Conditional repeatability and the variance explained by reaction norm variation in random slope models

Holger Schielzeth[1], Shinichi Nakagawa[2]

[1]Institute of Ecology and Evolution, Friedrich Schiller University Jena, Germany

[2]Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia

Data deposition: There is no data to be deposited. R scripts for simulations are available on Github (https://github.com/hschielzeth/RandomSlopeR2).

Address for correspondence

Holger Schielzeth, Institute of Ecology and Evolution, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany, holger.schielzeth@uni-jena.de

Abstract

Individuals differ in average phenotypes, but also in sensitivity to environmental variation. Such variation is biologically relevant, because it reflects variation in reaction norms. Between-individual variation in average phenotypes is typically quantified as random-intercept variation in linear mixed-effects models or as intra-class correlations (also known as repeatability). Similarly, context-sensitivity can be modelled as random-slope variation. However, random-slope variation implies that between-individual variation varies across the range of a covariate (environment, context, time or age) and has thus been called 'conditional' repeatability. While studies fitting random-slope models are on a rapid increase, there is a lack of a general concept for the quantification of context-sensitive between-individual variation. We here propose to put reaction-norm (random-slope) variation in perspective of the total phenotypic variance and suggest a way of standardization that we call random-slope coefficient of determination R_S^2. Furthermore, we illustrate that instead of the random-intercept variance, the average repeatability across an environmental gradient will be a biologically more relevant description of between-individual variation and we call this the marginalized repeatability R_{mar}. We provide simple equation to calculated key descriptors of conditional repeatabilities, clarify the difference between random-intercept variation and average between-individual variation and make recommendations for comprehensive reporting. Most importantly, reporting should include means and variances of covariates. While we introduce the concept with individual-variation in mind, the framework is equally applicable to other type of between-group/cluster variation that varies across some (environmental) gradient.

Keywords: intra-class correlation, between-individual variation, context-sensitivity, random-slope mixed effects models, conditional repeatability, reaction norm variation, standardized reporting

Introduction

Repeatabilities *R* and coefficients of variation R^2 allow a decomposition of the sources of biological variance in some response, feature or trait of interest. Repeatabilities (also known as intra-class correlations, ICC) are mostly concerned with a decomposition of random-effect variances (Nakagawa & Schielzeth 2010; Wolak *et al.* 2012). Repeatabilities have become particularly relevant in the study of labile and repeatedly expressed phenotypes (Bell *et al.* 2009). Coefficients of determination R^2 are used to quantify the variance explained in the fixed part of the model (marginal R^2; *sensu* (Nakagawa & Schielzeth 2013)). Repeatabilities and coefficients of variation are thus complementary quantities, one focusing on the random the other on the fixed part of the model.

Both, repeatabilities and coefficients of determination, quantify sources of variation in relation to the total variance in a response. To make this more concrete, we want to focus on variance decomposition in a context of the study of phenotypic variation, although the concepts are easily transferred to other systems. Imagine some flexible phenotype of some organism (this may be some physiological, endocrinological or behavioral trait, see e.g. (Nespolo & Franco 2007; Bell *et al.* 2009) that has been measured across multiple individuals with repeated observation per individual. Observed phenotypes y_{ij} are thus clustered within individuals *i* with repeated observations *j* per individual. The phenotypic equation represents a variance decomposition model that consists of a mechanistic (fixed effects) and idiosyncratic (random effects) part (Allegue *et al.* 2017):

$$y_{ij} = \alpha + \beta_1 \cdot x_{1,ij} + \beta_2 \cdot x_{2,ij} + \ldots + u_i + e_{ij}$$

Where i indexes individuals, j indexes observations, the terms $\beta \cdot x$ represent fixed effect predictors x and their slopes β with numbers indexing different predictors (of which there may be more than the two shown here). The sum of the terms $\beta \cdot x$ may be summarized as the linear predictor $\eta_{ij} = \sum \beta_k \cdot x_k$ and represents the fixed part of the model. The terms u and e are the random components, where u represents deviations of individuals from the population mean and e represents deviations of observations from individual means. Individual-level deviations u and observation-level deviations e are typically assumed to be normally distributed with mean of zero and variance estimated from the data. Since the linear predictor also explains some phenotypic variance, there are three variance components, σ^2_η, σ^2_u and σ^2_e, that can be interpreted as sources of biological variation.

We had previously discussed the various options for estimating ratios of these variance components (Nakagawa & Schielzeth 2010, 2013). Here we are concerned with another element, random-slope variation that blurs the distinction between the random and the fixed part of the model. Random slopes are interactions between fixed-effect covariates and random-effects: Slopes that vary by random-effect level (Gelman & Hill 2007; Dingemanse & Dochtermann 2013). Random-slopes are often important, because they allow to control for pseudoreplication in the estimation of the population slope (Schielzeth & Forstmeier 2009; Gurka *et al.* 2011). More importantly, however, they represent phenotypically plastic responses to an environment and are therefore relevant in the study of reaction norms (Nussey *et al.* 2007; Dingemanse *et al.* 2010). Random-slope models have become popular

in the study of ecology and evolution, because they reflect phenotypic plasticity such as an organisms' ability to cope with environmental changes.

Random-slope variation disrupts the concept of "the repeatability" (Biro & Stamps 2015). With random-slopes, the amount of between-individual variation, for example, depends on levels of the covariate, thus on context, environment, age or time. We have therefore introduced the term "conditional repeatability" for repeatabilities that vary by covariates (Nakagawa & Schielzeth 2010). Although random-slope models have become popular, we are not aware of any universal standardized measure of random-slope variation. And since meta-analyses on the magnitude of random-slope variation are missing, we also know very little about the magnitude of random-slope variation in natural systems. It therefore needs a system for estimating random-slope variation in a manner that is comparable across study system: It needs a method of standardization.

Johnson (2014) has introduced equations to estimate repeatabilities form random slope models. Johnson's method is based on the multiplication of the random-intercept random-slope variance-covariance matrix $\mathbf{\Sigma}$ with the design matrix $\mathbf{X}$ for the fixed effects (called $\mathbf{Z}$ in (Johnson 2014)). Effectively, Johnson's repeatability estimates average repeatabilities across the range of the covariate. As such, Johnston's repeatability is different from and typically larger than then random-intercept variation estimated in the model. In other words: With random-slope variation, the random-intercept variation is no longer a comprehensive parameter that describes the magnitude of individual differences (or differences among other types of groups). Random-intercept variation merely describes individual-differences at a single point of the environment.

Rights and Sterba (Rights & Sterba 2019a, b) have introduced a neat, comprehensive system for calculating various R^2 measures from mixed-effects models. Unlike Johnson (2014), they use the variances and covariances of fixed effects rather than the design matrices for the estimation. This offers a more concise version for reporting and it is therefore the approach that we adopt below. However, although Rights and Sterba (Rights & Sterba 2019a, b) strongly argue for single-source R^2 as the main focus of reporting and interpretation, we think that this does not capture the most biologically relevant estimates. An inherent feature of reaction norms that vary across contexts is that the between-individual variance is variable across contexts and an additive decomposition (as shown in (Rights & Sterba 2019a, b)) is no longer most efficient.

We here discuss the concept of conditional repeatabilities in some details. We start with theoretical considerations and the proceed to application. In particular, we introduce a calculation for variance-standardized random-slope variances and make recommendations for comprehensive reporting.

Theory

We first consider a phenotypic equation expressed as a mixed effect regression model and assume that all parameters are known with certainty.

$$y_{ij} = \alpha + (\beta + v_i) \cdot x_{ij} + u_i + \phi + e_{ij}$$

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma}\right), \ \mathbf{\Sigma} = \begin{bmatrix} V_u & C_{uv} \\ C_{uv} & V_v \end{bmatrix}$$

$$e_{ij} \sim N(0, V_R)$$

$$x_{ij} \sim D(\mu, V_x)$$

Where y_{ij} is the response of interest, α is the global intercept (an estimate of the population mean if covariates are centered), x_{ij} is a covariate (context, environment, age or time) that varies within individuals (an observation-level predictor), β is the population mean response to the covariate, u_i is the deviation of mean individual trait values from the population mean, v_i is the deviation of individual slopes from the population mean slope and e_{ij} are residual deviations. ϕ is merely introduced as a placeholder for other additive variance components such as additional fixed or random effects. The variance explained by ϕ is V_ϕ. e_{ij} are normally distributed with a variance V_R and u_i as well as v_i are multivariate normal distributed with variances of V_u and V_v, respectively, and a covariance of C_{uv}. The covariate x_{ij} is arbitrarily distributed (as symbolized by *D*) with mean μ and variance of V_x. We will first assume that covariances are grand-mean centered, such that $\mu = 0$ and lift the constraint later.

The equation translates into the following variance components (Allegue *et al.* 2017; Rights & Sterba 2019b):

Variance explained by fixed effects: $V_F = \beta^2 \cdot V_x$
Variance explained by individual: $V_I = V_u + V_v \cdot V_x$
Other variance components: V_ϕ
Residual variance: V_R
Phenotypic variance: $V_P = V_F + V_I + V_\phi + V_R$

V_I refers to the total variance explained by individual identity, including random intercept and random slope variance. This is an interesting quantity that summarizes individual differences, an important topic of current research (Réale *et al.* 2007; Stamps 2016), but it is different from (generally larger than) V_u if random-slope variation is non-zero.

We note that the phenotypic variance V_P as we calculate it here as the sum of additive variance components might differ slightly from the variance in response values as estimated from the raw data (Rights & Sterba 2019a). The difference is that the sum of the variance components aims to estimate the population variance while the variance in raw response values represents to variance in the sample. Since the population variance is what is relevant to biological interpretation (de Villemereuil *et al.* 2018), the sum of additive components is usually preferable. However, if components are not fully additive, this may lead to misestimation. It is therefore important to specify the variance decomposition correctly.

One component of the total between-individual variation V_I is the variance uniquely explained by random slopes:

$$V_S = V_v \cdot V_x$$

Note that we so far assume $\mu = 0$, which can be easily achieved by centering covariates prior to the analysis. Covariate centering is generally advisable when in random-slope models, because uncentered covariates tend to produces large covariation between random-slopes and random-intercepts, which often leads to convergence problems. If the covariate was not centered, then

$$V_S = V_v \cdot V_x + \mu^2 \cdot V_v$$

$$V_I = V_u + V_v \cdot V_x + \mu^2 \cdot V_v$$

The between-individual variation in average phenotypes is a little more difficult, because as a conditional repeatability it varies across the range of the covariate. We can calculate the amount of between-individual variation for any point x as (figure 1):

$$V_{I,x} = V_u + 2x \cdot C_{uv} + x^2 \cdot V_v .$$

The minimum value of $V_{I,x}$ (from where it increases in either direction) is reached at:

$$x_{\min} = \frac{-C_{uv}}{V_v} .$$

It follows that the minimum value of $V_{I,x}$ is:

$$\min(V_{I,x}) = V_u - \frac{C_{uv}^2}{V_v} .$$

With context-sensitive responses, it is difficult to conceptualize a pure among-individual variation in elevation across the entire gradient. We think that V_I is the best descriptor of overall individual differences (figure 1). One might be tempted to use $V_{I*} = V_I - V_S = V_u$ as an estimator of elevation, but this is just the between-individual variance at the point where the covariate is zero. Whether this is a meaningful value, depends on how the covariate is centered. The value might be representative for an average covariate value with mean-centered covariates (Schielzeth 2010). However, whether or not this is also the minimum value of between-individual variation depends on the intercept-slope covariance that, as a property of the population, is usually beyond experimental control.

Standardization

The above equations offer obvious ways of variance-standardization for both average between-individual variation V_I and random-slope variation V_S. We propose to call the variance-standardized average between-individual variation V_I the marginalized repeatability R_{mar}, because it marginalizes (averages) across the environmental gradient:

$$R_{mar} = \frac{V_I}{V_P} = \frac{V_u + V_v \cdot V_x + \mu^2 \cdot V_v}{V_F + V_I + V_\phi + V_R}$$

This value is typically larger than the variance-standardized random-intercept variation. We propose to call the variance-standardized random-slope variation the random-slope coefficient of determination R_S^2 (cf. $R_t^{2(v)}$; (Rights & Sterba 2019b)).

$$R_S^2 = \frac{V_S}{V_P} = \frac{V_v \cdot V_x + \mu^2 \cdot V_v}{V_F + V_I + V_\phi + V_R}$$

Variance-standardization puts the variance explained by individual components in perspective of the total phenotypic variance, which, in our experience, is what ecologists and evolutionary biologists are usually interested in (see (de Villemereuil *et al.* 2018) for a discussion). However, it has been argued that variance-standardization may produces different values not because of differences in the numerator, but because of differences in the denominator (Houle 1992). An alternative way of standardization is therefore standardization by the square of the mean trait value, if the trait is ratio-scale and the assumption that the variance explained increases with the square of the mean is reasonable (Houle *et al.* 2011). For mean-standardization, V_P has to be replaced by $\bar{y}^2$ (or α if all covariates are mean-centered) in the equations above.

Reporting

We write this article partly to encourage complete reporting for future meta-analyses of phenotypic plasticity. A first important message is that the mean of and variance in the covariate are important quantities that allow putting reaction norms and reaction norm variation in perspective of the phenotypic variance. One way to standardize random slopes is the use of variance-standardized covariates $x' = \frac{x - \bar{x}}{\sigma_x}$, in which case $\bar{x}' = 0$ and $V_{x'} = 1$. Alternatively, or better additionally, raw mean $\bar{x}$ and variance V_x should be reported. Furthermore, as we show below, it is important to report covariation among fixed effect. A second important message is that the correlation among random slopes and random intercepts is an important parameter that can be biologically interpreted and should be reported. Negative correlations show that between-individual variation is lower at high covariate values and *vice versa*.

For meta-analysis, it would also need some measures of uncertainty in all relevant estimates. This could be easily achieved by applying our equations to samples from the posterior distributions of models fit in a Bayesian framework (Gelman *et al.* 2004). In a likelihood framework, it could be achieved by parametric bootstrapping (Faraway 2014). However, some of the sampling variances are small in comparison to others and might not need to be available with full uncertainty estimates. For example, estimates of covariate means and variances are estimated with far higher precision than estimates of random effect variances. In fact, it might sometimes be useful to use means and variances for environmental covariates form independent data if the data were collected in an experimental setting where variance in the covariate was manipulated. (However, beware of out-of-sample predictions, see e.g. (Morrissey & Ruxton 2018))

Praxis

We now relax the assumption that quantities are known with certainty and focus on estimation. We have implemented simple simulation to illustrate some critical issues. We do not aim to present a full exploration of the full parameter space, that is potentially vast and varies between applications. For simulations that explored the power of different sampling designs for estimating for estimating random-slope variation we refer to (Martin *et al.*) and (van de Pol).

In brief, we implemented a data generating function for a simple random-slope model with a single grouping factor and two covariates. Random-slopes act on an observation-level predictor x. Furthermore, a group-level covariate ϕ was introduced with an associated slope γ. Random-slopes and random-intercepts were generated from a multivariate normal distribution with means of zero and covariance matrix $\mathbf{\Sigma}$. The generating phenotypic equation was:

$$y_{ij} = \alpha + (\beta + v_i) \cdot x_{ij} + u_i + \gamma \cdot \phi_{ij} + e_{ij}$$

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma} \right), \ \mathbf{\Sigma} = \begin{bmatrix} V_u & C_{uv} \\ C_{uv} & V_v \end{bmatrix}$$

$$e_{ij} \sim N(0, V_R)$$

$$x_{ij} \sim N(\mu_x, V_x)$$

$$\phi_{ij} \sim N(\mu_\phi, V_\phi)$$

We used the following parameter settings: $\alpha = 3$, $\beta = -0.5$, $\gamma = 0.5$, $V_R = 1$, $V_u = 1$, $V_v = 0.5$, $C_{uv} = 0.3$, $V_x = 1.2$, $V_\phi = 1$, $\mu_x = 0.5$, $\mu_\phi = -0.5$. Covariate values x were drawn from uniform distributions. These values were relatively arbitrarily chosen and effects are purposefully rather strong in order to demonstrate general patterns. The broad patterns are largely insensitive to the detailed choice of values. We simulated a population of 60 individuals with an average of 10 observations per individual, which we consider a moderate sample size. For each of 200 simulation runs we fitted the regression model, estimated the parameters and used the above equations to quantify important parts of conditional repeatability. Data were generated in R 3.6.2 (R Core Team 2020) and analyzed using random-slope models fitted in lme4 1.1-21 (Bates *et al.* 2015).

The basic setting showed that the conditional repeatability was quite accurately estimated, with only minor bias in $x_{\min}$ and slight downward bias in V_I (figure 2a). The accuracy in the estimation of model parameters and derived quantities is unequally distributed with some (like α, μ_x, V_x and V_R) being estimated with high accuracy while others (like C_{uv}, V_F V_ϕ and $x_{\min}$) being estimated with much less accurately (figure 3). We also simulated a reduced sample size of 30 individuals and an average of 3 observations per individual. The small-sample simulation resulted in more pronounced biased in $x_{\min}$ and V_I as comparted to the moderate sample size scenario and also a downward bias in the estimate of V_u (figure 2b).

We then assessed estimates when only random intercepts and no random slopes were fitted to the same data (note that data were generated with random-slope variation). These

models do not estimate conditional repeatabilities, but rather an overall repeatability. Interestingly, the random-intercept variance was close to (and only a little lower than) the total individual variance V_I as determined by the phenotypic equation (figure 4a). Evidently, this value was substantially larger than the V_u that was simulated, hence random-slope variation in the data was converted to between-individual variation in intercepts in the model. Yet, total individual variation as defined in our framework was not overestimated.

We further explored the effect of measurement error in the covariate on estimates of individual components. Linear models assume that covariates are measured without error (Snijders & Bosker 2011), hence this might sound like a non-sensible attempt. However, measurement error is inevitable, for example when within-subject centering is used to separate within and between-individual responses to some covariate (Raudenbush & Bryk 2002; van de Pol & Wright 2009). Simulations with the same moderate sample size as above and a rather large measured error of 30% in the covariate resulted in an underestimation of reaction norm variation V_S and underestimation of total between-individual variation V_I, but an overestimation of V_u (figure 4b). Hence, some of the reaction norm variation was converted to random-intercept variation. This is sometimes unavoidable as in the case of within-subject centering. If the amount of error can be estimated, it could, in principle, be corrected for.

Multiple and correlated predictors

We have above introduced equations that use means and variances of covariate x to quantify conditional repeatabilities. Johnson (Johnson) had introduced a different approach that uses the specific design matrices for the fixed effects instead. This approach will be particularly useful when fixed effect predictors are correlated, since positive correlations will inflate the contribution of a predictor to the phenotypic variance beyond $\beta \cdot V_x$.

$$V_I = Tr(\mathbf{X}\Sigma\mathbf{X}')/n$$

Where $\mathbf{X}$ is the model matrix for the intercept and the fixed effect of interest (typically a column of 1s for the intercept and a column of covariate values x for all observations), $\mathbf{X}'$ is the transpose of $\mathbf{X}$, n is the total number of observations and $\mathbf{\Sigma}$ is the random intercept-slope covariance matrix as defined above and as estimated from the data, Tr signifies the trace (the sum of the diagonal elements) of the resultant square matrix. To put it simply, V_I is calculated as the predicted amount of between-individual varaition associated with all observations and averaged across the dataset.

Johnson‘s (2014) approach can be useful for computation, but the use of means and variances of covariates makes reporting much easier. Our simulations show that means and variances are no less accurate, and yield unbiased estimates in most cases. Furthermore, it is also possible to estimate the variance in the fixed part explained by correlated predictors as:

$$V_F = \beta_1^2 \cdot V_{x_1} + \beta_2^2 \cdot V_{x_2} + 2 \cdot \beta_1 \cdot \beta_2 \cdot C_{x_1,x_2}$$

$$\mathbf{\Omega} = \begin{bmatrix} V_{x_1} & C_{x_1,x_2} \\ C_{x_1,x_2} & V_{x_2} \end{bmatrix}$$

where x_1 and x_2 are two fixed effect predictors with variances of V_{x_1} and V_{x_2}, respectively, and predictor covariance of C_{x_1,x_2} and β_1 as well as β_2 are the respective slopes. The correlation among predictors is summarized in the predictor variance-covariance matrix $\mathbf{\Omega}$. The predictor varaince-covariance matrix $\mathbf{\Omega}$ is suitable for concise reporting.

Conclusions

We present equations that allow the description of conditional between-individual variances. Most importantly, we introduce a way of standardized reporting of reaction norm variation, clarify the difference between random-intercept variation and average (marginalized) between individual variation and make recommendations for comprehensive reporting. By putting reaction norm variation in perspective of the phenotypic variance, we aim to promote comprehensive variance decomposition of natural variation in traits of interest. We hope that these tools will stimulate more research on context-sensitivity of individual (or other group-level) variation and will allow data for future meta-analyses to accumulate.

Acknowledgements

HS were supported by the German Research Foundation (DFG) as part of the SFB TRR 212 (NC³) (funding INST 215/543-1, 396782608). SN was supported by the ARC Discovery Project grant (DP180100818).

Data availability statement

There is no data to be deposited. R scripts for simulations are available on https://github.com/hschielzeth/RandomSlopeR2.

Author contributions

HS conceived the idea, implemented the simulations and drafted the manuscript; SN helped to refine the basic idea and substantially revise the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

References

Allegue H, Araya-Ajoy YG, Dingemanse NJ, Dochtermann NA, Garamszegi LZ, Nakagawa S, Réale D, Schielzeth H, Westneat DF (2017) Statistical Quantification of Individual Differences (SQuID): an educational and statistical tool for understanding multilevel phenotypic data in linear mixed models. *Methods in Ecology and Evolution* **8**, 257-267.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1-48.

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. *Animal Behaviour* **77**, 771-783.

Biro PA, Stamps JA (2015) Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. *Animal Behaviour* **105**, 223-230.

de Villemereuil P, Morrissey M, Nakagawa S, Schielzeth H (2018) Fixed effect variance and the estimation of repeatabilities and heritabilities: Issues and solutions. *Journal of Evolutionary Biology* **31**, 621-632.

Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology* **82**, 39-54.

Dingemanse NJ, Kazem AJ, Réale D, Wright J (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution* **25**, 81-89.

Faraway JJ (2014) *Linear models in R*, 2nd. CRC Press, Boca Raton.

Gelman A, Carlin JB, Stern HS, Rubin DR (2004) *Bayesian Data Analysis* Chapman & Hall, Boca Raton.

Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models* Cambridge University Press, Cambridge, U.K.

Gurka MJ, Edwards LJ, Muller KE (2011) Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine* **30**, 2696-2707.

Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195-204.

Houle D, Pelabon C, Wagner GP, Hansen TF (2011) Measurement and meaning in biology. *Quarterly Review of Biology* **86**, 3-34.

Johnson PCD (2014) Extension of Nakagawa & Schielzeth's R^{2GLMM} to random slopes models. *Methods in Ecology and Evolution* **5**, 944-946.

Martin JGA, Nussey DH, Wilson AJ, Réale D (2011) Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods in Ecology and Evolution* **2**, 362-374.

Morrissey MB, Ruxton GD (2018) Multiple regression is not multiple regressions: the meaning of multiple regression and the non-problem of collinearity. *Philosophy, Theory, and Practice in Biology* **10**, 3.

Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews* **85**, 935-956.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133-142.

Nespolo RF, Franco M (2007) Whole-animal metabolic rate is a repeatable trait: a meta-analysis. *Journal of Experimental Biology* **210**, 3877-3878.

Nussey DH, Wilson AJ, Brommer JE (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology* **20**, 831-844.

R Core Team (2020) *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna.

Raudenbush SW, Bryk AS (2002) *Hierarchical linear models: applications and data analysis methods*, 2nd. Sage Publications, London, U.K.

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. *Biological Reviews* **82**, 291-318.

Rights JD, Sterba SK (2019a) New recommendations on the Use of R-squared differences in multilevel model comparisons. *Multivariate Behavioral Research*, doi: 10.1080/00273171.00272019.01660605.

Rights JD, Sterba SK (2019b) Quantifying Explained Variance in Multilevel Models: An Integrative Framework for Defining R-Squared Measures. *Psychological Methods* **24**, 309-338.

Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* **1**, 103–113.

Schielzeth H, Forstmeier W (2009) Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology* **20**, 416-420.

Snijders TAB, Bosker RJ (2011) *Multilevel analysis: an introduction to basic and advanced multilevel modeling*, 2nd. Sage, London.

Stamps JA (2016) Individual differences in behavioural plasticities. *Biological Reviews* **91**, 534-567.

van de Pol M (2012) Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution* **3**, 268-280.

van de Pol MV, Wright J (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour* **77**, 753-758.

Wolak ME, Fairbairn DJ, Paulsen YR (2012) Guidelines for estimating repeatability. *Methods in Ecology and Evolution* **3**, 129-137.

Figure legends

Figure 1. Conceptual display of conditional repeatabilities. The figure shows the between individual variance as it depends on the value of the covariate. The random-intercept variance V_u is always estimated at a covariate value of zero. The minimum between-individual variance $\min(V_{I,x})$ is reached at $x_{\min}$. The average between-individual variance is V_I and is usually larger than V_u if there is random-slope variation.

Figure 2. Estimation error in conditional repeatabilities with (a) moderate ($N_{ind} = 60$, $N_{obs} = 600$) and (b) small ($N_{ind} = 30$, $N_{obs} = 90$) sample size. The figure shows the same quantities as Figure 1 with the data-generating (true) values shown black. Predicted conditional repeatabilities from 200 replications are shown in grey. Estimated V_I, V_u and $x_{\min} / \min(V_{I,x})$ are shown in orange, blue and green, respectively, with thin lines representing single iterations and bold lines average values.

Figure 3. Proportional bias in estimates for various model estimates and derived parameters for a base simulation setting (see main text) with 200 replicates. Results are broadly similar for a range of parameter values when the analysis equation matches with the data generation process. The interquartile range is shown in dark shading, the 95% envelope in moderate shading and the total range in light shading. Dots represent individual estimates from 200 replicates and red lines show mean values. Sim = Expectations based on simulation settings, Est = Estimation based on model fit.

Figure 4. Estimation error in conditional repeatabilities (a) when there is error in the covariate x (30% in this case) and (b) when the analysis model is fitted without random slopes. The figure shows the same quantities as Figure 1 with the data-generating (true) values shown black. Predicted conditional repeatabilities from 200 replications are shown in grey. Estimated V_I, V_u and $x_{\min} / \min(V_{I,x})$ are shown in orange, blue and green, respectively, with thin lines representing single iterations and bold lines average values. The case of missing random-slopes does not estimate conditional repeatabilities, but the estimated between-individual variance V_u closely approaches the simulated average between-individual variance V_I.

Figure 1

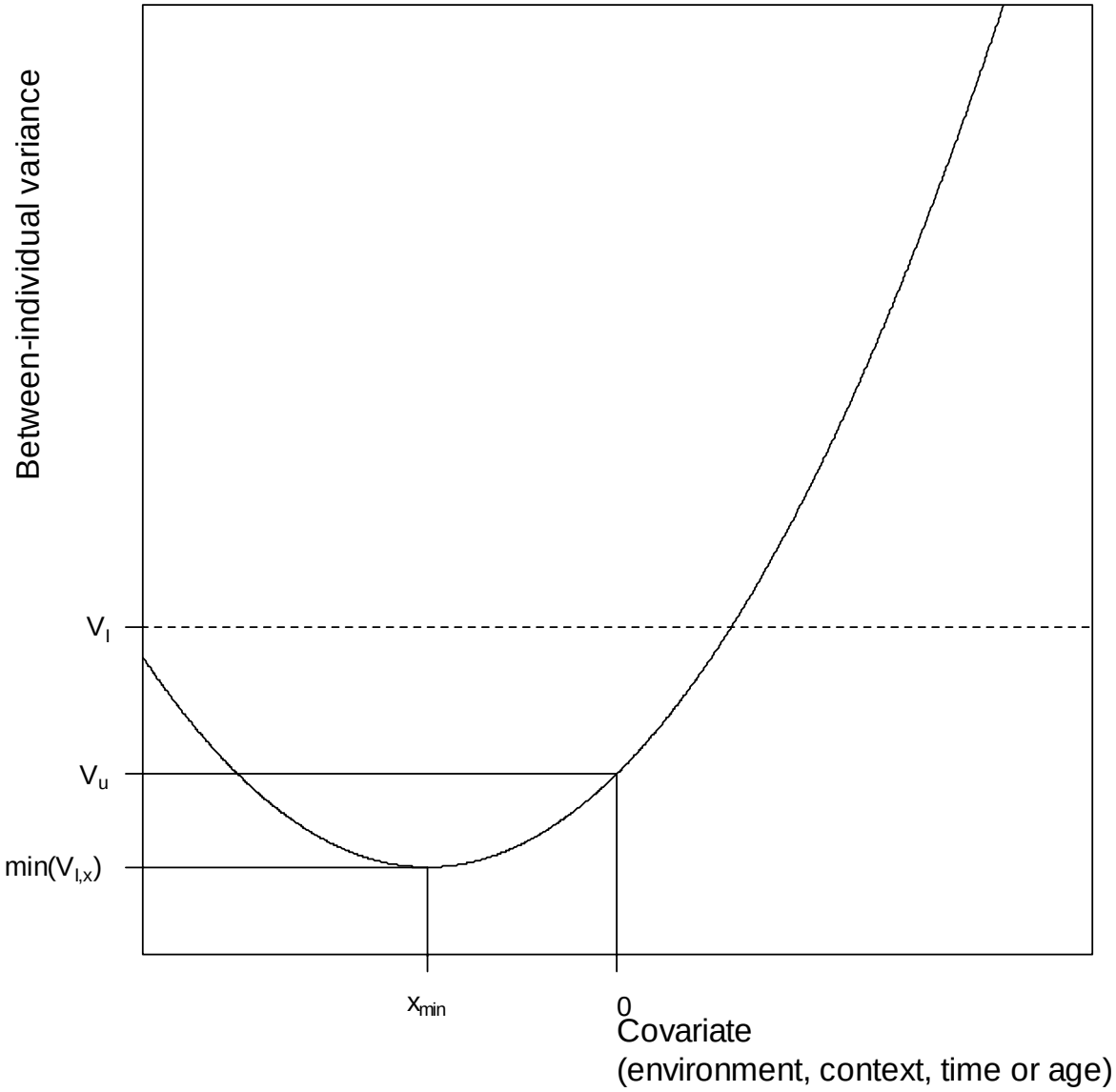

Figure 2

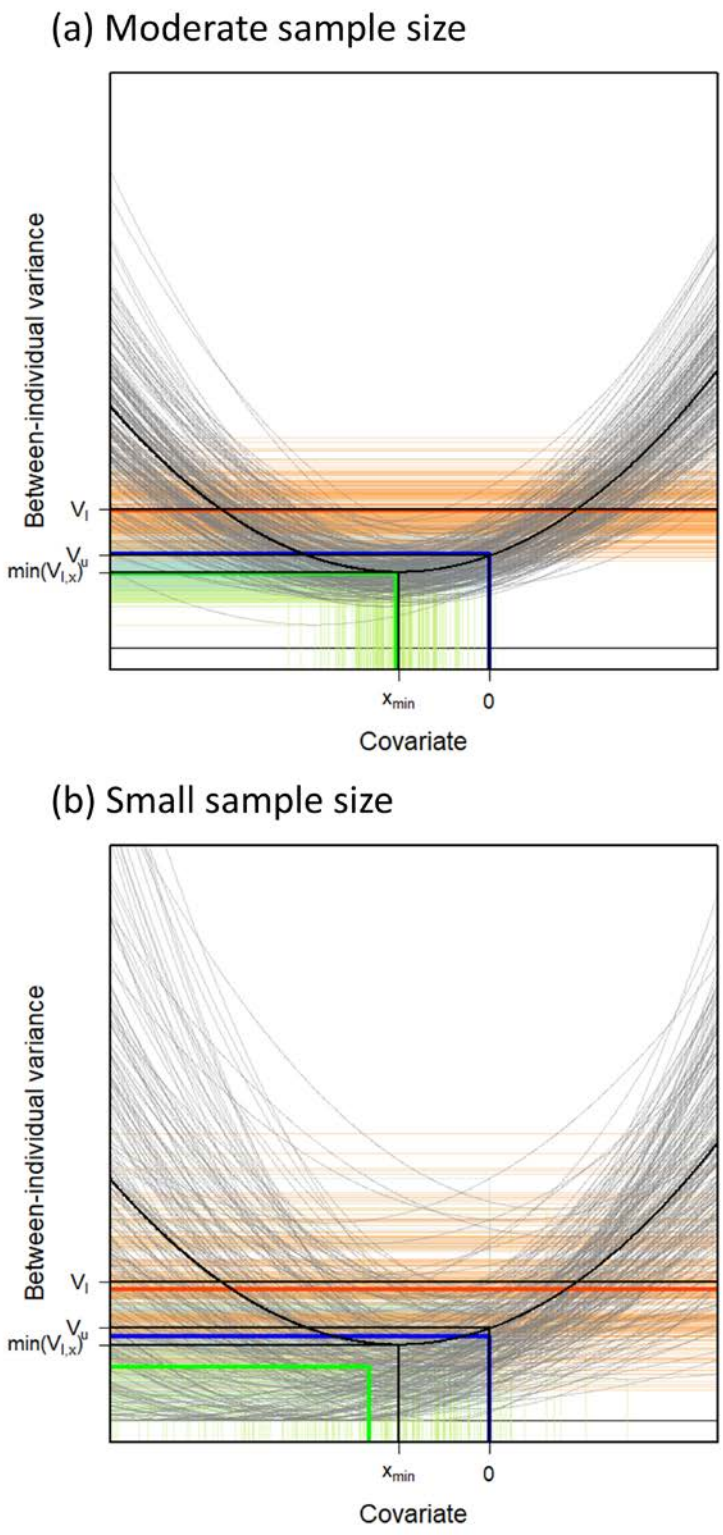

Figure 3

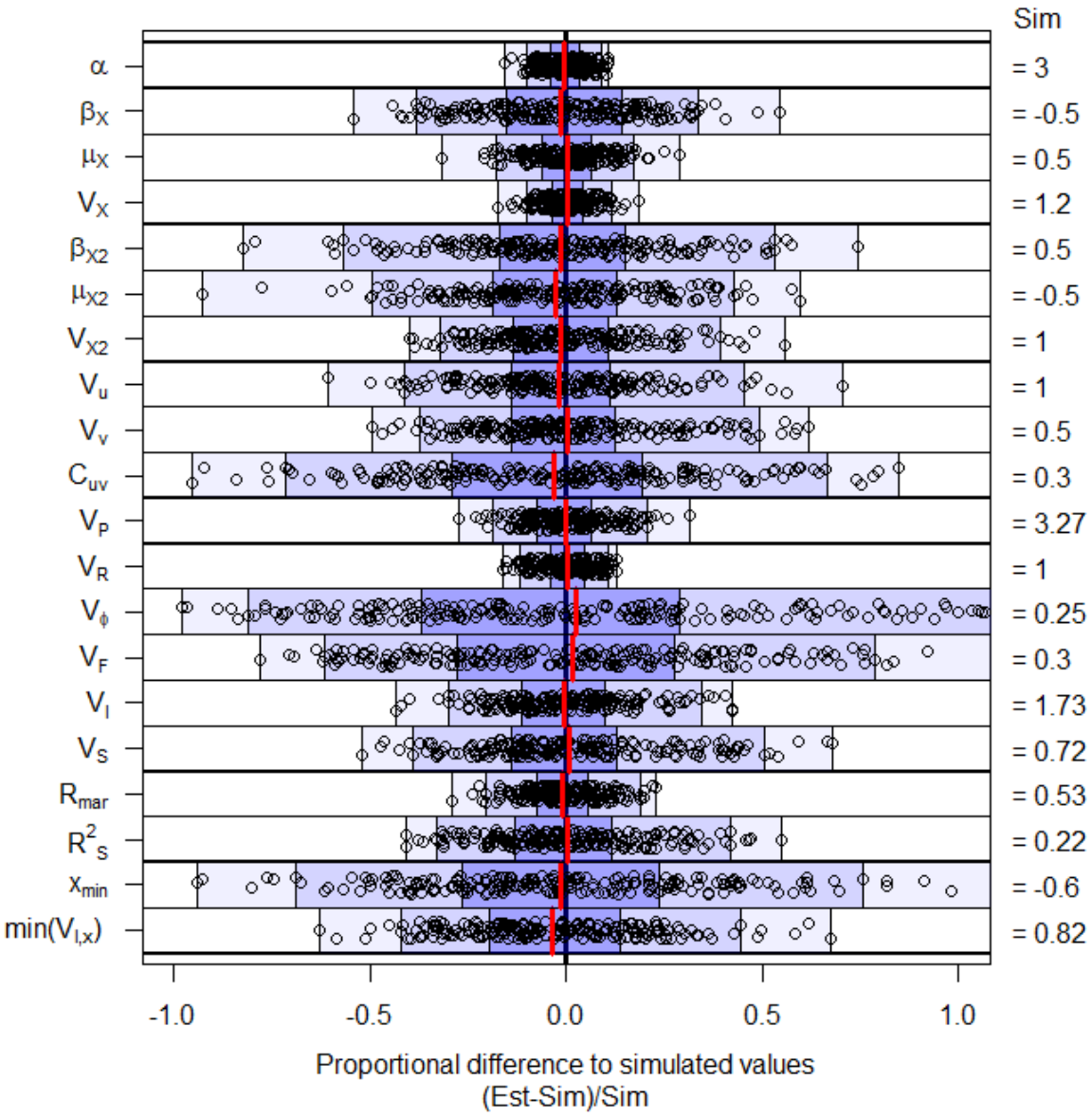

Figure 4