Extending R^2 and intra-class correlation coefficient from generalized linear mixed-effects models: capturing and characterizing biological variation

Shinichi Nakagawa[1,2] and Holger Schielzeth[3]

[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] 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

1

1   **Abstract**

2   The coefficient of determination R^2 quantifies the proportion of variance explained by a statistical

3   model and is an important summary statistic of biological interest. However, estimating R^2 for

4   (generalized) linear mixed models (GLMMs) remains challenging. We have previously introduced

5   a version of R^2 that we called R^2_{GLMM} for Poisson and binomial GLMMs using biological examples,

6   but not for other distributional families. Similarly, we earlier discussed how to estimate intra-class

7   correlation coefficients ICC using only Poisson and binomial GLMMs. In this article, we expand

8   our methods to all the other non-Gaussian distributions such as negative binomial and gamma

9   distributions, which are common in biological data. While expanding our approach, we highlight

10   two useful concepts for biologists, Jensen's inequality and the delta method, both of which help us

11   in understanding the properties of GLMMs. Jensen's inequality has important implications for

12   biologically more meaningful interpretation of GLMMs, while the delta method allows a general

13   derivation of distribution-specific variances. We also discuss some special considerations for

14   binomial GLMMs with binary or proportion data. We illustrate the implementation of our extension

15   by worked examples from the field of ecology and evolution in the R environment although our

16   method can be used regardless of statistical environments.

17   ***Key words:*** *repeatability, regression, heritability, goodness of fit, information criteria,*

18   *variance explained, model fit, variance decomposition, reliability analysis.*

19

2

1. Introduction

20

21 One of the main purposes of linear modellinåg is to understand the sources of variation in biological

22 data. In this context, it is not surprising that the coefficient of determination R^2 is a commonly

23 reported statistic because it represents the proportion of variance explained by a linear model. The

24 intra-class correlation coefficient ICC is a related statistic that quantifies the proportion of variance

25 explained by a grouping (random) factor in multilevel/hierarchical data. In the field of ecology and

26 evolution, a type of ICC is often referred to as repeatability R, where the grouping factor is often

27 individuals that have been phenotyped repeatedly [1, 2] . We have reviewed methods for estimating

28 R^2 and ICC in the past, with a particular focus on non-Gaussian response variables in the context of

29 biological data [2, 3]. These previous articles featured generalized linear mixed-effects models

30 (GLMMs) as the most versatile engine for estimating R^2 and ICC (specifically R^2_{GLMM} and

31 ICC_{GLMM}). Our descriptions were limited to random-intercept GLMMs, but Johnson [4] has

32 recently extended the methods to random-slope GLMMs, widening the applicability of these

33 statistics (see also, [5, 6]).

34 However, at least one important issue seems to remain. Currently these two statistics are only

35 described for binomial and Poisson GLMMs. Although these two types of GLMMs are arguably the

36 most popular [7], there are other common families of distributions in biology, such as negative

37 binomial and gamma distributions [8, 9]. In this article, we revisit and extend R^2_{GLMM} and ICC_{GLMM}

38 to more distributional families with a particular focus on negative binomial and gamma

39 distributions. In this context, we discuss Jensen's inequality and two variants of the delta method,

40 which are rarely known among biologists. However, these concepts are useful not only for

41 generalizing our previous methods, but also for interpreting the results of GLMMs for biologists.

42 Furthermore, we refer to some special considerations when obtaining R^2_{GLMM} and ICC_{GLMM} from

43 binomially GLMMs for binary and proportion data, which we did not discuss in the past [2, 3]. We

44 provide worked examples inspired from the field of ecology and evolution, focusing on

45 implementation in the R environment [10] and finish by referring to two alternative approaches for

46 obtaining R^2 and ICC from GLMMs along with a cautionary note.

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

48 To start with, we present R^2_{GLMM} and ICC$_{GLMM}$ for a simple case of Gaussian error distributions

49 based on a linear mixed-effects model (LMM, hence also referred to as R^2_{LMM} and ICC$_{LMM}$).

50 Imagine a two-level dataset where the first level corresponds to observations and the second level to

51 some grouping factor (e.g. individuals) with k fixed effect covariates. The model can be written as

52 (model 1):

53 $$y_{ij} = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + \varepsilon_{ij},$$ (2.1)

54 $$\alpha_i \sim \text{Gaussian}(0, \ \sigma_\alpha^2),$$ (2.2)

55 $$\varepsilon_{ij} \sim \text{Gaussian}(0, \ \sigma_\varepsilon^2),$$ (2.3)

56 where y_{ij} is the jth observation of the ith individual, x_{hij} is the jth value of the ith individual for the

57 hth of k fixed effects predictors, β_0 is the (grand) intercept, β_h is the regression coefficient for the

58 hth predictor, α_i is an individual-specific effect, assumed to be normally distributed in the

59 population with the mean and variance of 0 and σ_α^2, ε_{ij} is an observation-specific residual, assumed

60 to be normally distributed in the population with mean and variance of 0 and σ_ε^2, respectively. For

61 this model, we can define two types of R^2 as:

62 $$R^2_{\text{LMM}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2},$$ (2.4)

63 $$R^2_{\text{LMM}(c)} = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2},$$ (2.5)

64 $$\sigma_f^2 = \text{var}\left(\sum_h^k \beta_h x_{hij}\right),$$ (2.6)

4

65   where $R^2_{\mathrm{LMM}(m)}$ represents the marginal R^2, which is the variance accounted for by the fixed effects,

66    $R^2_{\mathrm{LMM}(c)}$ represents the conditional R^2, which is the variance explained by both fixed and random

67   effects, and σ^2_f is the variance explained by fixed effects [11, 12]. Since marginal and conditional

68   R^2 differ only in whether the random effect variance is included in the numerator, we avoid

69   redundancy and present equations only for marginal R^2 in the following.

70   Similarly, there are two types of ICC:

71    $$\mathrm{ICC}_{\mathrm{LMM(adj)}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\varepsilon} \qquad\qquad (2.7)$$

72    $$\mathrm{ICC}_{\mathrm{LMM}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_f + \sigma^2_\varepsilon} \qquad\qquad (2.8)$$

73   If no fixed effects are included, the two versions are identical and represent unadjusted ICC, but if

74   fixed effects are fitted, $\mathrm{ICC}_{\mathrm{LMM(adj)}}$ represents adjusted ICC, while $\mathrm{ICC}_{\mathrm{LMM}}$ represented unadjusted

75   ICC (*sensu* [2]). Since the two versions of ICC differ only in whether the fixed effect variance,

76   calculated as in equation (2.6), is included in the denominator, we avoid redundancy and present

77   equations only for adjusted ICC in the following.

78   One of the main difficulties in extending R^2 from LMMs to GLMMs is defining the residual

79   variance σ^2_ε. For binomial and Poisson GLMMs with an additive dispersion terms, we have

80   previously stated that σ^2_ε is equivalent to $\sigma^2_e + \sigma^2_d$ where σ^2_e is the variance for the additive

81   overdispersion term, and σ^2_d is the distribution-specific variance [2, 3]. Here overdispersion

82   represents the excess variation relative to what is expected from a certain distribution and can be

83   estimated by fitting an observation-level random effect (OLRE; see, [13, 14]). Alternatively,

84   overdispersion in GLMMs can be implemented using a multiplicative overdispersion term [15]. In

85   such an implementation, we stated that σ^2_ε is equivalent to $\omega \cdot \sigma^2_d$ where ω is a multiplicative

86   dispersion parameter estimated from the model [2]. However, obtaining σ^2_d for specific

87   distributions is not always possible, because in many families of GLMMs the parameters are less

5

88   clearly separated into a parameter for the expectation of the mean and a parameter for the

89   (over)dispersion. It turns out that binomial and Poisson distributions are special cases where σ_d^2 can

90   be usefully calculated, because either all overdispersion is modelled by an OLRE (additive

91   overdispersion) or by a single multiplicative overdispersion parameter (multiplicative

92   overdispersion). However, as we will show below, we can always obtain the GLMM version of σ_ε^2

93   (on the latent scale) directly. We refer to this generalised version of σ_ε^2 as 'the observation-level

94   variance' here rather than the residual variance (but we keep the notation σ_ε^2).

3. Extension of R^2_{GLMM} and ICC$_{\text{GLMM}}$

96   We now define R^2_{GLMM} and ICC$_{\text{GLMM}}$ for an overdispersed Poisson (also known as quasi-Poisson)

97   GLMM, because the overdispersed Poisson distribution is similar to the negative binomial

98   distribution at least in their uses[9, 16]. Imagine count data repeatedly measured from a number of

99   individuals with associated data on k covariates. We fit an overdispersed Poisson (OP) GLMM with

100   the log link function (model 2):

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

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

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

104   where y_{ij} is the jth observation of the ith individual and y_{ij} follows an overdispersed Poisson

105   distribution with two parameters, λ_{ij} and ω, $\ln(\lambda_{ij})$ is the latent value for the jth observation of the ith

106   individual, ω is the overdispersion parameter (when the multiplicative dispersion parameter ω is 1,

107   the model becomes a standard Poisson GLMM), α_i is an individual-specific effect, assumed to be

108   normally distributed in the population with the mean and variance of 0 and σ_α^2, respectively (as in

109   model 1), and the other symbols are the same as above. For such a model, we can define $R^2_{\text{GLMM}(m)}$

110   and (adjusted) ICC$_{\text{GLMM}}$ as:

6

111 $$R^2_{\text{OP}-\ln(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \ln(1+\omega/\lambda)},$$ (3.4)

112 $$\text{ICC}_{\text{OP}-\ln} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \ln(1+\omega/\lambda)},$$ (3.5)

113 where the subscript of R^2 and ICC denote the distributional family, here OP-ln for overdispersed

114 Poisson distribution with log link, the term $\ln(1+\omega/\lambda)$ corresponds to the observation-level

115 variance σ^2_ε (Table 1, for derivation see Appendix S1), ω is the overdispersion parameter, and λ is

116 the mean value of λ_{ij}. We discuss how to obtain λ below.

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

118 $$y_{ij} \sim \text{NB}(\lambda_{ij},\ \theta),$$ (3.6)

119 $$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$$ (3.7)

120 $$\alpha_i \sim \text{Gaussian}(0,\ \sigma^2_\alpha),$$ (3.8)

121 where y_{ij} is the jth observation of the ith individual and y_{ij} follows a negative binomial distribution

122 with two parameters, λ_{ij} and θ, where θ is the shape parameter of the negative binomial distribution

123 (given by the software often as the dispersion parameter), and the other symbols are the same as

124 above. $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} for this model can be calculated as:

125 $$R^2_{\text{NB}-\ln(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \ln(1+1/\lambda+1/\theta)},$$ (3.9)

126 $$\text{ICC}_{\text{NB}-\ln} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \ln(1+1/\lambda+1/\theta)},$$ (3.10)

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

128 $$y_{ij} \sim \text{gamma}(\lambda_{ij},\ \nu),$$ (3.11)

129 $$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$$ (3.12)

130 $$\alpha_i \sim \text{Gaussian}(0,\ \sigma^2_\alpha),$$ (3.13)

131   where y_{ij} is the jth observation of the ith individual and y_{ij} follows a gamma distribution with two

132   parameters, λ_{ij} and v, where v is the shape parameter of the gamma distribution (sometimes

133   statistical programs report $1/v$ instead of v; also note that the gamma distribution can be

134   parameterized in alternative ways, Table 1), $R^2_{\text{GLMM}(m)}$ and (adjusted) ICC_{GLMM} can be calculated

135   as:

$$R^2_{\text{gamma}-\ln(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \ln(1 + 1/v)},$$ (3.15)

$$\text{ICC}_{\text{gamma}-\ln} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \ln(1 + 1/v)},$$ (3.16)

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

139   For overdispersed Poisson, negative binomial and gamma GLMMs with log link, the observation-

140   level variance σ^2_ε can be obtained via the variance of the log-normal distribution, as described

141   above (see Appendix S1). There are two more alternative methods to obtain the same target: the

142   delta method and the trigamma function. The two alternatives have different advantages and will be

143   discussed in some detail below.

144   The delta method for variance approximation uses a first order Taylor series expansion, which is

145   often employed to approximate the standard error (error variance) for transformations (or functions)

146   of a variable x when the (error) variance of x itself is known (see [17]; for an accessible reference

147   for biologists, [18]). A simple case of the delta method for variance approximation can be written

148   as:

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

150   where x is a random variable (typically represented by observations), f represents a function (e.g.

151   log or square-root), var denotes variance, and d/dx is a (first) derivative with respect to variable x.

152   Taking derivatives of any function can be easily done using the R environment (examples can be

153   found in the Appendices). It is the delta method that Foulley and colleagues [19] used to derive the

154   distribution specific variance σ_d^2 for Poisson GLMMs as $1/\lambda$: Given that $\mathrm{var}[\lambda_{ij}] = \lambda$ in Poisson

155   distributions and $d\ln(\lambda)/dx = 1/\lambda$, it follows that $\mathrm{var}[\ln(\lambda_{ij})] \approx \lambda(1/\lambda)^2$ (note that for Poisson

156   distributions without overdispersion, σ_d^2 is equal to σ_ε^2 because $\sigma_e^2 = 0$). One clear advantage of

157   the delta method is its flexibility, and we can easily obtain the observation-level variance σ_ε^2 for all

158   kinds of distributions/link functions. For example, by using the delta method, it is straightforward to

159   obtain σ_ε^2 for the Tweedie (compound Poisson-gamma) distribution, which has been used to model

160   non-negative real numbers in ecology (e.g., [20, 21]). For the Tweedie distribution, the variance on

161   the observed scale has the relationship $\mathrm{var}[y] = \varphi\mu^p$ where μ is the mean on the observed scale and

162   φ is the dispersion parameter, comparable to λ and ω in equation (3.1), and p is a positive constant

163   called an index parameter. Therefore, when used with the log-link function, an approximated σ_ε^2

164   value can be obtained by $\varphi\mu^{(p-2)}$ according to equation (4.1). The log-normal approximation

165    $\ln(1 + \varphi\mu^{(p-2)})$ is also possible (see Appendix S1; cf. Table 1).

166   The use of the trigamma function ψ_1 is limited to distributions with log link, but it should provide

167   the most accurate estimate of the observation level variance σ_ε^2. This is because the variance of a

168   gamma-distributed variable on the log scale is equal to $\psi_1(v)$ where v is the shape parameter of the

169   gamma distribution [22] and hence σ_ε^2 is $\psi_1(v)$. At the level of the statistical parameters (Table 1;

170   on the 'expected data' scale; *sensu* [23]; see their Figure 1), Poisson and negative binomial

171   distributions can be both seen special cases of gamma distributions, and σ_ε^2 can be obtained using

172   the trigamma function (Table 1). For example, σ_ε^2 for the Poisson distribution is $\psi_1(\lambda)$ with the

173   speciality that in the case of Poisson distributions $\sigma_\varepsilon^2 = \sigma_d^2$. As we show in Appendix S2, $\ln(1 + 1/\lambda)$

174   (log-normal approximation), $1/\lambda$ (delta method approximation) and $\psi_1(\lambda)$ (trigamma function) are

175   similar if λ is greater than 2. Nonetheless, our recommendation is to use the trigamma function for

176   obtaining σ_{ε}^2 whenever this is possible.

177   We note that in calculations of heritability (which can be seen as a type of ICC although in a strict

178   sense, it is not; see [23]) using negative binomial GLMMs, the trigamma function has been

179   previously used to obtain observation-level variance ([22, 24]; cf. [23]). Table 1 summarises

180   observation-level variance σ_{ε}^2 for overdispersed Poisson, negative binomial and gamma

181   distributions for commonly used link functions.

5. How to estimate λ from data

183   Imagine a Poisson GLMM with log link and additive overdispersion fitted as an observation-level

184   random effect (model 5):

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

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

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

188     $e_{ij} \sim \text{Gaussian}(0, \sigma_{e}^2)$,   (5.4)

189   where y_{ij} is the jth observation of the ith individual, and follows a Poisson distribution with the

190   parameter λ_{ij}, e_{ij} is an additive overdispersion term for jth observation of the ith individual, and the

191   other symbols are the same as above. Using the log-normal approximation $R^2_{\text{GLMM}(m)}$ and (adjusted)

192   ICC_{GLMM} can be calculated as:

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

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

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

196   link (Table 1).

197   In our earlier papers, we proposed to use the exponential of the intercept (from the intercept-only

198   model or models with centred fixed factors) $\exp(\beta_0)$ as an estimator of λ [2, 3]. We also suggested

199   that it is possible to use the mean of observed values y_{ij}. Unfortunately, these two recommendations

200   are often inconsistent with each other. This is because, given the model 5 (and all the models in the

201   previous section), the following relationships hold:

202    $$\exp(\beta_0) \le \mathrm{E}[y_{ij}]_,$$ (5.7)

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

204    $$\mathrm{E}[y_{ij}] = \mathrm{E}[\lambda_{ij}]_,$$ (5.9)

205   where E represents the expected value (i.e., mean) on the observed scale, β_0 is the mean value on

206   the latent scale (i.e. β_0 from the intercept-only model), σ_τ^2 is the total variance on the latent scale

207   (e.g., $\sigma_\alpha^2 + \sigma_e^2$ in the models 1and 5, and σ_α^2 in models 2-4[2]; see also [25]). In fact, $\exp(\beta_0)$ gives

208   the median value of y_{ij} rather than the mean of y_{ij}, assuming a Poisson distribution. Thus, the use of

209    $\exp(\beta_0)$ will often overestimate σ_d^2, providing conservative (smaller) estimates of R^2 and ICC,

210   compared to when using averaged y_{ij}, which is a better estimate of $\mathrm{E}[y_{ij}]$. Quantitative differences

211   between the two approaches may often be negligible, but when λ is small, the difference can be

212   substantial so the choice of the method needs to be reported for reproducibility (Appendix S2). Our

213   new recommendation is to obtain λ via equation (5.8). When sampling is balanced (i.e. observations

214   are equally distributed across individuals and covariates), equation (5.8) and the mean of the

215   observed values will give similar values, but when unbalanced, method equation (5.8) is preferable.

216   This recommendation for obtaining λ also applies to negative binomial GLMMs (see Table 1).

11

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

217

218 A general form of equation (5.7) is known as Jensen's inequality, $g(\overline{x}) \leq \overline{g(x)}$ where g is a convex

219 function. Hence, the transformation of the mean value is equal to or larger than the mean of

220 transformed values (the opposite is true for a concave function; that is, $g(\overline{x}) \geq \overline{g(x)}$; [26]). In fact,

221 whenever the function is not strictly linear, simple application of the inverse link function (or back-

222 transformation) cannot be used to translate the mean on the latent scale into the mean value on the

223 observed scale. This inequality has important implications for the interpretation of results from

224 GLMMs, and also generalized linear models GLMs and linear models with transformed response

225 variables.

226 Although log-link GLMMs (e.g., model 5) have an analytical formula, equation (5.8), this is not

227 usually the case. Therefore, converting the latent scale values into observation-scale values requires

228 simulation using the inverse link function. However, the delta method for bias correction can be

229 used as a general approximation to account for Jensen's inequality when using link functions or

230 transformations. This application of the delta method uses a second order Taylor series expansion

231 [17, 27]. A simple case of the delta method for bias correction can be written as:

232 $$E[f(x)] \approx f(x) + 0.5\sigma_\tau^2 \frac{d^2}{dx^2} f(x),$$ (6.1)

233 where d^2/dx^2 is a second derivative with respect to the variable x and the other symbols are as in

234 equations (4.1) and (5.8). By employing this bias correction delta method (with

235 $d^2 \exp(x)/dx^2 = \exp(x)$), we can approximate equation (5.8) using the same symbols as in

236 equations (5.7)-(5.9):

237 $$E[\lambda_{ij}] = E[\exp(\beta_0)] \approx \exp(\beta_0) + 0.5\sigma_\tau^2 \exp(\beta_0)$$ (6.2)

12

238   The comparison between equation (5.8) (exact) and equation (6.2) (approximate) is shown in

239   Appendix S3. The approximation is most useful when the exact formula is not available as in the

240   case of a binomial GLMM with logit link (model 6):

241    $y_{ij} \sim \text{binomial}(p_{ij},\ n_{ij})$,   (6.3)

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

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

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

245   where y_{ij} is the number of 'success' in n_{ij} trials by the ith individual at the jth occasion (for binary

246   data, n_{ij} is always 1), p_{ij} is the underlying probability of success, and the other symbols are the same

247   as above.

248   To obtain corresponding values between the latent scale and data (observation) scale, we need to

249   account for Jensen's inequality (note the logit function combines of concave and convex sections).

250   For example, the overall intercept, β_0 on the latent scale could be transformed not just with the

251   inverse (anti) logit function ($\text{logit}^{-1}(x) = \exp(x)/(1+\exp(x))$) but also the bias corrected

252   approximation. For the case of the binomial GLMM, we can use this approximation below given

253   that $d^2\text{logit}^{-1}(x)/dx^2 = \exp(x)(1-\exp(x))/(1+\exp(x))^3$:

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

255   We can replace β_0 with any value obtained from the fixed part of the model (i.e. $\beta_0 + \sum \beta_h x_{hij}$).

256   Another approximation proposed by Zeger and colleagues [28] produces similar (but slightly better)

257   estimates than equation (6.7). Using our notation, this approximation can be written as:

13

258 $$E[p_{ij}] \approx \text{logit}^{-1}\left(\beta_0 \sqrt{1 + \left(\frac{16\sqrt{3}}{15\pi}\right)^2 \sigma_\tau^2}^{-1} \right).$$ (6.8)

259 A comparison between equations (6.7) and (6.8) is also shown in Appendix S3. This approximation

260 uses the exact solution for the inverse probit function, which can be written for a model like model

261 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):

262 $$E[p_{ij}] = \text{probit}^{-1}\left(\beta_0 \sqrt{1 + \sigma_\tau^2}^{-1} \right).$$ (6.9)

263 Simulation will give the most accurate conversions when no exact solutions are available. The use

264 of the delta method for bias correction accounting for Jensen's inequity is a very general and

265 versatile approach that is applicable for any distribution with any link function (see Appendix S3)

266 and can save computation time. We note that the accuracy of the delta method (both variance

267 approximation and bias correction) depends on the form of the function f, the conditions for and

268 limitation of the delta method are described in the article by Oehlert [27].

269 ## 7. Special considerations for binomial GLMMs

270 The observation-level variance σ_ε^2 can be thought of as being added to the latent scale on which

271 other variance components are also estimated in a GLMM (equations (3.2), (3.7), (3.12), (5.2) and

272 (6.4) for models 2-6). Since the proposed R^2_{GLMM} and ICC $_{\text{GLMM}}$ are ratios between variance

273 components and their additive combinations, we can show using the delta method that R^2_{GLMM} and

274 ICC $_{\text{GLMM}}$ calculated via σ_ε^2 approximate to those of R^2 and ICC on the observation (original) scale

275 (shown in Appendix S4). In some cases, there exist specific formulas for ICC on the observation

276 scale [2]. In the past, we distinguished between ICC on the latent scale and on the observation scale

277 [2]. Such a distinction turns out to be strictly appropriate only for binomial distributions but not for

278 Poisson distributions (and probably also not for other non-Gaussian distributions). This is because

14

279   the property of what we have called the distribution-specific variance σ_d^2 for binomial distributions

280   (e.g. $\pi^2/3$ for binomial error distribution with the logit link function) is quite different from what we

281   have discussed as the observation-level variance σ_ε^2 although these two types of variance are

282   related conceptually (i.e., both represents variance due to non-Gaussian distributions with specific

283   link functions). Let us explain this further.

284   A binomial distribution with a mean of p (the proportion of successes) has a variance of $p(1-p)$ and

285   we find that the observation-level variance is $1/(p(1-p))$ using the delta method on the logit-link

286   function (see Table 2). This observation-level variance $1/(p(1-p))$ is clearly different from the

287   distribution-specific variance $\pi^2/3$. As with the observation-level variance for the log-Poisson model

288   (which is $1/\lambda$ and changes with λ; note that we would have called $1/\lambda$ the distribution-specific

289   variance; [2, 3]), the observation-level variance of the binomial distribution changes as p changes

290   (see Appendix S5), suggesting these two observation-level variances ($1/\lambda$ and $1/(p(1-p))$) are

291   analogous while the distribution-specific variance $\pi^2/3$ is not. Further, the minimum value of

292    $1/(p(1-p))$ 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 will

293   always produce larger values than those using $\pi^2/3$. Consequently, Browne and colleagues [15]

294   showed that ICC values (or variance partition coefficients, VPCs) estimated using $\pi^2/3$ were higher

295   than corresponding ICC values on the observation (original) scale using logistic-binomial GLMMs

296   (see also [29]). Then, what is $\pi^2/3$?

297   Three common link functions in binomial GLMMs (logit, probit and complementary log-log) all

298   have corresponding distributions on the latent scale: the logistic distribution, standard normal

299   distribution and Gumbel distribution, respectively. Each of these distributions has a theoretical

300   variance, namely, $\pi^2/3$, 1 and $\pi^2/6$, respectively (Table 2). As far as we are aware, these theoretical

301   variances only exist for binomial distributions. It is important to notice that, for example, the

302   meaning of $1/(p(1-p))$, which is the variance on the latent scale that approximates to the variance

303   due to binomial distributions on the observation scale is distinct from the meaning of $\pi^2/3$, which is

304 the variance of the latent distribution (i.e., the logistic distribution) according to which the original

305 data are theoretically distributed on the logit scale. We need distinguishing these theoretical

306 (distribution-specific) variances from the observation-level variance. Put another way, R^2 and ICC

307 values using the theoretical distribution-specific variance can rightly be called the latent (link) scale

308 (*sensu* [2]) while, as mentioned above, R^2 and ICC values using the observation-level variance

309 estimate the counterparts on the observation (original) scale (cf. [23]). The use of the theoretical

310 distribution-specific variance will almost always provide different values of R^2_{GLMM} and ICC $_{GLMM}$

311 from those using the observation-level obtained via the delta method (see Appendix S5). In any

312 case, we should be aware that binomial GLMMS are special cases for obtaining R^2_{GLMM} and ICC

313 $_{GLMM}$ from binomial GLMMs.

8. Worked examples: revisting the beetles

315 In the following, we present a worked example by expanding the beetle dataset that was generated

316 for the previous work [3]. In brief, the dataset represents a hypothetical species of beetle that has the

317 following life cycle: larvae hatch and grow in the soil until they pupate, and then adult beetles feed

318 and mate on plants. Larvae are sampled from 12 different populations ('Population'; see Figure 1).

319 Within each population, larvae are collected at two different microhabitats ('Habitat'): dry and wet

320 areas as determined by soil moisture. Larvae are exposed to two different dietary treatments

321 ('Treatment'): nutrient rich and control. The species is sexually dimorphic and can be easily sexed

322 at the pupa stage ('Sex'). Male beetles have two different color morphs: one dark and the other

323 reddish brown ('Morph', labeled as A and B in Figure 1). Sexed pupae are housed in standard

324 containers until they mature ('Container'). Each container holds eight same-sex animals from a

325 single population, but with a mix of individuals from the two habitats ($N_{[container]} = 120$; $N_{[animal]} =$

326 960).

327 We have data on the five phenotypes, two of them sex-limited: (i) the number of eggs laid by each

328 female after random mating which we had generated previously using Poisson distributions (with

16

329   additive dispersion) and we revisit here for analysis with quasi-Poisson models (i.e. multiplicative

330   dispersion), (ii) the incidence of endo-parasitic infections that we generated as being negative

331   binomial distributed, (iii) body length of adult beetles which we had generated previously using

332   Gaussian distributions and that we revisit here for analysis with gamma distributions, (iv) time to

333   visit five predefined sectors of an arena (employed as a measure of exploratory tendencies) that we

334   generated as being gamma distributed, and (v) the two male morphs, which was again generated

335   with binomial distributions. We will use this simulated dataset to estimate R^2_{GLMM} and ICC $_{\text{GLMM}}$.

336   All data generation and analyses were conducted in R 3.3.1 [10]. We used functions to fit GLMMs

337   from the three R packages: 1) the *glmmadmb* function from glmmADMB [30], 2) the *glmmPQL*

338   function from MASS [31], and 3) the *glmer* and *glmer.nb* functions from lme4 [32]. In Table 1, we

339   only report results from *glmmADMB* because this is the only function that can fit models with all

340   relevant distributional families. All scripts and results are provided as an electronic supplement

341   (Appendix S6). In addition, Appendix S6 includes an example of a model using the Tweedie

342   distribution, which was fitted by the *cpglmm* function from the cplm package [21]. Notably, our

343   approach for R^2_{GLMM} is kindly being implemented in the *rsquared* function in the R package,

344   piecewiseSEM [33]. Another important note is that we often find less congruence in GLMM results

345   from the different packages than those of linear mixed-effects models, LMM. Thus, it is

346   recommended to run GLMMs in more than one package to check robustness of the results although

347   this may not always be possible.

348   In all the models, estimated regression coefficients and variance components are very much in

349   agreement with what is expected from our parameter settings (Table 1 and Appendix S6). When

350   comparing the null and full models, which had 'sex' as a predictor, the magnitudes of the variance

351   component for the container effect always decrease in the full models. This is because the variance

352   due to sex is confounded with the container variance in the null model. As expected, (unadjusted)

353   ICC values from the null models are usually smaller than adjusted ICC values from the full models

354   because the observation-level variance (analogous to the residual variance) was smaller in the full

17

355 models, implying that the denominator of equation (3.2) shrinks. However, the numerator also

356 becomes smaller for ICC values for the container effect from the parasite, size and exploration

357 models so that adjusted ICC values are not necessarily larger than unadjusted ICC values.

358 Accordingly, adjusted $ICC_{[container]}$ is smaller in the parasite and size models but not in the

359 exploration model. The last thing to note is that for the morph models (binomial mixed models),

360 both R^2 and ICC $_{values}$ are larger when using the distribution-specific variance rather than the

361 observation-level variance, as discussed above (Table 3; also see Appendix S4).

9. Alternatives and a cautonary note

363 Here we extended our simple methods for obtaining R^2_{GLMM} and ICC $_{GLMM}$ for Poisson and

364 binomial GLMMs to other types of GLMMs such as negative binomial and gamma. We have

365 described three different ways of obtaining the observational-level variance and how to obtain the

366 key rate parameter λ for Poisson and negative binomial distributions. We discussed important

367 considerations which arise for estimating R^2_{GLMM} and ICC $_{GLMM}$ with binomial GLMMs. As we

368 have shown, the merit of our approach is not only its ease of implementation but also that our

369 approach encourages researchers to pay more attention to variance components at different levels.

370 Research papers in the field of ecology and evolution often report only regression coefficients but

371 not variance components of GLMMs [3].

372 We would like to highlight two recent studies that provide alternatives to our approach. First, Jaeger

373 and colleagues [5] have proposed R^2 for fixed effects in GLMMs, which they referred to as $R^2_{\beta*}$ (an

374 extension of an R^2 for fixed effects in linear mixed models or R^2_{β} by Edwards and colleagues [34]).

375 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

376 distribution (error structure) with any link function. Jaeger and colleagues highlight that in the

377 framework of $R^2_{\beta*}$, they can easily obtain semi-partial R^2, which quantifies the relative importance

378 of each predictor (fixed effect). As they demonstrate by simulation, their method potentially gives a

379 very reliable tool for model selection. One current issue for this approach is that implementation

380   does not seem as simple as our approach. We note that our R^2_{GLMM} framework could also provide

381   semi-partial R^2 via commonality analysis (see [35]; note that unique variance for each predictor in

382   commonality analysis corresponds to semi-partial $R^{2;}$; [36]).

383   Second, de Villemereuil and colleagues [23] provided a framework with which one can estimate

384   exact heritability using GLMMs at different scales (e.g. data and latent scales). Their method can be

385   extended to obtain exact ICC values on the data (observation) scale, which is analogous to, but not

386   the same as, our ICC $_{GLMM}$ using the observation-level variance, σ^2_ε described above. Further, this

387   method can, in theory, be extended to estimate R^2_{GLMM} on the data (observation) scale. One

388   potential difficulty is that the method of de Villemereuli and colleagues. is exact but that a

389   numerical method is used to solve relevant equations so one will require a software package (e.g.,

390   the QGglmm package; [23]).

391   Finally, we finish by repeating what we said at the end of our original R^2 paper [3]. Both R^2 and

392   ICC are indices that are likely to reflect only one or a few aspects of a model fit to the data and

393   should not be used for gauging the quality of a model. We encourage biologists use R^2 and ICC in

394   conjunctions with other indices like information criteria (e.g. AIC, BIC and DIC), and more

395   importantly, with model diagnostics such as checking for model assumptions, heteroscedasticity

396   and sensitivity to outliers.

Authors' contribuions

397

398   SN conceived ideas and conducted analysis. Both developed the ideas further, and contributed to

399   writing and editing of the manuscript.

Competing interests

400

401   We have not competing interests

19

Funding

Acknowledgements

409

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. 2016 An R^2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, 10.1080/02664763.02662016.01193725. (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: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.

[11] Snijders, T. & Bosker, R. 1999 *Multilevel Analysis: an Introduction to basic and advanced multilevel modeling*. London, Sage.

21

435   [12] Snijders, T. & Bosker, R. 2011 *Multilevel Analysis: an Introduction to basic and advanced*

436   *multilevel modeling*. 2nd ed. London, Sage.

437   [13] Harrison, X.A. 2014 Using observation-level random effects to model overdispersion in count

438   data in ecology and evolution. *Peerj* **2**, e616. (doi:ARTN e61610.7717/peerj.616).

439   [14] Harrison, X.A. 2015 A comparison of observation-level random effect and Beta-Binomial

440   models for modelling overdispersion in Binomial data in ecology & evolution. *Peerj* **3**, e1114.

441   (doi:ARTN e111410.7717/peerj.1114).

442   [15] Browne, W.J., Subramanian, S.V., Jones, K. & Goldstein, H. 2005 Variance partitioning in

443   multilevel logistic models that exhibit overdispersion. *J R Stat Soc a Stat* **168**, 599-613.

444   [16] Gelman, A. & Hill, J. 2006 *Data analysis using regression and multilevel/hierarchical models*

445   Cambridge, Cambridge University Press.

446   [17] Ver Hoef, J.M. 2012 Who Invented the Delta Method? *Am Stat* **66**, 124-127.

447   (doi:10.1080/00031305.2012.687494).

448   [18] Powell, L.A. 2007 Approximating variance of demographic parameters using the delta method:

449   A reference for avian biologists. *Condor* **109**, 949-954. (doi:Doi 10.1650/0010-

450   5422(2007)109[949:Avodpu]2.0.Co;2).

451   [19] Foulley, J.L., Gianola, D. & Im, S. 1987 Genetic Evaluation of Traits Distributed as Poisson-

452   Binomial with Reference to Reproductive Characters. *Theor. Appl. Genet.* **73**, 870-877. (doi:Doi

453   10.1007/Bf00289392).

454   [20] Foster, S.D. & Bravington, M.V. 2013 A Poisson-Gamma model for analysis of ecological

455   non-negative continuous data. *Environ Ecol Stat* **20**, 533-552. (doi:10.1007/s10651-012-0233-0).

456   [21] Zhang, Y.W. 2013 Likelihood-based and Bayesian methods for Tweedie compound Poisson

457   linear mixed models. *Stat Comput* **23**, 743-757. (doi:10.1007/s11222-012-9343-7).

458   [22] Tempelman, R.J. & Gianola, D. 1999 Genetic analysis of fertility in dairy cattle using negative

459   binomial mixed models. *J. Dairy Sci.* **82**, 1834-1847.

460   [23] de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. 2016 General Methods for

461   Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. *Genetics* **204**, 1281-

462   1294. (doi:10.1534/genetics.115.186536).

463   [24] Matos, C.A.P., Thomas, D.L., Gianola, D., Tempelman, R.J. & Young, L.D. 1997 Genetic

464   analysis of discrete reproductive traits in sheep using linear and nonlinear models .1. Estimation of

465   genetic parameters. *J. Anim. Sci.* **75**, 76-87.

466   [25] Carrasco, J.L. 2010 A generalized concordance correlation coefficient based on the variance

467   components generalized linear mixed models for overdispersed count data. *Biometrics* **66**, 897-904.

468   (doi:Doi 10.1111/J.1541-0420.2009.01335.X).

469   [26] Rao, C.R. 2002 *Linear statistical inference and its applications*. 2nd ed. ed. New York, John

470   Wiley & Sons.

471   [27] Oehlert, G.W. 1992 A note on the delta method. *Am Stat* **46**, 27-29. (doi:Doi

472   10.2307/2684406).

473   [28] Zeger, S.L., Liang, K.Y. & Albert, P.S. 1988 Models for Longitudinal Data - a Generalized

474   Estimating Equation Approach. *Biometrics* **44**, 1049-1060. (doi:Doi 10.2307/2531734).

475   [29] Goldstein, H., Browne, W. & Rasbash, J. 2002 Partitioning variation in multilevel models.

476   *Understanding Statistics* **1**, 223-231.

477   [30] Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen,

478   A. & Sibert, J. 2012 AD Model Builder: using automatic differentiation for statistical inference of

479   highly parameterized complex nonlinear models. *Optim Method Softw* **27**, 233-249.

480   (doi:10.1080/10556788.2011.597854).

481   [31] Venables, W.N. & Ripley, B.D. 2002 *Modern applied statistics with S*. 4 ed. New York,

482   Springer.

483   [32] Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. 2015 Fitting Linear Mixed-Effects

484   Models Using lme4. *J Stat Softw* **67**, 1-48.

485   [33] Lefcheck, J.S. 2016 PIECEWISESEM: Piecewise structural equation modelling in R for

486   ecology, evolution, and systematics. *Methods Ecol Evol* **7**, 573-579. (doi:10.1111/2041-

487   210x.12512).

488   [34] Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F. & Schabenberger, O. 2008 An R^2

489   statistic for fixed effects in the linear mixed model. *Stat Med* **27**, 6137-6157. (doi:Doi

490   10.1002/Sim.3429).

491   [35] Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. 2014

492   Using commonality analysis in multiple regressions: a tool to decompose regression effects in the

493   face of multicollinearity. *Methods Ecol Evol* **5**, 320-328. (doi:Doi 10.1111/2041-210x.12166).

494   [36] Nimon, K.F. & Oswald, F.L. 2013 Understanding the Results of Multiple Linear Regression:

495   Beyond Standardized Regression Coefficients. *Organ Res Methods* **16**, 650-674.

496   (doi:10.1177/1094428113493929).

497

498

499

500

501

Table 1. The observation-level variance σ_ε^2 for the three distributional families: quasi-Poisson (overdispersed Poisson), negative binomial and gamma with the three different methods for deriving σ_ε^2: the delta method, long-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 (OP: overdispersed Poisson)	$OP(\lambda, \omega)$	$E[y] = \lambda$	log	$\dfrac{\omega}{\lambda}$	$\ln\left(1 + \dfrac{\omega}{\lambda}\right)$	$\psi_1\left(\dfrac{\lambda}{\omega}\right)$
Poisson (when $\omega = 1$)	$\lambda > 0$ $\omega > 0$	$\text{var}[y] = \lambda\omega$	square-root	0.25ω	-	
Negative binomial (NB)	$NB(\lambda, \theta)$	$E[y] = \lambda$	log	$\dfrac{1}{\lambda} + \dfrac{1}{\theta}$	$\ln\left(1 + \dfrac{1}{\lambda} + \dfrac{1}{\theta}\right)$	$\psi_1\left(\left[\dfrac{1}{\lambda} + \dfrac{1}{\theta}\right]^{-1}\right)$

25

$\lambda > 0$ $\theta > 0$		$\mathrm{var}[y] = \lambda + \dfrac{\lambda^2}{\theta}$	square-root	$0.25\left(1 + \dfrac{\lambda}{\theta}\right)$	-	
Gamma	$\mathrm{gmma}(\lambda,\ \nu)$	$E[y] = \lambda$	log	$\dfrac{1}{\nu}$	$\ln\left(1 + \dfrac{1}{\nu}\right)$	$\psi_1(\nu)$
$\lambda > 0$ $\nu > 0$		$\mathrm{var}[y] = \dfrac{\lambda^2}{\nu}$	inverse (reciprocal)	$\dfrac{1}{\nu\lambda^2}$	-	
Gamma (alternative parameterization)	$\mathrm{gamma}(\nu,\ \kappa)$	$E[y] = \dfrac{\nu}{\kappa}$	log	$\dfrac{1}{\nu}$	$\ln\left(1 + \dfrac{1}{\nu}\right)$	$\psi_1(\nu)$
$\nu > 0$ $\kappa > 0$		$\mathrm{var}[y] = \dfrac{\nu}{\kappa^2}$	inverse (reciprocal)	$\dfrac{\kappa^2}{\nu^3}$	-	

$\mathrm{var}[\ln(x)] = \psi_1(\nu) = \sum_{n=1}^{\infty} 1/(\nu+n)$ when x follows gamma distribution. In the R environment, the function, *trigamma* can be used to obtain $\psi_1(\nu)$.

27

Table 2. The distribution-specific variance σ_d^2 and observation-level variance σ_ε^2 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	Distribution-specific variance	Observation-level variance using the delta method (min. values and corresponding p)
Binomial (Bernoulli; $n = 1$)	binomial(p, n) $0 < p < 1$ $n >= 1$ (integers)	logit	$\ln\left(\dfrac{p}{1-p}\right)$	$\dfrac{\pi^2}{3} \sim 3.29$ (logistic distribution)	$\dfrac{1}{p(1-p)}$ (min = 4; $p = 0.5$)
	$E[y] = np$ var$[y] = np(1-p)$	probit ($\Phi(p)$)	$\sqrt{2}\,\mathrm{erf}^{-1}(2p-1)$	1 (standard normal distribution)	$2\pi p(1-p)\left(\exp\left[\mathrm{erf}^{-1}(2p-1)\right]^2\right)^2$ (min $\sim$ 1.57; $p = 0.5$)

cloglog

(complimentary

log-log)

$\ln(-\ln(1-p))$

$$\frac{\pi^2}{6} \sim 1.65$$

(Gumbel distribution)

$$\frac{p}{\left(\ln(1-p)\right)^2 (1-p)}$$

(min ~ 1.54; $p \sim 0.8$;

~ 2.08; $p = 0.5$)

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

Table 3. 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 $N_{[animal]}=960$.

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]
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	-	-

30

				[-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
$R^2_{\mathrm{GLMM}(m)}$	-	10.01%	-	48.83%	-	49.54%	-	78.34%	-	3.98% (4.57%)
$R^2_{\mathrm{GLMM}(c)}$	-	47.19%	-	86.91%	-	72.52%	-	93.34%	3.98% (4.57%)	27.46% (31.55%)
ICC[Population]	25.50%	31.47%	11.62%	34.89%	11.38%	33.17%	3.40%	26.94%	19.49% (22.63%;)	20.96% (24.23%)

ICC[Container]	5.98%	9.84%	61.30%	39.53%	59.57%	12.37%	17.34%	42.34%	2.67% (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 distribution-specific variance were used; note that ones in brackets use the distribution-specific variance for R^2 and ICC.

ICC[Container] is not a typical 'repeatability' but the proportion of variance due to the container effect beyond the population variance.

Figure legends

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