erroneously significant results (y-axis within plots) for differences between sexes (simulation 1) at different rates of pseudoreplication (number of replicates per individual; x-axis within plots), different magnitudes of variation between individual hairs (between hairs standard deviation, left to right across plots), different magnitudes of residual variance (residual standard deviation, bottom to top across plots), and different magnitudes of variation between individuals (different lines and points within plots). Open points show results when there was no variation between individuals, and increasingly darker points indicate increasing magnitudes of differences between individuals (between individual standard deviation increased from 0 to 1 with an increment of 0.2). The plot is based on simulations whereby per combination of parameters 1,000 data sets were simulated and analyzed using a mixed model controlling for hair ID. Note that data were generated with no differences between sexes, and, hence, if pseudoreplication were no issue all false positive rates should fall close to or below 0.05 (top edge of gray boxes).

rates fell close to the expected 0.05 (full model: 0.044; sex: 0.054; season: 0.040; pooled across all parameter combinations).

Simulation 2: Type I Error Rates in the Presence of Autocorrelation

Autocorrelation within hair samples (simulation 2a) did not obviously affect type I error rate, and this was the case regardless of the number of hair sections per individual, the number of individuals, or the magnitude of autocorrelation (Fig. 2). However, autocorrelation within individuals (simulation 2b) strongly affected type I error rate whereby type I error rates increased with the number of hair samples per individual and the magnitude of within individual autocorrelation (Fig. 3). The number of individuals did not strongly affect type I error rates.

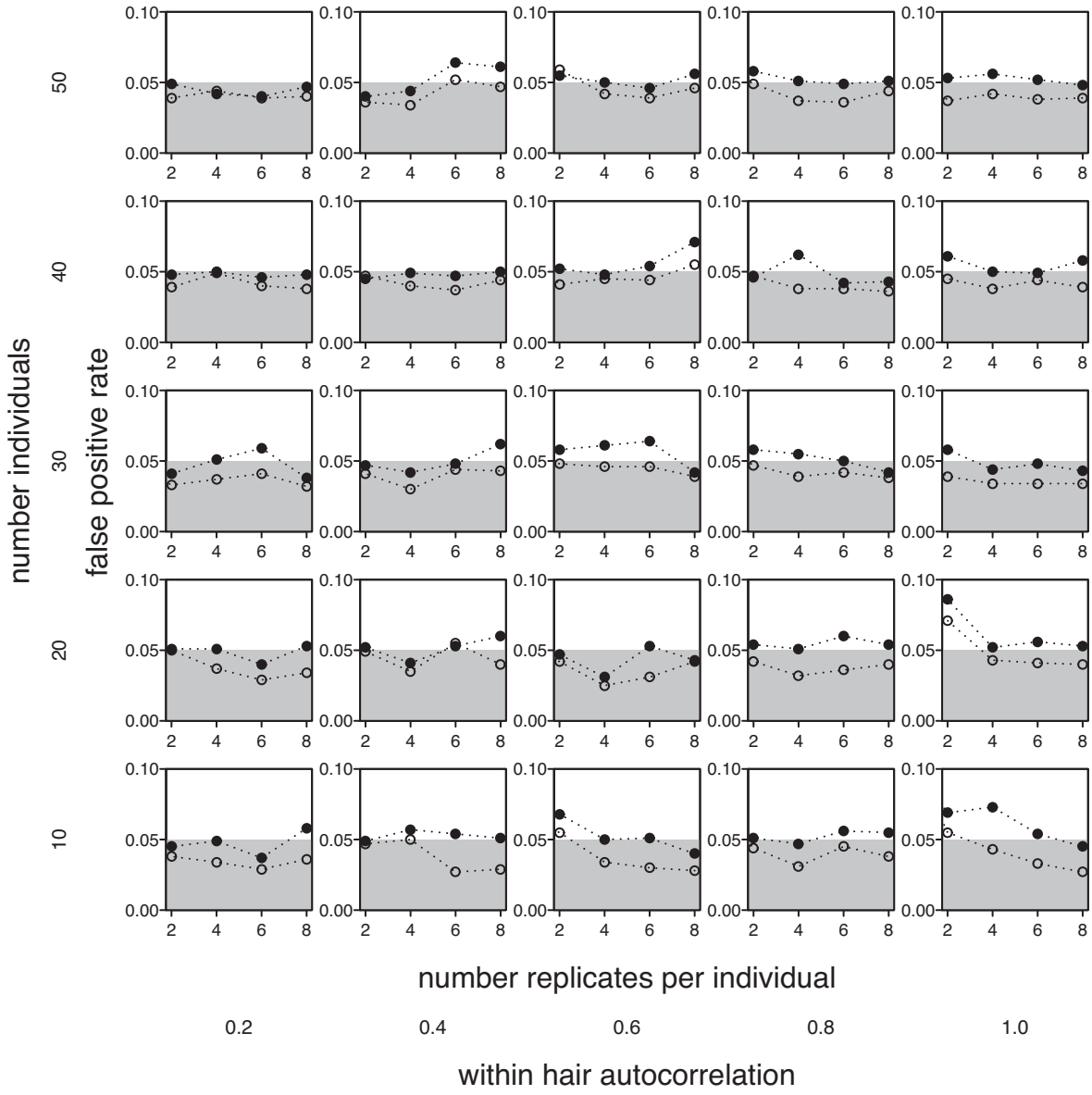

Fig. 2. Type I error rates for the effect of season when measures show temporal autocorrelation within hairs but no seasonal variation (simulation 2a). Indicated are type I error rates (y-axis within plots) for different numbers of hairs per individual (x-axis within plots), strength of within hair autocorrelation (left to right across plots), and numbers of individuals (bottom to top across plots). Filled circles show type error rates for analyses pseudoreplicating at the level of individual, and open circles show type error rates for analyses appropriately accounting for individual. Each pair of dots depicted on top of one another within the same plot is based on the same set of 1,000 simulated data sets. Note that type I error rate was invariably close to the nominal level of 0.05 (top edge of gray boxes).

Accounting for non-independence by fitting the appropriate model with random effects of individual and random slopes of season within individual lead to type I error rates close to the nominal level of 0.05 (Fig. 3).

Simulation 3: Type II Error Rates When Testing Seasonality

The majority of the simulated data sets revealed significance for the within-subjects predictor of season, obviously because we simulated strong seasonal effects. However, when considering only those 54 combinations of parameters where the proportion of significant findings differed between the pseudoreplicated and the appropriately controlled analysis, we found that the power of the non-pseudoreplicated analysis was on average larger (Wilcoxon test: $T^{+} = 1132.5$, $N = 54$, $P < 0.001$; Fig. 4). While the difference in power was usually not very large, certain simulations revealed the proportion of significant

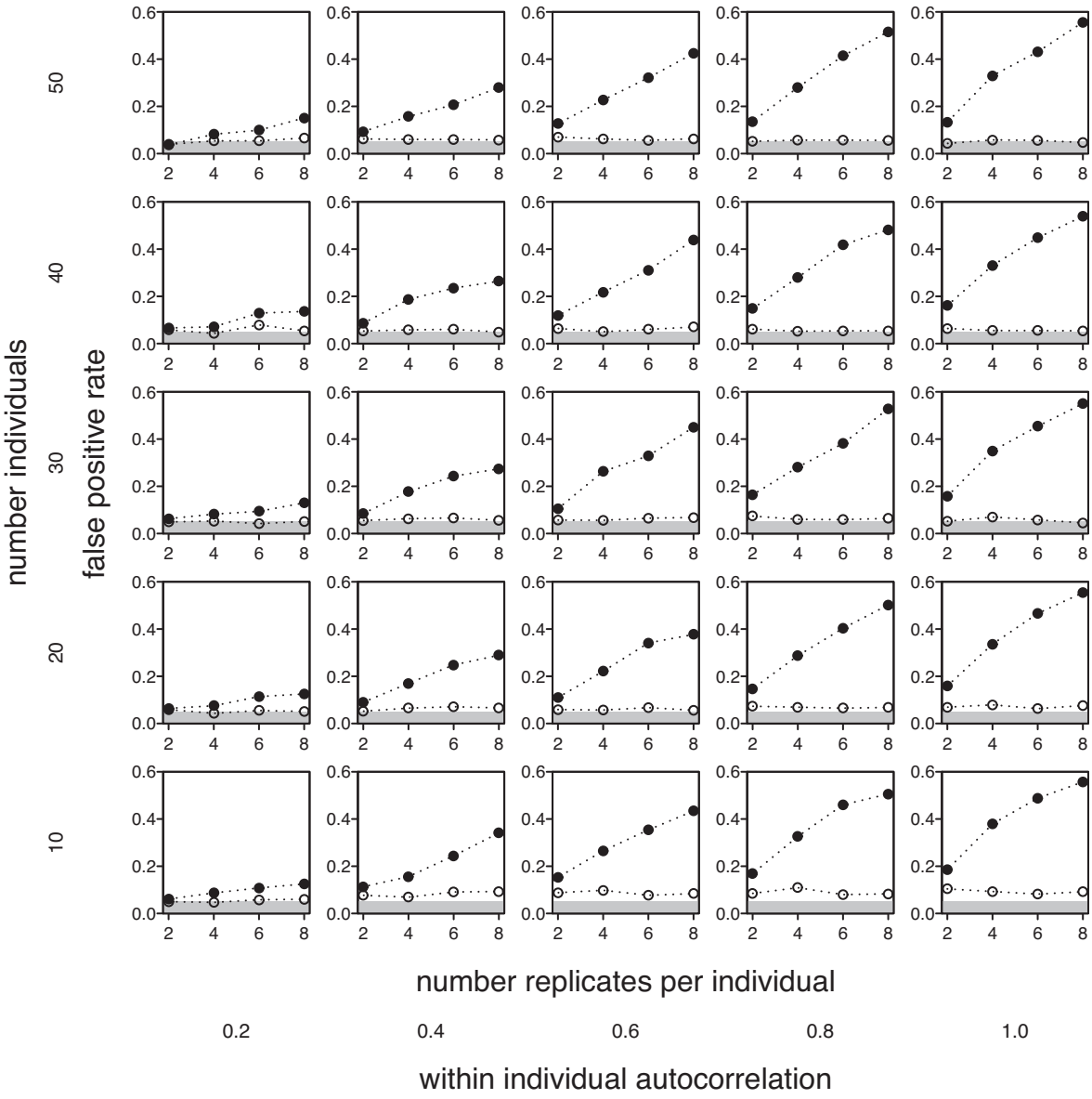

Fig. 3. Type I error rates when measures show temporal autocorrelation within individuals (simulation 2b). Indicated are type I error rates (y-axis within plots) for different numbers of hairs per individual (x-axis within plots), strength of within individual autocorrelation (left to right across plots), and numbers of individuals (bottom to top across plots). Filled circles show type error rates for analyses pseudoreplicating at the level of individual, and open circles show type error rates for analyses appropriately accounting for individual. Each pair of dots depicted on top of one another within the same plot is based on the same set of 1,000 simulated data sets. Note that for pseudoreplicated analyses type I error rates clearly increased with the number of samples per individual and the magnitude of within individual autocorrelation but were hardly affected by the number of individuals. Note also that analyses appropriately accounting for pseudoreplication invariably revealed type I error rates close to the nominal level of 0.05 (top edge of gray boxes).

results to increase by up to 0.25 when accounting for pseudoreplication.

Simulation 4: Unbalanced Sample Sizes

Repeating simulations 1 and 2b with more realistic, unbalanced, contributions of individuals to the data sets revealed elevated type I error rates, too. Error rates were higher for the between-subjects fixed effect of sex and for sex as well as for the within-subjects effect of season increased with sample size (Fig. 5). When accounting for pseudoreplication, type I error rates were close to the nominal level of 0.05 (although slightly elevated at small sample sizes). Determining the number of individuals sampled at least once and the average number of samples per individual revealed that even small sample sizes (e.g., 20 to 40 hairs of 50 individuals) resulted in pseudoreplication although at a very moderate level (appendix, Fig. S3).

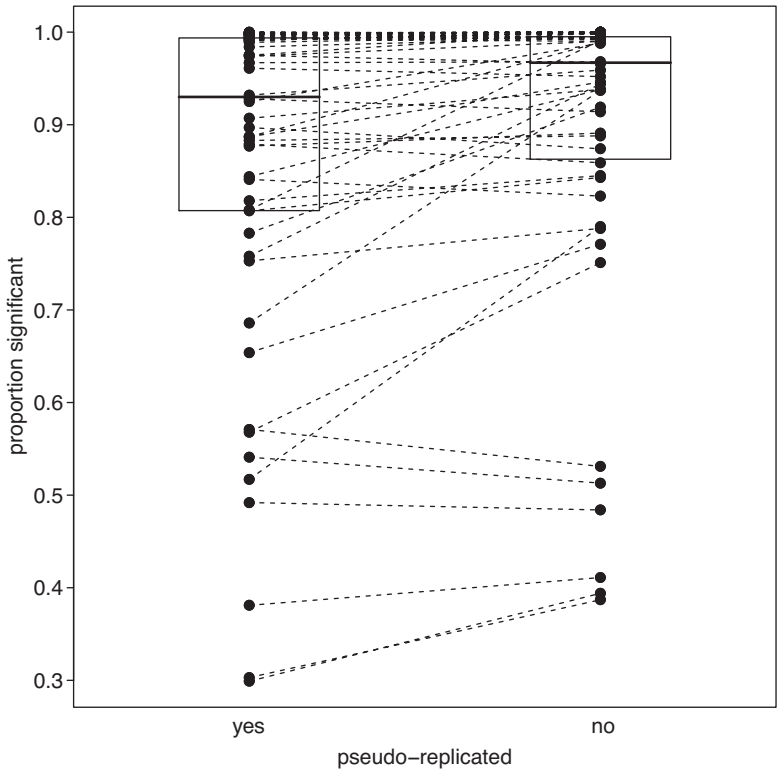

Fig. 4. Proportion significant findings for a within subjects' predictor (here, season) in a pseudoreplicated analysis (left) and when controlling for pseudoreplication (right; simulation 3). Each pair of dots connected by a dashed line shows results based on the same 1,000 data sets simulated with the same set of parameters. All simulated data sets included an effect of season and, hence, should reveal significance. Note that pseudoreplicated analyses had a decreased power (i.e., increased type II error rate).

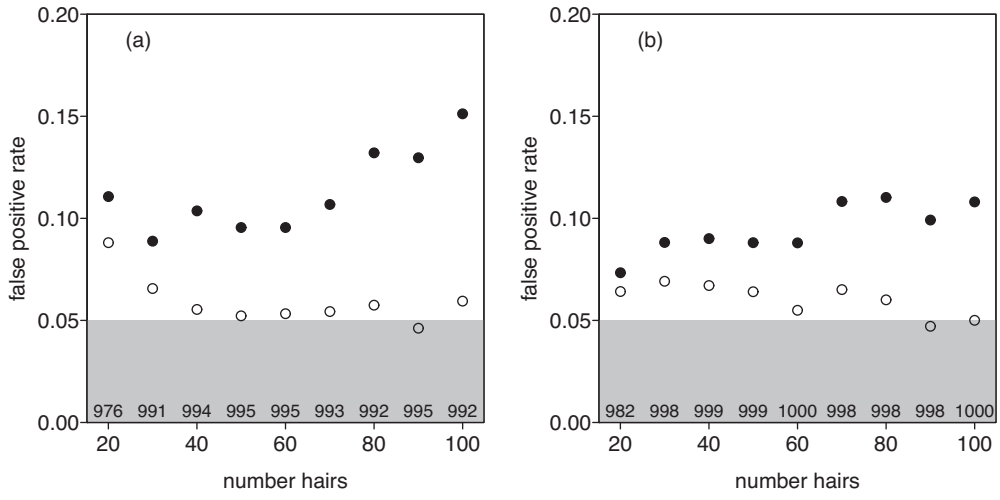

Fig. 5. Type I error rates in case of unbalanced contributions of individuals and moderate levels of pseudoreplication (simulation 4). The simulations conducted were identical to simulation 1, simulating pseudoreplication at the level of a between-subjects effect (a) and simulation 2b, simulating pseudoreplication at the level of a within-subjects effect (b), with the exception that the contribution of the different individuals was not balanced but rather determined by randomly sampling from 50 individuals a certain number of hairs (x-axis). Filled circles show type error rates for analyses pseudo-replicating at the level of individual, and open circles show type error rates for analyses appropriately accounting for individual. The numbers above the x-axis indicate the number of simulated data sets evaluated (numbers smaller than 1,000 are due to models not converging which we excluded before evaluating the results). Each pair of dots depicted on top of one another within the same plot is based on the same set of 1,000 simulated data sets. Note that type I error rates clearly increased with the number of samples per individual (filled circles) and that analyses appropriately accounting for pseudoreplication (open circles) revealed type I error rates close to the nominal level of 0.05 (top edge of gray boxes).

Nevertheless, even such moderate levels lead to elevated type I error rates.

DISCUSSION

Our simulations revealed that pseudoreplicated analyses can lead to greatly inflated probabilities to detect effects in stable isotope ratios of hair which do not exist (simulations 1, 2, and 4). As such they are in line with many other studies which investigated the effects of pseudoreplication, which revealed the exact same findings [e.g., Machlis et al., 1985; Mundry & Sommer, 2007]. However, most studies investigating the effects of pseudoreplicating on the level of individual investigated the effects of between subjects predictors whereas we also investigated whether pseudoreplication can also affect the probability of erroneous significance for within-subjects predictors. Our findings with regard to this part of the study revealed that, indeed, pseudoreplication can also lead to greatly inflated type I error rates for a within-subjects predictor (simulation 2b). Finally, we also investigated type II error probabilities for a within subjects predictor variable and found that pseudoreplication can lead to greatly inflated probabilities of erroneous non-significances, that is, missing to detect an actual effect in stable isotope ratios of hair. Hence, our findings are largely in line with those of other investigations of the effects of pseudoreplication by finding elevated type I error rates for within-subjects predictors but, to our knowledge, expand them by finding elevated type II error rates for within-subjects predictors (but see Barr et al., 2013). Importantly, our findings are not driven by differences in sample size as can be seen from the fact that models appropriately accounting for individual differences were not affected by pseudoreplication.

It is also important to emphasize that, despite assuming replicate observations of the same individuals to occur in the form of different "hair samples" of the same animal, our results are not specific to hair keratin samples. Instead we are confident that the results would equally hold for whichever are the specific samples taken, be it hairs or fecal samples from unidentified individuals, or whatever else.

While our finding of elevated type I error rates for between-subjects predictors is concordant with those of other studies of the effects of pseudoreplication in the sense that they lead to an elevated type I error rate, our results regarding the effects of pseudoreplication for the assessment of the significance of within-subjects predictors (here tested for the case of temporal autocorrelation/seasonality of isotope values within hair samples) are, to our knowledge, novel (but see Barr et al., 2013) and somewhat worrisome. In fact, we found that for such predictors pseudoreplication can lead to false significance as well as to erroneous non-significance, and that the probabilities of these two types of errors depend on various factors such as the magnitude of variation between individuals, the magnitude of the effect of the within-subjects predictor and autocorrelation, or the number of observations per individual. As a consequence, and since these different sources of variation will usually be unknown for samples from unhabituated primates, results obtained for a within-subjects predictor will most likely be uninterpretable since significance as well as non-significance can either be real or an artefact of pseudoreplication. This is in sharp contrast to results obtained for a between-subjects predictor, which can be an artefact of pseudoreplication when they indicate significance, but, to our knowledge, not when they indicate non-significance [see also Machlis et al., 1985].

We here based our investigation on inference based on *P*-values. The question may arise whether other ways of drawing inference (e.g., Bayesian or information theory based inference) may alleviate the issue and provide alternative means of hypothesis testing. Unfortunately, this is not the case. In fact, the assumption of independence of data (or residuals) is fundamental to all statistical approaches currently existing [e.g., Burnham & Anderson, 2002; McCarthy, 2007], and with regard to the consequences of pseudoreplication the only difference between the three statistical philosophies is how it manifests: while in null-hypothesis significance testing pseudoreplication manifests in the form of wrong *P*-values (and wrong standard errors of estimates as well as confidence intervals), in information theory based inference it leads to wrong values of the information criterion used (e.g., Akaike's Information Criterion), and in Bayesian inference it leads to wrong posterior probabilities. This similarity is due to the fact that all three approaches are based on a largely identical mathematical machinery to assess how well the model fits the response.

Given the severe consequences of pseudoreplication it might be worth assessing the probability of sampling any given individual repeatedly. Of course, this probability depends to a large extend on the sampling schema in relation to the life style (e.g., social organization and home range size) of the investigated species. For instance, when sampling in an area being considerably smaller than the usual home range of the individuals, the probability of having replicate samples of the same individual(s) will be large (as compared to when sampling in an area being considerably larger than the usual home range size of the target species), and this probability will further increase with increasing sample size. The probability of having at least one individual sampled more than once (thus violating the assumption of independence) is equivalent to the "birthday problem" [e.g., Frey, 2006] and equals $1-(\prod_{i=0}^{n-1}(n-i))/N^n$ where N is the size of the population and n is the

TABLE I. Probability of Having Sampled at Least One Individual Twice as a Function of the Size of the Population Sampled From (Columns) and the Number of Samples Taken (Rows)

	Population size																					
	10	11	12	13	14	15	16	17	18	19	20	25	30	35	40	45	50	60	70	80	90	100
Sample size																						
2	0.100	0.091	0.083	0.077	0.071	0.067	0.062	0.059	0.056	0.053	0.050	0.040	0.033	0.029	0.025	0.022	0.020	0.017	0.014	0.012	0.011	0.010
3	0.280	0.256	0.236	0.219	0.204	0.191	0.180	0.170	0.160	0.152	0.145	0.117	0.098	0.084	0.074	0.066	0.059	0.049	0.042	0.037	0.033	0.030
4	0.496	0.459	0.427	0.399	0.375	0.353	0.333	0.316	0.300	0.286	0.273	0.223	0.188	0.163	0.143	0.128	0.116	0.097	0.083	0.073	0.065	0.059
5	0.698	0.656	0.618	0.584	0.553	0.525	0.500	0.477	0.456	0.436	0.419	0.347	0.296	0.258	0.229	0.205	0.186	0.157	0.136	0.120	0.107	0.097
6	0.849	0.812	0.777	0.744	0.713	0.684	0.656	0.631	0.607	0.585	0.564	0.478	0.414	0.364	0.325	0.294	0.268	0.227	0.198	0.175	0.156	0.142
7	0.940	0.915	0.889	0.862	0.836	0.810	0.785	0.761	0.738	0.716	0.695	0.603	0.531	0.473	0.426	0.388	0.356	0.305	0.266	0.237	0.213	0.193
8	0.982	0.969	0.954	0.936	0.918	0.899	0.879	0.859	0.840	0.821	0.802	0.714	0.640	0.579	0.527	0.483	0.446	0.386	0.340	0.303	0.274	0.250
9	0.996	0.992	0.985	0.976	0.965	0.953	0.940	0.926	0.911	0.896	0.881	0.806	0.736	0.675	0.621	0.575	0.535	0.468	0.415	0.373	0.338	0.310
10	1.000	0.998	0.996	0.992	0.987	0.981	0.974	0.965	0.956	0.945	0.935	0.876	0.815	0.759	0.707	0.660	0.618	0.548	0.490	0.444	0.405	0.372
11		1.000	0.999	0.998	0.996	0.994	0.990	0.986	0.980	0.974	0.967	0.925	0.877	0.828	0.780	0.736	0.695	0.623	0.563	0.513	0.471	0.435
12			1.000	1.000	0.999	0.998	0.997	0.995	0.992	0.989	0.985	0.958	0.922	0.882	0.840	0.800	0.762	0.692	0.632	0.580	0.535	0.497
13				1.000	1.000	1.000	0.999	0.999	0.997	0.996	0.994	0.978	0.953	0.922	0.888	0.853	0.819	0.754	0.695	0.643	0.597	0.557
14					1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.990	0.973	0.951	0.925	0.896	0.866	0.807	0.752	0.701	0.656	0.615
15						1.000	1.000	1.000	1.000	1.000	0.999	0.995	0.986	0.971	0.951	0.928	0.904	0.852	0.801	0.753	0.709	0.669
16							1.000	1.000	1.000	1.000	1.000	0.998	0.993	0.983	0.969	0.952	0.932	0.889	0.844	0.800	0.758	0.718
17								1.000	1.000	1.000	1.000	0.999	0.997	0.991	0.982	0.969	0.954	0.919	0.880	0.840	0.801	0.763
18									1.000	1.000	1.000	1.000	0.999	0.995	0.989	0.981	0.970	0.942	0.909	0.874	0.838	0.804
19										1.000	1.000	1.000	0.999	0.998	0.994	0.988	0.981	0.959	0.932	0.902	0.871	0.839
20											1.000	1.000	1.000	0.999	0.997	0.993	0.988	0.972	0.951	0.925	0.898	0.870
25												1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.993	0.985	0.975	0.962
30													1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.996	0.992
35														1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
40															1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
45																1.000	1.000	1.000	1.000	1.000	1.000	1.000
50																	1.000	1.000	1.000	1.000	1.000	1.000
60																		1.000	1.000	1.000	1.000	1.000
70																			1.000	1.000	1.000	1.000
80																				1.000	1.000	1.000
90																					1.000	1.000
100																						1.000

number of samples taken from it. The probabilities of having at least one individual sampled more than once can be surprisingly large, even for few samples taken from large populations (Table I).

Since pseudoreplication has such adverse consequences for the interpretation of the results, the question arises how it can be avoided. The only isotope study using hair samples from largely unidentified individuals of great apes could at least partly test for the effect of "individual" on stable isotope differences between gorillas and chimpanzees as well as on the effect of season. At least one semi-habituated silverback gorilla was identified and sampled repeatedly and the results of three different models comprising the full dataset, the dataset without the identified silverback and a dataset only comprising the identified silverback (only testing for season) could be compared and did not reveal any contradicting results [Oelze et al., 2014]. While this was an attempt to find a workaround solution, we here suggest that more rigid steps should be considered during study design.

For future stable isotope studies in primates we mainly see two options to control for individual in statistical analyses: a) adjusting the sampling strategy in the field to avoid re-sampling of individuals and/or b) genotyping of hair and other tissue samples for isotope analyses. The first option, adjusting the sampling strategy, implies to space sampling locations such that the distances between them clearly exceed the typical home range diameter of a given primate community. For animals like great apes, which each day build new night nests and different individuals frequently nest in close vicinity to one another, hair (and dung) samples collected from different fresh nests of the same nest group will certainly derive from different individuals. Potentially one can even sample from more than one nest group in a given territory if it can be assured that these where build by different communities or foraging groups in the same night. Nevertheless, this attempt may be logistically demanding. As hair samples of great apes commonly cover a time period of 6–10 months [Oelze, This volume], temporal variation and thus seasonality can only be assessed within this given time frame. In case of such sampling schema, nest, sampling location (GPS coordinates) or nest group identification need to be included as a random effect into the respective model in case repeated samples taken from the same nest, location or nest group, respectively, are included into the analysis. The other option is to genetically sequence hair samples, for instance using the root of the hair or fresh feces commonly found in or under the night nests of great apes [Morin et al., 2001]. Genetic monitoring of endangered primate species, for example, via capture-recapture techniques, is an important approach in primate research and conservation [Arandjelovic et al., 2010, 2011]. Combined efforts in the field can aid the collection of hair and fecal samples.

In summary, we demonstrated that pseudoreplication can severely affect type I and type I error rates. Therefore, statistical analyses of data likely comprising repeated observations of individuals but not accounting for this fact bear a large risk of producing erroneous findings. Such erroneous findings can manifest as false significances (in case of between- and within-subjects predictors) as well false non-significances (in case of a within-subjects predictor), and they are fairly likely to happen, even at very moderate levels of pseudoreplication. Hence, every effort must be taken to ensure independent samples or to identify the individuals from which samples were obtained. If this is not possible results may be artefacts of pseudoreplication and lack interpretability.

ACKNOWLEDGMENTS

We thank Hjalmar Kühl and Erin Wessling for fruitful discussions and comments on an earlier version of this manuscript. We are particularly grateful for the fantastic IT support at the Max Planck Institute for Evolutionary Anthropology. Without the computation capacities provided and the help of Alexander Födisch, Andreas Walther and Rainer Benz who made it possible to use them in a seamless fashion it would have been pretty impossible to conduct this project. We also thank two anonymous reviewers and the editor Donald C. Dunbar for fruitful comments on earlier versions of this paper.

REFERENCES

Aiken LS, West SG. 1991. Multiple regression: testing and interpreting interactions. Newbury Park: Sage. p 224.

Arandjelovic M, Head J, Kühl H, et al. 2010. Effective non-invasive genetic monitoring of multiple wild western gorilla groups. Biological Conservation 143:1780–1791.

Arandjelovic M, Head J, Rabanal LI, et al. 2011. Non-invasive genetic monitoring of wild central chimpanzees. PLoS ONE 6:e14761.

Baayen RH. 2008. Analyzing linguistic data. Cambridge: Cambridge University Press. p 368.

Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. Journal of Memory and Language 68:255–278.

Blumenthal SA, Chritz KL, Rothman JM, Cerling TE. 2012. Detecting intraannual dietary variability in wild mountain gorillas by stable isotope analysis of feces. Proceedings of the National Academy of Sciences 109:21277–21282.

Bolker BM, Brooks ME, Clark CJ, et al. 2008. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127–135.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference. 2nd ed. Berlin: Springer. p 488.

Cerling TE, Wittemyer G, Ehleringer JR, Remien CH, Douglas-Hamilton I. 2009. History of Animals using Isotope Records (HAIR): a 6-year dietary history of one family of

African elephants. Proceedings of the National Academy of Sciences 106:8093–8100.
Crowley BE, Rasoazanabary E, Godfrey LR. 2014. Stable isotopes complement focal individual observations and confirm dietary variability in reddish-gray mouse lemurs (Microcebusgriseorufus) from southwestern Madagascar. American Journal of Physical Anthropology 155:77–90.
Cohen J, Cohen P. 1983. Applied multiple regression/correlation analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum Associates, Inc. p 691.
Dobson AJ. 2002. An introduction to generalized linear models. Boca Raton: Chapman & Hall/CRC. p 320.
Fahy GE, Richards M, Riedel J, Hublin J-J, Boesch C. 2013. Stable isotope evidence of meat eating and hunting specialization in adult male chimpanzees. Proceedings of the National Academy of Sciences 110:5829–5833.
Fruth B, Hohmann G. 1996. Nest building in the great apes: the great leap forward? In: McGrew WC, Marchant L, Nishida T, editors. Great ape societies. New York: Cambridge University Press. p 225–240.
Forstmeier W, Schielzeth H. 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behavioral Ecology and Sociobiology 65:47–55.
Frey B. 2006. Statistics hacks. Sebastopol, CA: O'Reilly. p 356.
Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press. p 607.
Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.
Loudon JE, Grobler P, Sponheimer M, et al. 2014. Using the stable carbon and nitrogen isotope compositions of vervet monkeys (*Chlorocebus pygerythrus*) to examine questions in ethnoprimatology. PloS ONE 9:e100758.
Morin PA, Chambers KE, Boesch C, Vigilant L. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytes verus). Molecular Ecology 10:1835–1844.
McCarthy MA. 2007. Bayesian methods in ecology. Cambridge: Cambridge University Press. p 225.
McCullagh P, Nelder JA. 1989. Generalized linear models. London: Chapman and Hall. p 532.
Machlis L, Dodd PWD, Fentress JC. 1985. The pooling fallacy: problems arising when individuals contribute more than one observation to the data set. Zeitschrift Für Tierpsychologie 68:201–214.
Mundry R. 2014. Statistical issues and assumptions of phylogenetic generalised least squares (PGLS). In: Garamszegi LZ, editor. Modern phylogenetic methods and their application in evolutionary biology. Heidelberg: Springer. p 131–153.
Mundry R, Sommer C. 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. Animal Behaviour 74:965–976.
Oelze VM, Fuller BT, Richards MP, et al. 2011. Exploring the contribution and significance of animal protein in the diet of bonobos by stable isotope ratio analysis of hair. Proceedings of the National Academy of Sciences 108:9792–9797.
Oelze VM, Head JS, Robbins MM, Richards M, Boesch C. 2014. Niche differentiation and dietary seasonality among sympatric gorillas and chimpanzees in Loango National Park (Gabon) revealed by stable isotope analysis. Journal of Human Evolution 66:95–106.
Oelze VM. This volume. Reconstructing temporal variation in great ape diets: a methodological framework for isotope analyses in non-invasively collected hair. American Journal of Primatology.
Quinn GP, Keough MJ. 2002. Experimental designs and data analysis for biologists. Cambridge: Cambridge University Press. p 527.
Schielzeth H, Forstmeier W. 2009. Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology 20:416–420.
Schoeninger MJ, Iwaniec UT, Glander KE. 1997. Stable isotope ratios indicate diet and habitat use in New World monkeys. American Journal of Physical Anthropology 103:69–83.
Schoeninger MJ, Moore J, Sept JM. 1999. Subsistence strategies of two "savanna" chimpanzee populations: the stable isotope evidence. American Journal of Primatology 49:297–314.
Schwertl M, Auerswald K, Schäufele R, Schnyder H. 2005. Carbon and nitrogen stable isotope composition of cattle hair: ecological fingerprints of production systems? Agriculture, Ecosystems and Environment 109:153–165.
Siegel S, Castellan NJ. 1988. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill. p 312.
Sponheimer M, Loudon JE, Codron D, et al. 2006. Do "savanna" chimpanzees consume C-4 resources? Journal of Human Evolution 51:128–133.
Waller BM, Warmelink L, Liebal K, Micheletta J, Slocombe KE. 2013. Pseudoreplication: a widespread problem in primate communication research. Animal Behaviour 86:483–488.
Zar JH. 1999. Biostatistical analysis. 4th ed. New Jersey: Prentice Hall. p 944.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.