bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded

Shinichi Nakagawa[1, 2], Paul C. D. Johnson[3] and Holger Schielzeth[4]

[1] Evolution & Ecology Research Centre, and School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2] Diabetes and Metabolism Division, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia

[3] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Graham Kerr Building, Glasgow G12 8QQ, UK

[4] Population Ecology Group, Institute of Ecology, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany

Running head: R^2 and ICC from GLMMs

Address correspondence to S. Nakagawa. Email: s.nakagawa@unsw.edu.au; H. Schielzeth. Email: holger.schielzeth@uni-jena.de

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The coefficient of determination R^2 quantifies the proportion of variance explained by a statistical model and is an important summary statistic of biological interest. However, estimating R^2 for generalized linear mixed models (GLMMs) remains challenging. We have previously introduced a version of R^2 that we called R^2_{GLMM} for Poisson and binomial GLMMs, but not for other distributional families. Similarly, we earlier discussed how to estimate intra-class correlation coefficients ICC using Poisson and binomial GLMMs. In this article, we expand our methods to all other non-Gaussian distributions, in particular to negative binomial and gamma distributions that are commonly used for modelling biological data. While expanding our approach, we highlight two useful concepts for biologists, Jensen's inequality and the delta method, both of which help us in understanding the properties of GLMMs. Jensen's inequality has important implications for biologically meaningful interpretation of GLMMs, while the delta method allows a general derivation of variance associated with non-Gaussian distributions. We also discuss some special considerations for binomial GLMMs with binary or proportion data. We illustrate the implementation of our extension by worked examples from the field of ecology and evolution in the R environment. However, our method can be used across disciplines and regardless of statistical environments.

Key words: *repeatability, heritability, goodness of fit, model fit, variance decomposition, reliability analysis.*

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

One of the main purposes of linear modelling is to understand the sources of variation in biological data. In this context, it is not surprising that the coefficient of determination R^2 is a commonly reported statistic because it represents the proportion of variance explained by a linear model. The intra-class correlation coefficient ICC is a related statistic that quantifies the proportion of variance explained by a grouping (random) factor in multilevel/hierarchical data. In the field of ecology and evolution, a type of ICC is often referred to as repeatability R, where the grouping factor is often individuals that have been phenotyped repeatedly [1, 2]. We have reviewed methods for estimating R^2 and ICC in the past, with a particular focus on non-Gaussian response variables in the context of biological data [2, 3]. These previous articles featured generalized linear mixed-effects models (GLMMs) as the most versatile engine for estimating R^2 and ICC (specifically R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$). Our descriptions were limited to random-intercept GLMMs, but Johnson [4] has recently extended the methods to random-slope GLMMs, widening the applicability of these statistics (see also, [5, 6]).

However, at least one important issue seems to remain. Currently these two statistics are only described for binomial and Poisson GLMMs. Although these two types of GLMMs are arguably the most popular [7], there are other families of distributions that are commonly used in biology, such as negative binomial and gamma distributions [8, 9]. In this article, we revisit and extend R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ to more distributional families with a particular focus on negative binomial and gamma distributions. In this context, we discuss Jensen's inequality and two variants of the delta method, which are hardly known among biologists. These concepts are useful not only for generalizing our previous methods, but also for interpreting the results of GLMMs. Furthermore, we refer to some special considerations when obtaining R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ from binomial GLMMs for binary and proportion data, which we did not discuss in the past [2, 3]. We provide worked examples inspired from the field of ecology and evolution, focusing on implementation in

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the R environment [10] and finish by referring to two alternative approaches for obtaining R^2 and ICC from GLMMs along with a cautionary note.

2. Definitions of R^2_{GLMM}, $\mathrm{ICC}_{\mathrm{GLMM}}$ and overdispersion

To start with, we present R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ for a simple case of Gaussian error distributions based on a linear mixed-effects model (LMM, hence also referred to as R^2_{LMM} and $\mathrm{ICC}_{\mathrm{LMM}}$). Imagine a two-level dataset where the first level corresponds to observations and the second level to some grouping/clustering factor (e.g. individuals with repeated measurements) with k fixed effect covariates. The model can be written as (referred to as Model 1):

$$y_{ij} = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + \varepsilon_{ij}, \tag{2.1}$$

$$\alpha_i \sim \mathrm{Gaussian}(0,\ \sigma_\alpha^2), \tag{2.2}$$

$$\varepsilon_{ij} \sim \mathrm{Gaussian}(0,\ \sigma_\varepsilon^2), \tag{2.3}$$

where y_{ij} is the jth observation of the ith individual, x_{hij} is the jth value of the ith individual for the hth of k fixed effects predictors, β_0 is the (grand) intercept, β_h is the regression coefficient for the hth predictor, α_i is an individual-specific effect, assumed to be normally distributed in the population with the mean and variance of 0 and σ_α^2, ε_{ij} is an observation-specific residual, assumed to be normally distributed in the population with mean and variance of 0 and σ_ε^2, respectively. For this model, we can define two types of R^2 as:

$$R^2_{\mathrm{LMM}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{2.4}$$

$$R^2_{\mathrm{LMM}(c)} = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{2.5}$$

$$\sigma_f^2 = \mathrm{var}\left(\sum_{h}^{k} \beta_h x_{hij}\right), \tag{2.6}$$

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $R^2_{\mathrm{LMM}(m)}$ represents the marginal R^2, which is the proportion of the total variance explained by the fixed effects, $R^2_{\mathrm{LMM}(c)}$ represents the conditional R^2, which is the proportion of the variance explained by both fixed and random effects, and σ_f^2 is the variance explained by fixed effects [11]. Since marginal and conditional R^2 differ only in whether the random effect variance is included in the numerator, we avoid redundancy and present equations only for marginal R^2 in the following. Similarly, there are two types of ICC:

$$\mathrm{ICC}_{\mathrm{LMM(adj)}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} \tag{2.7}$$

$$\mathrm{ICC}_{\mathrm{LMM}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_f^2 + \sigma_\varepsilon^2} \tag{2.8}$$

If no fixed effects are fitted (other than the intercept), $\sigma_f^2 = 0$ so that $\mathrm{ICC}_{\mathrm{LMM(adj)}}$ equals $\mathrm{ICC}_{\mathrm{LMM}}$. In such a case, the ICC should not be called 'adjusted'(*sensu* [2]). For an ICC value to be adjusted for a source of variance, that variance must be more than 0 and omitted from the ICC calculation. Since the two versions of ICC differ only in whether the fixed effect variance, calculated as in equation (2.6), is included in the denominator, we avoid redundancy and present equations only for adjusted ICC in the following.

One of the main difficulties in extending R^2 from LMMs to GLMMs is defining the residual variance σ_ε^2. For binomial and Poisson GLMMs with an additive dispersion term, we have previously stated that σ_ε^2 is equivalent to $\sigma_e^2 + \sigma_d^2$ where σ_e^2 is the variance for the additive overdispersion term, and σ_d^2 is the distribution-specific variance [2, 3]. Here, overdispersion represents the excess variation relative to what is expected from a certain distribution and can be estimated by fitting an observation-level random effect (OLRE; see, [12, 13]). Alternatively, overdispersion in GLMMs can be implemented using a multiplicative overdispersion term [14]. In such an implementation, we stated that σ_ε^2 is equivalent to $\omega \cdot \sigma_d^2$ where ω is a multiplicative

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dispersion parameter estimated from the model [2]. However, obtaining σ_d^2 for specific distributions is not always possible, because in many families of GLMMs, σ_ε^2 (observation-level variance) cannot be clearly separated into σ_e^2 (overdispersion variance) and σ_d^2 (distribution-specific variance). It turns out that binomial and Poisson distributions are special cases where σ_d^2 can be usefully calculated, because either all overdispersion is modelled by an OLRE (additive overdispersion) or by a single multiplicative overdispersion parameter (multiplicative overdispersion). This is not the case for other families. However, as we will show below, we can always obtain the GLMM version of σ_ε^2 (on the latent scale) directly. We refer to this generalised version of σ_ε^2 as 'the observation-level variance' here rather than the residual variance (but we keep the notation σ_ε^2). Note that the observation-level variance, σ_ε^2, should not be confused with the variance associated with OLRE, which estimates σ_e^2 and can be considered to be a part of σ_ε^2.

3. Extension of R^2_{GLMM} and ICC_{GLMM}

We now define R^2_{GLMM} and ICC_{GLMM} for a quasi-Poisson (may also be referred to as overdispersed-Poisson) GLMM, because the quasi-Poisson distribution is an extension of Poisson distribution [15, 16], and is similar to the negative binomial distribution at least in their common applications [9, 17]. Imagine count data repeatedly measured from a number of individuals with associated data on k covariates. We fit a quasi-Poisson (QP) GLMM with the log link function (Model 2):

$$y_{ij} \sim \text{QP}(\lambda_{ij},\ \omega), \tag{3.1}$$

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i, \tag{3.2}$$

$$\alpha_i \sim \text{Gaussian}(0,\ \sigma_\alpha^2), \tag{3.3}$$

where y_{ij} is the jth observation of the ith individual and y_{ij} follows a quasi-Poisson distribution with two parameters, λ_{ij} and ω [15, 16], $\ln(\lambda_{ij})$ is the latent value for the jth observation of the ith

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

individual, ω is the overdispersion parameter (when the multiplicative dispersion parameter ω is 1, the model becomes a standard Poisson GLMM), α_i is an individual-specific effect, assumed to be normally distributed in the population with the mean and variance of 0 and σ_α^2, respectively (as in Model 1), and the other symbols are the same as above. Quasi-Poisson distributions have a mean of λ and a variance of $\lambda\omega$ (Table 1). For such a model, we can define $R^2_{\mathrm{GLMM}(m)}$ and (adjusted) $\mathrm{ICC}_{\mathrm{GLMM}}$ as:

$$R^2_{\mathrm{QP-ln}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1+\omega/\lambda)}, \tag{3.4}$$

$$\mathrm{ICC}_{\mathrm{QP-ln}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1+\omega/\lambda)}, \tag{3.5}$$

where the subscript of R^2 and ICC denote the distributional family, here QP-ln for quasi-Poisson distribution with log link, the term $\ln(1+\omega/\lambda)$ corresponds to the observation-level variance σ_ε^2 (Table 1, for derivation see Appendix S1), ω is the overdispersion parameter, and λ is the mean value of λ_{ij}. We discuss how to obtain λ below (Section 5).

The calculation is very similar for a negative binomial (NB) GLMM with the log link (Model 3):

$$y_{ij} \sim \mathrm{NB}(\lambda_{ij},\ \theta), \tag{3.6}$$

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i, \tag{3.7}$$

$$\alpha_i \sim \mathrm{Gaussian}(0,\ \sigma_\alpha^2), \tag{3.8}$$

where y_{ij} is the jth observation of the ith individual and y_{ij} follows a negative binomial distribution with two parameters, λ_{ij} and θ, where θ is the shape parameter of the negative binomial distribution (given by the software often as the dispersion parameter), and the other symbols are the same as above. The parameter θ is sometimes referred to as 'size'. Negative binomial distributions have a mean of λ and a variance of $\lambda + \lambda^2/\theta$ (Table 1). $R^2_{\mathrm{GLMM}(m)}$ and (adjusted) $\mathrm{ICC}_{\mathrm{GLMM}}$ for this model can be calculated as:

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$R^2_{\mathrm{NB-ln}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + 1/\lambda + 1/\theta)}, \tag{3.9}$$

$$\mathrm{ICC}_{\mathrm{NB-ln}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + 1/\lambda + 1/\theta)}, \tag{3.10}$$

Finally, for a gamma GLMM with the log link (Model 4):

$$y_{ij} \sim \mathrm{gamma}(\lambda_{ij},\ \nu), \tag{3.11}$$

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i, \tag{3.12}$$

$$\alpha_i \sim \mathrm{Gaussian}(0,\ \sigma_\alpha^2), \tag{3.13}$$

where y_{ij} is the jth observation of the ith individual and y_{ij} follows a gamma distribution with two parameters, λ_{ij} and ν, where ν is the shape parameter of the gamma distribution (sometimes statistical programs report $1/\nu$ instead of ν; also note that the gamma distribution can be parameterized in alternative ways, Table 1). Gamma distributions have a mean of λ and a variance of λ^2/ν (Table 1). $R^2_{\mathrm{GLMM}(m)}$ and (adjusted) $\mathrm{ICC}_{\mathrm{GLMM}}$ can be calculated as:

$$R^2_{\mathrm{gamma-ln}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + 1/\nu)}, \tag{3.15}$$

$$\mathrm{ICC}_{\mathrm{gamma-ln}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + 1/\nu)}, \tag{3.16}$$

4. Obtaining the observation-level variance by the 'first' delta method

For overdispersed Poisson, negative binomial and gamma GLMMs with log link, the observation-level variance σ_ε^2 can be obtained via the variance of the log-normal distribution (Appendix S1). This is the approach that has led to the terms presented above. There are two more alternative methods to obtain the same target: the delta method and the trigamma function. The two alternatives have different advantages and we will therefore discuss them in some detail in the following.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The delta method for variance approximation uses a first order Taylor series expansion, which is often employed to approximate the standard error (error variance) for transformations (or functions) of a variable *x* when the (error) variance of *x* itself is known (see [18]; for an accessible reference for biologists, [19]). The delta method for variance approximation can be written as:

$$\mathrm{var}[f(x)] \approx \mathrm{var}[x]\left(\frac{d}{dx}f(x)\right)^2, \tag{4.1}$$

where *x* is a random variable (typically represented by observations), *f* represents a function (e.g. log or square-root), var denotes variance, and *d*/*dx* is a (first) derivative with respect to variable *x*. Taking derivatives of any function can be easily done using the R environment (examples can be found in the Appendices). It is the delta method that Foulley and colleagues [20] used to derive the distribution-specific variance σ_d^2 for Poisson GLMMs as $1/\lambda$ (see also [21]). Given that $\mathrm{var}[y] = \lambda$ in the case of Poisson distributions and $d\ln(\lambda)/dx = 1/\lambda$, it follows that $\mathrm{var}[\ln(y)] \approx \lambda(1/\lambda)^2 = 1/\lambda$ (note that for Poisson distributions without overdispersion, σ_d^2 is equal to σ_ε^2 because $\sigma_e^2 = 0$).

One clear advantage of the delta method is its flexibility. We can easily obtain the observation-level variance σ_ε^2 for all kinds of distributions/link functions. For example, by using the delta method, it is straightforward to obtain σ_ε^2 for the Tweedie distribution, which has been used to model non-negative real numbers in ecology (e.g., [22, 23]). For the Tweedie distribution, the variance on the observed scale has the relationship $\mathrm{var}[y] = \varphi\mu^p$ where μ is the mean on the observed scale and φ is the dispersion parameter, comparable to λ and ω in equation (3.1), and *p* is a positive constant called an index parameter. Therefore, when used with the log-link function, σ_ε^2 can be approximated by $\varphi\mu^{(p-2)}$ according to equation (4.1). The log-normal approximation $\ln(1+\varphi\mu^{(p-2)})$ is also possible (see Appendix S1; Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The use of the trigamma function ψ_1 is limited to distributions with log link, but it is considered to provide the most accurate estimate of the observation level variance σ_ε^2 in those cases. This is because the variance of a gamma-distributed variable on the log scale is equal to $\psi_1(\nu)$ where ν is the shape parameter of the gamma distribution [24] and hence σ_ε^2 is $\psi_1(\nu)$. At the level of the statistical parameters (Table 1; on the 'expected data' scale; *sensu* [25]; see their Figure 1), Poisson and negative binomial distributions can both be seen as special cases of gamma distributions, and σ_ε^2 can be obtained using the trigamma function (Table 1). For example, σ_ε^2 for the Poisson distribution is $\psi_1(\lambda)$ with the speciality that in the case of Poisson distributions $\sigma_\varepsilon^2 = \sigma_d^2$. As we show in Appendix S2, ln(1+1/λ) (log-normal approximation), 1/λ (delta method approximation) and $\psi_1(\lambda)$ (trigamma function) give similar results when λ is greater than 2. Our recommendation is to use the trigamma function for obtaining σ_ε^2 whenever this is possible.

The trigamma function has been previously used to obtain observation-level variance in calculations of heritability (which can be seen as a type of ICC although in a strict sense, it is not; see [25]) using negative binomial GLMMs ([24, 26]; cf. [25]). Table 1 summarises observation-level variance σ_ε^2 for overdispersed Poisson, negative binomial and gamma distributions for commonly used link functions.

5. How to estimate λ from data

For some calculations, we require an estimate of the global expected value λ. Imagine a Poisson GLMM with log link and additive overdispersion fitted as an observation-level random effect (Model 5):

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}), \tag{5.1}$$

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{p} \beta_h x_{hij} + \alpha_i + e_{ij}, \tag{5.2}$$

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\alpha_i \sim \text{Gaussian}(0,\ \sigma_\alpha^2), \tag{5.3}$$

$$e_{ij} \sim \text{Gaussian}(0,\ \sigma_e^2), \tag{5.4}$$

where y_{ij} is the jth observation of the ith individual, and follows a Poisson distribution with the parameter λ_{ij}, e_{ij} is an additive overdispersion term for jth observation of the ith individual, and the other symbols are the same as above. Poisson distributions have a mean of λ and a variance of λ (cf. Table 1). Using the log-normal approximation $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} can be calculated as:

$$R^2_{\text{P-ln}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_e^2 + \ln(1+1/\lambda)}, \tag{5.5}$$

$$\text{ICC}_{\text{P-ln}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2 + \ln(1+1/\lambda)}, \tag{5.6}$$

where, as mentioned above, the term $\ln(1+1/\lambda)$ is σ_ε^2 (or σ_d^2) for Poisson distributions with the log link (Table 1).

In our earlier papers, we proposed to use the exponential of the intercept, $\exp(\beta_0)$ (from the intercept-only model) as an estimator of λ [2, 3]; note that $\exp(\beta_0)$ from models with any fixed effects will often be different from $\exp(\beta_0)$ from the intercept-only model. We also suggested that it is possible to use the mean of observed values y_{ij}. Unfortunately, these two recommendations are often inconsistent with each other. This is because, given Model 5 (and all the models in the previous section), the following relationships hold:

$$\exp(\beta_0) \le \text{E}[y_{ij}], \tag{5.7}$$

$$\text{E}[\lambda_{ij}] = \exp(\beta_0 + 0.5\sigma_\tau^2), \tag{5.8}$$

$$\text{E}[y_{ij}] = \text{E}[\lambda_{ij}], \tag{5.9}$$

where E represents the expected value (i.e., mean) on the observed scale, β_0 is the mean value on the latent scale (i.e. β_0 from the intercept-only model), σ_τ^2 is the total variance on the latent scale

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(e.g., $\sigma_\alpha^2 + \sigma_e^2$ in Models 1 and 5, and σ_α^2 in Models 2-4 [2]; see also [27]). In fact, exp(β_0) gives the median value of y_{ij} rather than the mean of y_{ij}, assuming a Poisson distribution. Thus, the use of exp(β_0) will often overestimate σ_d^2, providing smaller estimates of R^2 and ICC, compared to when using averaged y_{ij} (which is usually a better estimate of E[y_{ij}]). Quantitative differences between the two approaches may often be negligible, but when λ is small, the difference can be substantial so the choice of the method needs to be reported for reproducibility (Appendix S2). Our new recommendation is to obtain λ via equation (5.8), which is the Poisson parameter averaged across cluster-level parameters (λ_i for each individual in our example; [[17, 20, 28]]). Thus, obtaining λ via equation (5.8) will be more accurate than estimating λ by calculating the average of observed values although these two methods will give very similar or identical values when sampling is balanced (i.e., observations are equally distributed across individuals and covariates). This recommendation for obtaining λ also applies to negative binomial GLMMs (see Table 1).

6. Jensen's inequality and the 'second' delta method

A general form of equation (5.7) is known as Jensen's inequality, $g(\bar{x}) \leq \overline{g(x)}$ where g is a convex function. Hence, the transformation of the mean value is equal to or larger than the mean of transformed values (the opposite is true for a concave function; that is, $g(\bar{x}) \geq \overline{g(x)}$; [29]). In fact, whenever the function is not strictly linear, simple application of the inverse link function (or back-transformation) cannot be used to translate the mean on the latent scale into the mean value on the observed scale. This inequality has important implications for the interpretation of results from GLMMs, and also generalized linear models GLMs and linear models with transformed response variables.

Although log-link GLMMs (e.g., Model 5) have an analytical solution, equation (5.8), this is not usually the case. Therefore, converting the latent scale values into observation-scale values requires simulation using the inverse link function. However, the delta method for bias correction can be

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

used as a general approximation to account for Jensen's inequality when using link functions or transformations. This application of the delta method uses a second order Taylor series expansion [18, 30]. A simple case of the delta method for bias correction can be written as:

$$\mathrm{E}[f(x)] \approx f(x) + 0.5\sigma_\tau^2 \frac{d^2}{dx^2} f(x), \tag{6.1}$$

where d^2/dx^2 is a second derivative with respect to the variable x and the other symbols are as in equations (4.1) and (5.8). By employing this bias correction delta method (with $d^2 \exp(x)/dx^2 = \exp(x)$), we can approximate equation (5.8) using the same symbols as in equations (5.7)-(5.9):

$$\mathrm{E}[\lambda_{ij}] = \mathrm{E}[\exp(\beta_0)] \approx \exp(\beta_0) + 0.5\sigma_\tau^2 \exp(\beta_0) \tag{6.2}$$

The comparison between equation (5.8) (exact) and equation (6.2) (approximate) is shown in Appendix S3. The approximation is most useful when the exact formula is not available as in the case of a binomial GLMM with logit link (Model 6):

$$y_{ij} \sim \mathrm{binomial}(n_{ij},\ p_{ij},), \tag{6.3}$$

$$\mathrm{logit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij}, \tag{6.4}$$

$$\alpha_i \sim \mathrm{Gaussian}(0,\ \sigma_\alpha^2), \tag{6.5}$$

$$e_{ij} \sim \mathrm{Gaussian}(0,\ \sigma_e^2), \tag{6.6}$$

where y_{ij} is the number of 'success' in n_{ij} trials by the ith individual at the jth occasion (for binary data, n_{ij} is always 1), p_{ij} is the underlying probability of success, and the other symbols are the same as above. Binomial distributions have a mean of p and a variance of $np(1–p)$; (Table 2).

To obtain corresponding values between the latent scale and data (observation) scale, we need to account for Jensen's inequality. The logit function used in binomial GLMMs combines of concave

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and convex sections, which the delta method deals with efficiently. The overall intercept, β_0 on the latent scale could therefore be transformed not with the inverse (anti) logit function ($\text{logit}^{-1}(x) = \exp(x)/(1+\exp(x))$), but with the bias-corrected delta method approximation. Given that $d^2\text{logit}^{-1}(x)/dx^2 = \exp(x)(1-\exp(x))/(1+\exp(x))^3$ in the case of the binomial GLMM with the logit-link function, the approximation can be written as:

$$\text{E}[y_{ij}] = \text{E}[\text{logit}^{-1}(\beta_0)] \approx \frac{\exp(\beta_0)}{1+\exp(\beta_0)} + 0.5\sigma_\tau^2 \frac{\exp(\beta_0)(1-\exp(\beta_0))}{(1+\exp(\beta_0))^3}. \tag{6.7}$$

We can replace β_0 with any value obtained from the fixed part of the model (i.e. $\beta_0 + \sum \beta_h x_{hij}$). McCulloch and colleagues [31] provide another approximation formula, which, by using our notation, can be written as:

$$\text{E}[y_{ij}] \approx \text{logit}^{-1}\left(\beta_0 - 0.5\sigma_\tau^2 \tanh\left(\frac{\beta_0(1+2\exp(-0.5\sigma_\tau^2)}{6}\right)\right) \tag{6.8}$$

Yet, another approximation proposed by Zeger and colleagues [32] can be written as:

$$\text{E}[y_{ij}] \approx \text{logit}^{-1}\left(\beta_0\left[\sqrt{1+\left(\frac{16\sqrt{3}}{15\pi}\right)^2\sigma_\tau^2}\right]^{-1}\right). \tag{6.9}$$

This approximation, equation (6.9), uses the exact solution for the inverse probit function, which can be written for a model like Model 6 but using the probit link: i.e., $\text{probit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij}$ in place of equation (6.4):

$$\text{E}[y_{ij}] = \text{probit}^{-1}\left(\beta_0\sqrt{1+\sigma_\tau^2}^{-1}\right). \tag{6.10}$$

A comparison between equations (6.7), (6.8) and (6.9) is also shown in Appendix S3 (it turns out equation (6.8) gives the best approximation). Simulation will give the most accurate conversions

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

when no exact solutions are available. The use of the delta method for bias correction accounting for Jensen's inequality is a very general and versatile approach that is applicable for any distribution with any link function (see Appendix S3) and can save computation time. We note that the accuracy of the delta method (both variance approximation and bias correction) depends on the form of the function *f*, the conditions for and limitation of the delta method are described in the article by Oehlert [30].

7. Special considerations for binomial GLMMs

The observation-level variance σ_ε^2 can be thought of as being added to the latent scale on which other variance components are also estimated in a GLMM (equations (3.2), (3.7), (3.12), (5.2) and (6.4) for Models 2-6). Since the proposed R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ are ratios between variance components and their sums, we can show using the delta method that R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ calculated via σ_ε^2 approximate to those of R^2 and ICC on the observation (original) scale (shown in Appendix S4). In some cases, there exist specific formulae for ICC on the observation scale [2]. In the past, we distinguished between ICC on the latent scale and on the observation scale [2]. Such a distinction turns out to be strictly appropriate only for binomial distributions but not for Poisson distributions (and probably also not for other non-Gaussian distributions). This is because the property of what we have called the distribution-specific variance σ_d^2 for binomial distributions (e.g. $\pi^2/3$ for binomial error distribution with the logit link function) is quite different from what we have discussed as the observation-level variance σ_ε^2 although these two types of variance are related conceptually (i.e., both represents variance due to non-Gaussian distributions with specific link functions). Let us explain this further.

A binomial distribution with a mean of p (the proportion of successes) has a variance of $p(1-p)/n$ (the variance for the number of successes is $np(1-p)$; see Table 2). We find that the observation-level variance is $1/(np(1-p))$ using the delta method on the logit-link function (see Table 2). This

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observation-level variance $1/(np(1-p))$, or $1/(p(1-p))$ for binary data, is clearly different from the distribution-specific variance $\pi^2/3$. As with the observation-level variance for the log-Poisson model (which is $1/\lambda$ and changes with λ; note that we would have called $1/\lambda$ the distribution-specific variance; [2, 3]), the observation-level variance of the binomial distribution changes as p changes (see Appendix S5), suggesting these two observation-level variances ($1/\lambda$ and $1/(np(1-p))$ are analogous while the distribution-specific variance $\pi^2/3$ is not. Further, the minimum value of $1/(p(1-p))$ is 4, which is larger than $\pi^2/3 \approx 3.29$, meaning that the use of $1/p(1-p)$ in R^2 and ICC for binary data will always produce larger values than those using $\pi^2/3$. Consequently, Browne and colleagues [14] showed that ICC values (or variance partition coefficients, VPCs) estimated using $\pi^2/3$ were higher than corresponding ICC values on the observation (original) scale using logistic-binomial GLMMs (see also [33]). Note that they only considered binary data, i.e., $1/(np(1-p))$ where $n = 1$, because all proportion data can be rearranged as binary responses with a grouping/clustering factor.

Then, what is $\pi^2/3$? Three common link functions in binomial GLMMs (logit, probit and complementary log-log) all have corresponding distributions on the latent scale: the logistic distribution, standard normal distribution and Gumbel distribution, respectively. Each of these distributions has a theoretical variance, namely, $\pi^2/3$, 1 and $\pi^2/6$, respectively, which we previous referred to as distribution-specific variances [2, 3] (Table 2). As far as we are aware, these theoretical variances only exist for binomial distributions. The meaning of $1/(np(1-p))$, which is the variance on the latent scale that approximates to the variance due to binomial distributions on the observation scale is distinct from the meaning of $\pi^2/3$, which is the variance of the latent distribution (i.e., the logistic distribution with the scale parameter being 1). The use of the theoretical variance will almost always provide different values of R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$ from those using the observation-level obtained via the delta method (see Appendix S5). This is because the use of $\pi^2/3$ implicitly assumes all data sets have the same observation-level variance regardless of mean (p) given the same number of trials (n). Therefore, we need distinguishing these theoretical

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variances from the observation-level variance. R^2 and ICC values using the theoretical distribution-specific variance might be rightly called the latent (link) scale (*sensu* [2]) whereas, as mentioned above, R^2 and ICC values using the observation-level variance estimate the counterparts on the observation (original) scale (cf. [25]).

8. Worked examples: revisting the beetles

In the following, we present a worked example by expanding the beetle dataset that was generated for previous work [3]. In brief, the dataset represents a hypothetical species of beetle that has the following life cycle: larvae hatch and grow in the soil until they pupate, and then adult beetles feed and mate on plants. Larvae are sampled from 12 different populations (‘Population’; see Figure 1). Within each population, larvae are collected at two different microhabitats (‘Habitat’): dry and wet areas as determined by soil moisture. Larvae are exposed to two different dietary treatments (‘Treatment’): nutrient rich and control. The species is sexually dimorphic and can be easily sexed at the pupa stage (‘Sex’). Male beetles have two different color morphs: one dark and the other reddish brown (‘Morph’, labeled as A and B in Figure 1). Sexed pupae are housed in standard containers until they mature (‘Container’). Each container holds eight same-sex animals from a single population, but with a mix of individuals from the two habitats ($N_{[container]} = 120$; $N_{[animal]} = 960$).

We have data on five phenotypes, two of them sex-limited: (i) the number of eggs laid by each female after random mating which we had generated previously using Poisson distributions (with additive dispersion) and we revisit here for analysis with quasi-Poisson models (i.e. multiplicative dispersion), (ii) the incidence of endo-parasitic infections that we generated as being negative binomial distributed, (iii) body length of adult beetles which we had generated previously using Gaussian distributions and that we revisit here for analysis with gamma distributions, (iv) time to visit five predefined sectors of an arena (employed as a measure of exploratory tendencies) that we generated as being gamma distributed, and (v) the two male morphs, which was again generated

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with binomial distributions (for the details of parameter settings, see Table 3). We will use this simulated dataset to estimate R^2_{GLMM} and $\mathrm{ICC}_{\mathrm{GLMM}}$.

All data generation and analyses were conducted in R 3.3.1 [10]. We used functions to fit GLMMs from the three R packages: 1) the *glmmadmb* function from glmmADMB [34], 2) the *glmmPQL* function from MASS [35], and 3) the *glmer* and *glmer.nb* functions from lme4 [36]. In Table 4, we only report results from *glmmadmb* because this is the only function that can fit models with all relevant distributional families. All scripts and results are provided as an electronic supplement (Appendix S6). In addition, Appendix S6 includes an example of a model using the Tweedie distribution, which was fitted by the *cpglmm* function from the cplm package [23]. Notably, our approach for R^2_{GLMM} is kindly being implemented in the *rsquared* function in the R package piecewiseSEM [37]. Another important note is that we often find less congruence in GLMM results from the different packages than those of linear mixed-effects models, LMM. For example, GLMM using the gamma error structure with the log-link function (Size and Exploration models), *glmmadmb* and *glmmPQL* produced very similar results, while *glmer* gave larger R^2 and ICC values than the former two functions (for more details, see Appendix S6; also see [38]). Thus, it is recommended to run GLMMs in more than one package to check robustness of the results although this may not always be possible.

In all the models, estimated regression coefficients and variance components are very much in agreement with what is expected from our parameter settings (compare Table 3 with Table 4; see also Appendix S6). When comparing the null and full models, which had ‘sex’ as a predictor, the magnitudes of the variance component for the container effect always decrease in the full models. This is because the variance due to sex is confounded with the container variance in the null model. As expected, (unadjusted) ICC values from the null models are usually smaller than adjusted ICC values from the full models because the observation-level variance (analogous to the residual variance) was smaller in the full models, implying that the denominator of, for example, equation (3.5) shrinks. However, the numerator also becomes smaller for ICC values for the container effect

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from the parasite, size and exploration models so that adjusted ICC values are not necessarily larger than unadjusted ICC values. Accordingly, adjusted $ICC_{[container]}$ is smaller in the parasite and size models but not in the exploration model. The last thing to note is that for the morph models (binomial mixed models), both R^2 and ICC values are larger when using the distribution-specific variance rather than the observation-level variance, as discussed above (Table 4; see also Appendix S4).

9. Alternatives and a cautonary note

Here we extend our simple methods for obtaining R^2_{GLMM} and ICC_{GLMM} for Poisson and binomial GLMMs to other types of GLMMs such as negative binomial and gamma. We describe three different ways of obtaining the observational-level variance and how to obtain the key rate parameter λ for Poisson and negative binomial distributions. We discuss important considerations which arise for estimating R^2_{GLMM} and ICC_{GLMM} with binomial GLMMs. As we have shown, the merit of our approach is not only its ease of implementation but also that our approach encourages researchers to pay more attention to variance components at different levels. Research papers in the field of ecology and evolution often report only regression coefficients but not variance components of GLMMs [3].

We would like to highlight two recent studies that provide alternatives to our approach. First, Jaeger and colleagues [5] have proposed R^2 for fixed effects in GLMMs, which they referred to as $R^2_{\beta*}$ (an extension of an R^2 for fixed effects in linear mixed models or R^2_{β} by Edwards and colleagues [39]). They show that $R^2_{\beta*}$ is a general form of our marginal R^2_{GLMM}; in theory, $R^2_{\beta*}$ can be used for any distribution (error structure) with any link function. Jaeger and colleagues highlight that in the framework of $R^2_{\beta*}$, they can easily obtain semi-partial R^2, which quantifies the relative importance of each predictor (fixed effect). As they demonstrate by simulation, their method potentially gives a very reliable tool for model selection. One current issue for this approach is that implementation does not seem as simple as our approach (see also [40]). We note that our R^2_{GLMM} framework could

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also provide semi-partial R^2 via commonality analysis (see [41]), since unique variance for each predictor in commonality analysis corresponds to semi-partial $R^{2;}$; [42].

Second, de Villemereuil and colleagues [25] have provided a framework with which one can estimate exact heritability using GLMMs at different scales (e.g. data and latent scales). Their method can be extended to obtain exact ICC values on the data (observation) scale, which is analogous to, but not the same as, our ICC_{GLMM} using the observation-level variance, σ_ε^2 described above. Further, this method can, in theory, be extended to estimate R^2_{GLMM} on the data (observation) scale. One potential difficulty is that the method of de Villemereuil and colleagues is exact but that a numerical method is used to solve relevant equations so one will require a software package (e.g., the QGglmm package; [25]). Relevantly, they have shown that heritability on the latent scale does not need σ_d^2 (distribution-specific) but only need σ_e^2 (overdispersion variance), which has interesting consequences in relation to our R^2_{GLMM} and ICC_{GLMM} (we briefly describes this possibility in Appendix S7; see also [40]).

Finally, we finish by repeating what we said at the end of our original R^2 paper [3]. Both R^2 and ICC are indices that are likely to reflect only one or a few aspects of a model fit to the data and should not be used for gauging the quality of a model. We encourage biologists use R^2 and ICC in conjunctions with other indices like information criteria (e.g. AIC, BIC and DIC), and more importantly, with model diagnostics such as checking for model assumptions, heteroscedasticity and sensitivity to outliers.

Authors' contribuions

SN conceived ideas, and conducted analysis with discussions with HS. All developed the ideas further, and contributed to writing and editing of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing interests

We have no competing interests.

Funding

SN was supported by an Australian Research Council Future Fellowship (FT130100268). HS was supported by an Emmy Noether fellowship from the German Research Foundation (DFG; SCHI 1188/1-1).

Acknowledgements

We thank Losia Lagisz for help in making Figure 1. This work has been benefited from discussion with Jarrod Hadfield, Pierre de Villemereuil, Alistair Senior, Joel Pick and Dan Noble. We would also like to thanks an anonymous reviewer, whose comments have improved our manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

[1] Lessells, C.M. & Boag, P.T. 1987 Unrepeatable repeatabilities - a common mistake. *Auk* **104**, 116-121.

[2] Nakagawa, S. & Schielzeth, H. 2010 Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* **85**, 935-956. (doi:10.1111/j.1469-185X.2010.00141.x).

[3] Nakagawa, S. & Schielzeth, H. 2013 A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* **4**, 133-142. (doi:10.1111/j.2041-210x.2012.00261.x).

[4] Johnson, P.C.D. 2014 Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods Ecol Evol* **5**, 944-946. (doi:10.1111/2041-210x.12225).

[5] Jaeger, B.C., Edwards, L.J., Das, K. & Sen, P.K. 2017 An R^2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics* **44**, 1086-1105. (doi:10.1080/02664763.2016.1193725).

[6] LaHuis, D.M., Hartman, M.J., Hakoyama, S. & Clark, P.C. 2014 Explained variance measures for multilevel models. *Organ Res Methods* **17**, 433-451. (doi:10.1177/1094428114541701).

[7] Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. 2009 Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* **24**, 127-135. (doi: 10.1016/J.Tree.2008.10.008).

[8] Bolker, B.M. 2008 *Ecological models and data in R*. Princeton, NJ, Princeton University Presss.

[9] Ver Hoef, J.M. & Boveng, P.L. 2007 Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* **88**, 2766-2772.

[10] R Development Core Team. 2016 R: A language and environment for statistical computing. (version 2.15.0 ed. Vienna, Austria, R Foundation for Statistical Computing.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[11] Snijders, T. & Bosker, R. 2011 *Multilevel Analysis: an Introduction to basic and advanced multilevel modeling*. 2nd ed. London, Sage.

[12] Harrison, X.A. 2014 Using observation-level random effects to model overdispersion in count data in ecology and evolution. *Peerj* **2**, e616. (doi:ARTN e61610.7717/peerj.616).

[13] Harrison, X.A. 2015 A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *Peerj* **3**, e1114. (doi:ARTN e111410.7717/peerj.1114).

[14] Browne, W.J., Subramanian, S.V., Jones, K. & Goldstein, H. 2005 Variance partitioning in multilevel logistic models that exhibit overdispersion. *J R Stat Soc a Stat* **168**, 599-613.

[15] Efron, B. 1986 Double Exponential-Families and Their Use in Generalized Linear-Regression. *J Am Stat Assoc* **81**, 709-721. (doi:10.2307/2289002).

[16] Gelfand, A.E. & Dalal, S.R. 1990 A note on overdispersed exponential-families. *Biometrika* **77**, 55-64. (doi:10.2307/2336049).

[17] Gelman, A. & Hill, J. 2006 *Data analysis using regression and multilevel/hierarchical models* Cambridge, Cambridge University Press.

[18] Ver Hoef, J.M. 2012 Who Invented the Delta Method? *Am Stat* **66**, 124-127. (doi:10.1080/00031305.2012.687494).

[19] Powell, L.A. 2007 Approximating variance of demographic parameters using the delta method: A reference for avian biologists. *Condor* **109**, 949-954. (doi: 10.1650/0010-5422(2007)109[949:Avodpu]2.0.Co;2).

[20] Foulley, J.L., Gianola, D. & Im, S. 1987 Genetic Evaluation of Traits Distributed as Poisson-Binomial with Reference to Reproductive Characters. *Theor. Appl. Genet.* **73**, 870-877. (doi: 10.1007/Bf00289392).

[21] Gray, B.R. & Burlew, M.M. 2007 Estimating trend precision and power to detect trends across grouped count data. *Ecology* **88**, 2364-2372.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[22] Foster, S.D. & Bravington, M.V. 2013 A Poisson-Gamma model for analysis of ecological non-negative continuous data. *Environ Ecol Stat* **20**, 533-552. (doi:10.1007/s10651-012-0233-0).

[23] Zhang, Y.W. 2013 Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Stat Comput* **23**, 743-757. (doi:10.1007/s11222-012-9343-7).

[24] Tempelman, R.J. & Gianola, D. 1999 Genetic analysis of fertility in dairy cattle using negative binomial mixed models. *J Dairy Sci* **82**, 1834-1847.

[25] de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. 2016 General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. *Genetics* **204**, 1281-1294. (doi:10.1534/genetics.115.186536).

[26] Matos, C.A.P., Thomas, D.L., Gianola, D., Tempelman, R.J. & Young, L.D. 1997 Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models .1. Estimation of genetic parameters. *J Anim Sci* **75**, 76-87.

[27] Carrasco, J.L. 2010 A generalized concordance correlation coefficient based on the variance components generalized linear mixed models for overdispersed count data. *Biometrics* **66**, 897-904. (doi: 10.1111/J.1541-0420.2009.01335.X).

[28] Foulley, J.L. & Im, S. 1993 A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models. *Genet Sel Evol* **25**, 101-107. (doi: 10.1051/gse:19930107).

[29] Rao, C.R. 2002 *Linear statistical inference and its applications*. 2nd ed. ed. New York, John Wiley & Sons.

[30] Oehlert, G.W. 1992 A note on the delta method. *Am Stat* **46**, 27-29. (doi: 10.2307/2684406).

[31] McCulloch, C.E., Searle, S.R. & Neuhaus, J.M. 2008 *Generalized, linear, and mixed models*. 2nd ed. ed. Hoboken, N.J., Wiley ; Chichester : John Wiley [distributor].

[32] Zeger, S.L., Liang, K.Y. & Albert, P.S. 1988 Models for Longitudinal Data - a Generalized Estimating Equation Approach. *Biometrics* **44**, 1049-1060. (doi: 10.2307/2531734).

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[33] Goldstein, H., Browne, W. & Rasbash, J. 2002 Partitioning variation in multilevel models. *Understand Stat* **1**, 223-231.

[34] Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. 2012 AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim Method Softw* **27**, 233-249. (doi:10.1080/10556788.2011.597854).

[35] Venables, W.N. & Ripley, B.D. 2002 *Modern applied statistics with S*. 4 ed. New York, Springer.

[36] Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. 2015 Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**, 1-48.

[37] Lefcheck, J.S. 2016 PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol Evol* **7**, 573-579. (doi:10.1111/2041-210x.12512).

[38] Brooks, M.E., Kristensen, K.K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. & Bolker, B.M. 2017 Modeling zero-inflated count data with glmmTMB. *bioRxiv* (doi: 10.1101/132753).

[39] Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F. & Schabenberger, O. 2008 An R^2 statistic for fixed effects in the linear mixed model. *Stat Med* **27**, 6137-6157. (doi: 10.1002/Sim.3429).

[40] Ives, A.R. 2017 R2s for correlated data: phylogenetic models, LMMs, and GLMMs. *bioRxiv* (doi: 10.1101/144170).

[41] Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. 2014 Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods Ecol Evol* **5**, 320-328. (doi: 10.1111/2041-210x.12166).

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[42] Nimon, K.F. & Oswald, F.L. 2013 Understanding the results of multiple linear regression: beyond standardized regression coefficients. *Organ Res Methods* **16**, 650-674. (doi:10.1177/1094428113493929).

Table 1. The observation-level variance σ_ε^2 for the three distributional families: quasi-Poisson, negative binomial and gamma with the three different methods for deriving σ_ε^2: the delta method, log-normal approximation and the trigamma function, ψ_1.

Family	Distributional parameters	Mean (E[y]) Variance (var[y])	Link function	Delta method	Log-normal approximation	Trigamma function
Quasi-Poisson (QP)	$\mathrm{OP}(\lambda,\ \omega)$	$\mathrm{E}[y]=\lambda$	log	$\frac{\omega}{\lambda}$	$\ln\left(1+\frac{\omega}{\lambda}\right)$	$\psi_1\left(\frac{\lambda}{\omega}\right)$
Poisson (when $\omega=1$)	$\lambda>0$ $\omega>0$	$\mathrm{var}[y]=\lambda\omega$	square-root	0.25ω	-	
Negative binomial (NB)	$\mathrm{NB}(\lambda,\ \theta)$	$\mathrm{E}[y]=\lambda$	log	$\frac{1}{\lambda}+\frac{1}{\theta}$	$\ln\left(1+\frac{1}{\lambda}+\frac{1}{\theta}\right)$	$\psi_1\left(\left[\frac{1}{\lambda}+\frac{1}{\theta}\right]^{-1}\right)$
	$\lambda>0$ $\theta>0$	$\mathrm{var}[y]=\lambda+\frac{\lambda^2}{\theta}$	square-root	$0.25\left(1+\frac{\lambda}{\theta}\right)$	-	

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gamma	gmma(λ, ν)	$\mathrm{E}[y]=\lambda$	log	$\frac{1}{\nu}$	$\ln\left(1+\frac{1}{\nu}\right)$	$\psi_1(\nu)$
	$\lambda > 0$ $\nu > 0$	$\mathrm{var}[y]=\frac{\lambda^2}{\nu}$	inverse (reciprocal)	$\frac{1}{\nu\lambda^2}$	-	
Gamma (alternative parameterization)	gamma(ν, κ)	$\mathrm{E}[y]=\frac{\nu}{\kappa}$	log	$\frac{1}{\nu}$	$\ln\left(1+\frac{1}{\nu}\right)$	$\psi_1(\nu)$
	$\nu > 0$ $\kappa > 0$	$\mathrm{var}[y]=\frac{\nu}{\kappa^2}$	inverse (reciprocal)	$\frac{\kappa^2}{\nu^3}$	-	

$\mathrm{var}[\ln(x)]=\psi_1(\nu)=\sum_{n=1}^{\infty}1/(\nu+n)$ when x follows gamma distribution. In the R environment, the function, *trigamma* can be used to obtain $\psi_1(\nu)$; also note that ν is known as a shape parameter while κ is as a rate parameter in gamma distribution.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. The distribution-specific (theoretical) variance σ_d^2 and observation-level variance σ_ε^2 using the delta method for binomial (and Bernoulli) distributions; note that only one of them should be used for obtaining R^2 and ICC.

Family	Distributional parameters, mean & variance	Link name	Link function	Theoretical (distribution-specific) variance	Observation-level variance (min. values and corresponding p given n = 1)
Binomial (Bernoulli; n = 1)	binomial(n, p) $0 < p < 1$ n > = 1 (integers) E[y] = np var[y] = $np(1-p)$ var[y/n] = $p(1-p)/n$	logit	$\ln\left(\frac{p}{1-p}\right)$	$\frac{\pi^2}{3} \approx 3.29$ (logistic distribution)	$\frac{1}{np(1-p)}$ (min = 4; p = 0.5)
		probit ($\Phi(p)$)	$\sqrt{2}\mathrm{erf}^{-1}(2p-1)$	1 (standard normal distribution)	$2\pi n^{-1}p(1-p)\left(\exp\left[\mathrm{erf}^{-1}(2p-1)\right]^2\right)^2$ (min ~ 1.57; p = 0.5)

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cloglog (complimentary log-log)	$\ln(-\ln(1-p))$	$\frac{\pi^2}{6} \approx 1.65$ (Gumbel distribution)	$\frac{p}{n\left(\ln(1-p)\right)^2(1-p)}$ (min ~ 1.54; p ~ 0.8; ~ 2.08; p = 0.5)

'erf^{-1}' is the inverse of the Gauss error function, which is often denoted as 'erf'.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Parameter settings of regression coefficients (*b*) and variance components (σ^2) for five data sets: 1) fecundity, 2) endoparasite, 3) size, 4) exploration and 5) morph; all parameters are set on the latent scale apart from the size data (see below).

Response	Intercept (*b*)	Sex (*b*)	Treatment (*b*)	Habitat (*b*)	Population (σ^2)	Container (σ^2)	Overdispersion (σ^2)
Fecundity: the number of eggs per female	1.1	-	0.5	0.1	0.4	0.05	0.1
Parasite: the number of endoparasites per individual	1.8	-2	-0.8	0.7	0.5	0.8	-
Size: the body length of an individual*	15	-3	0.4	0.15	1.3	0.3	1.2
Exploration: the time taken visiting five sectors for an individual	4	-1	2	-0.5	0.2	0.2	-
Morph colour morph of a male	-0.8	-	0.8	0.5	1.2	0.2	-

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Data for the six sets of models were simulated on the normal (Gaussian) scale but analysed assuming a gamma error structure with the log link so that estimations of these parameters will be on the log scale; note the overdispersion variance for this data is the residual variance.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. Mixed-effects model analysis of a simulated dataset estimating variance components and regression slopes for nutrient manipulations on fecundity, endoparasite loads, body length, exploration levels and male morph types; $N_{[population]}$=12, $N_{[container]}$=120 and $N_{[animal]}$=960 ($N_{[male]}$= $N_{[female]}$ = 480).

Model name	Fecundity models (log-link)		Parasite models (log-link)		Size models (log-link)		Exploration models (log-link)		Morph models (logit-link)	
	Quasi-Poisson mixed models		Negative binomial mixed models		Gamma mixed models		Gamma mixed models		Binomial (binary) mixed models	
	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model
Fixed effects	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*
	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]	[95% CI]
Intercept	1.630	1.261	0.766	1.752	2.682	2.737	4.752	4.056	-0.108	-0.740
	[1.379, 1.882]	[0.989, 1.532]	[0.330, 1.202]	[1.282, 2.223]	[2.616, 2.689]	[2.699, 2.775]	[4.555, 4.949]	[3.842, 4.269]	[-0.718, 0.501]	[-1.450, -0.030]
Treatment (experiment)	-	0.491	-	-0.768	-	0.033	-	2.007	-	0.840
		[0.391, 0.591]		[-0.870, -0.667]		[0.023, 0.044]		[1.965, 2.050]		[0.422, 1.258]

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Habitat (wet)	-	0.152	-	0.700	-	0.009	-	-0.560	-	0.414
		[0.055, 0.249]		[0.599, 0.801]		[-0.001, 0.019]		[-0.603, -0.518]		[0.002, 0.826]
Sex (male)	-	-	-	-2.198	-	-0.213	-	-1.105	-	-
				[-2.511, -1.884]		[-0.230, -0.196]		[-1.256, -0.955]		-
Random effects	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2	σ^2
Population	0.178	0.187	0.375	0.541	0.0026	0.0039	0.071	0.104	1.002	1.111
Container	0.042	0.059	1.976	0.613	0.0140	0.0014	0.364	0.163	0.136	0.186
Observation-level (Distribution-specific)	0.477	0.349	0.873	0.397	0.0069	0.0064	1.664	0.118	4.010 (3.290)	4.010 (3.290)
Fixed factors	-	0.066	-	1.479	-	0.0116	-	1.393	-	0.220
$R^2_{\mathrm{GLMM}(m)}$	-	9.96%	-	48.50%	-	49.54%	-	78.34%	-	3.98% (4.57%)

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

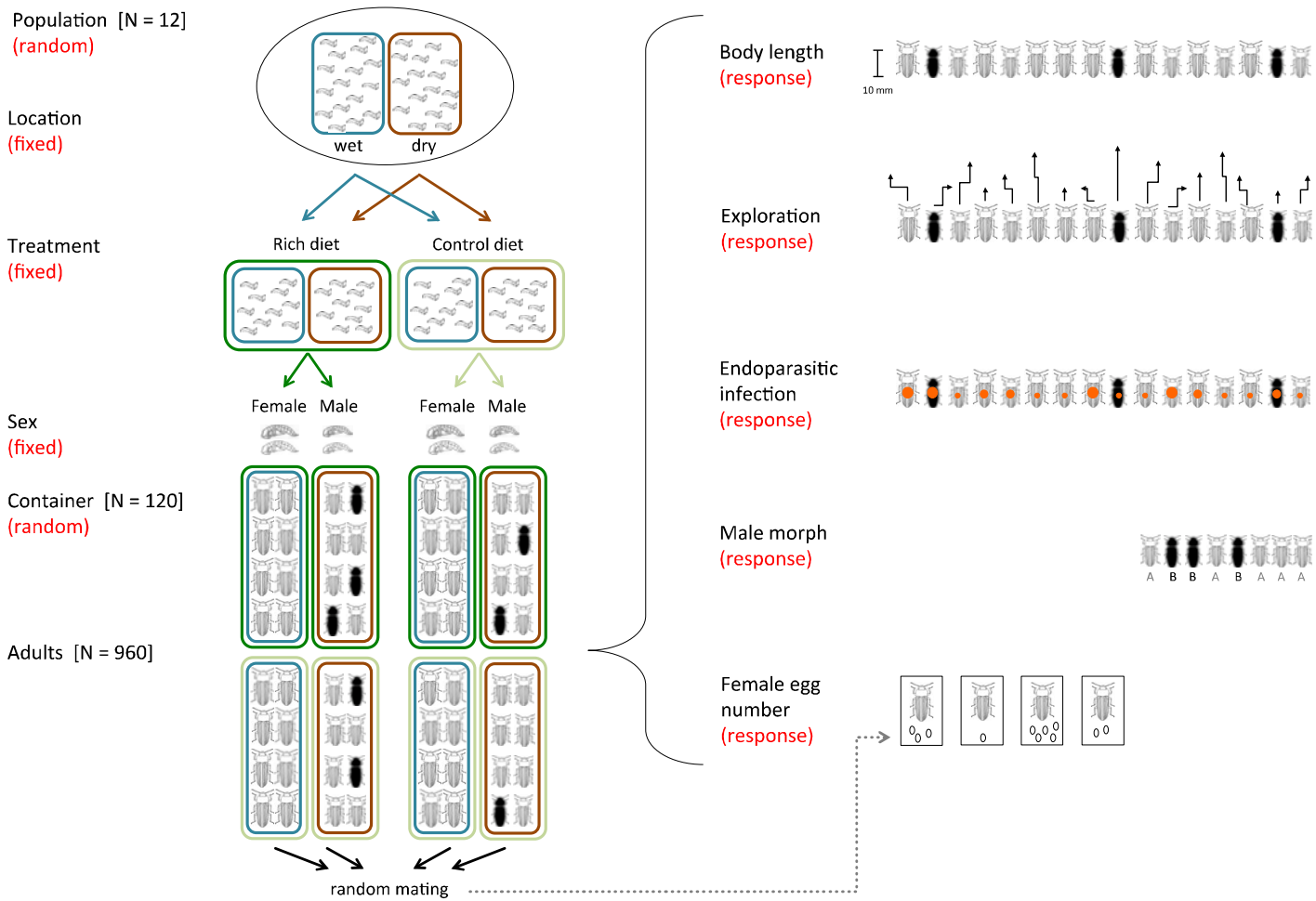
$R^2_{\mathrm{GLMM}(c)}$	-	46.95%	-	86.33%	-	72.52%	-	93.34%	-	27.46% (31.55%)
$\mathrm{ICC}_{[\mathrm{Population}]}$	25.33%	31.30%	11.53%	34.44%	11.38%	33.17%	3.40%	26.94%	19.48% (22.64%;)	20.95% (24.23%)
$\mathrm{ICC}_{[\mathrm{Container}]}$	5.94%	9.79%	60.80%	39.02%	59.57%	12.37%	17.34%	42.34%	2.64% (3.07%;)	3.50% (4.05%)
AIC	2498.8	2412.3	4342.6	3920.5	3379.9	3139.5	11223.8	9004.3	605.5	589.6

95 % CI (confidence intervals) were calculated by the *confint* function in lme4. The observation-level variance was obtained by using the trigamma function. In the Morph models, both the observation-level variance and (theoretical) distribution-specific variance were used; note that ones in brackets use the distribution-specific variance for R^2 and ICC. $\mathrm{ICC}_{[\mathrm{Container}]}$ is not a typical 'repeatability' but the proportion of variance due to the container effect beyond the population variance.

bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legends

Figure 1. A schematic of how hypothetical datasets are obtained (see the main text for details).



bioRxiv preprint doi: https://doi.org/10.1101/095851; this version posted July 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.