*An article submitted to Methods in Ecology and Evolution*

# Coefficient of determination $R^2$ and intra-class correlation coefficient ICC from generalized linear mixed-effects models revisited and expanded

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   **Abstract**

2    1. The coefficient of determination $R^2$ quantifies the proportion of variance explained by a

3       statistical model and is an important summary statistic of biological interest. However,

4       estimating $R^2$ for (generalized) linear mixed models (GLMMs) remains challenging. We

5       have previously introduced a version of $R^2$ that we called $R^2_{GLMM}$ for Poisson and binomial

6       GLMMs, but not for other distributional families.

7    2. Similarly, we earlier discussed how to estimate intra-class correlation coefficients ICC (also

8       known as repeatability in the field of ecology and evolution) using Poisson and binomial

9       GLMMs, but not for other distributional families. ICC is related to $R^2$ because they are both

10       ratios of variance components.

11    3. In this article we expand our method to additional non-Gaussian distributions, namely quasi-

12       Poisson, negative binomial and gamma GLMMs. However, in theory, our extension could

13       be applied to any distribution and we include an explanatory calculation for the Tweedie

14       distribution.

15    4. While expanding our approach, we highlight two useful concepts, Jensen's inequality and

16       the delta method, both of which help in understanding the properties of GLMMs. Jensen's

17       inequality has important implications for the interpretation GLMMs while the delta method

18       allows a general derivation of distribution-specific variances. We also discuss some special

19       considerations for binomial GLMMs with binary or proportion data.

20    5. We illustrate the implementation of our extension by worked examples in the R

21       environment. However, our method can be used regardless of statistical packages and

22       environments. We finish by referring to two alternative methods to our approach along with

23       a cautionary note.

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

25       *variance explained, intra-class correlation, model fit, variance decomposition, reliability*

26       *analysis.*

## Introduction

28  One of the main purposes of linear modelling is to understand the sources of variation in biological

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

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

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

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

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

34  individuals that have been phenotyped repeatedly (Lessells and Boag 1987, Nakagawa and

35  Schielzeth 2010). We have reviewed methods for estimating $R^2$ and ICC in the past (Nakagawa and

36  Schielzeth 2010, 2013), with a particular focus on non-Gaussian response variables, featuring

37  generalized linear mixed-effects models (GLMMs) as the most versatile engine for estimating $R^2$

38  and ICC (specifically $R^2_{GLMM}$ and $ICC_{GLMM}$). Our descriptions were limited to random-intercept

39  GLMMs, but Johnson (2014) has recently extended the methods to random-slope GLMMs,

40  widening the applicability of these statistics (see also, LaHuis *et al.* 2014; Jaeger et al. 2016).

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

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

43  most popular (Bolker *et al.* 2009), there are other commonly used families for GLMMs, such as

44  negative binomial and gamma distributions (Ver Hoef and Boveng 2007, Bolker 2008). In this

45  article, we revisit and extend $R^2_{GLMM}$ and $ICC_{GLMM}$ to more distributional families, in particular to

46  negative binomial and gamma distributions. In this context we discuss Jensen's inequality and two

47  variants of the delta method, which are useful not only for extending our method, but also for

48  interpreting the results of GLMMs in general. Furthermore, we refer to some special considerations

49  when obtaining $R^2_{GLMM}$ and $ICC_{GLMM}$ from binomially GLMMs for binary and proportion data,

50  which we did not discuss in the past (Nakagawa and Schielzeth 2010, 2013). We provide worked

51  examples focusing on implementation in the R environment (R Core Team 2016) and finish by

52    referring to two alternative approaches for obtaining $R^2$ and ICC from GLMMs along with a

53    cautionary note.

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

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

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

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

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

59    (model 1):

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

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

62    $$\varepsilon_{ij} \sim \text{Gaussian}(0,\ \sigma_\varepsilon^2),$$     eqn 3

63    where $y_{ij}$ is the $j$th observation of the $i$th individual, $x_{hij}$ is the $j$th value of the $i$th individual for the

64    $h$th of $k$ fixed effects predictors, $\beta_0$ is the (grand) intercept, $\beta_h$ is the regression coefficient for the

65    $h$th predictor, $\alpha_i$ is an individual-specific effect, assumed to be normally distributed in the

66    population with the mean and variance of 0 and $\sigma_\alpha^2$, $\varepsilon_{ij}$ is an observation-specific residual, assumed

67    to be normally distributed in the population with mean and variance of 0 and $\sigma_\varepsilon^2$, respectively. For

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

69    $$R^2_{LMM(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2},$$     eqn 4

70    $$R^2_{LMM(c)} = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2},$$     eqn 5

71    $$\sigma_f^2 = \text{var}\left(\sum_h^k \beta_h x_{hij}\right),$$     eqn 6

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

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

74    effects, and $\sigma^2_f$ is the variance explained by fixed effects (Snijders and Bosker 1999, 2011). Since

75    marginal and conditional $R^2$ differ only in whether the random effect variance is included in the

76    numerator, we avoid redundancy and present equations only for marginal $R^2$ in the following.

77    Similarly, there are two types of ICC:

78    $$\text{ICC}_{\text{LMM(adj)}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\varepsilon} \qquad\qquad \text{eqn 7}$$

79    $$\text{ICC}_{\text{LMM}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_f + \sigma^2_\varepsilon} \qquad\qquad \text{eqn 8}$$

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

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

82    ICC (*sensu* Nakagawa and Schielzeth 2010). Since the two versions of ICC differ only in whether

83    the fixed effect variance (calculated as in Equation 6) is included in the denominator, we avoid

84    redundancy and present equations only for adjusted ICC in the following.

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

86    variance $\sigma^2_\varepsilon$. For binomial and Poisson GLMMs with an additive dispersion terms, we have

87    previously stated that $\sigma^2_\varepsilon$ is equivalent to $\sigma^2_e + \sigma^2_d$ where $\sigma^2_e$ is the variance for the additive

88    overdispersion term, and $\sigma^2_d$ is the distribution-specific variance (Nakagawa and Schielzeth 2010,

89    2013). Here overdispersion represents the excess variation relative to what is expected from a

90    certain distribution and can be estimated by fitting an observation-level random effect (OLRE; see

91    Harrison 2014, 2015). Alternatively, overdispersion in GLMMs can be implemented using a

92    multiplicative overdispersion term (Browne *et al.* 2005). In such an implementation, we stated that

93    $\sigma^2_\varepsilon$ is equivalent to $\omega \cdot \sigma^2_d$ where $\omega$ is a multiplicative dispersion parameter estimated from the

94    model (Nakagawa and Schielzeth 2010). But obtaining $\sigma^2_d$ for specific distributions is not always

5

95     possible, because in many families of GLMMs the parameters are less clearly separated into a

96     parameter for the expectation of the mean and a parameter for the (over)dispersion. It turns out that

97     binomial and Poisson distributions are special cases where $\sigma_d^2$ can be usefully calculated, because

98     either all overdispersion is modelled by an OLRE (additive overdispersion) or by a single

99     multiplicative overdispersion parameter (multiplicative overdispersion). However, as we will show

100    below, we can always obtain the GLMM version of $\sigma_\varepsilon^2$ (on the latent scale) directly. We refer to

101    this generalised version of $\sigma_\varepsilon^2$ as 'the observation-level variance' here rather than the residual

102    variance (but we keep the notation $\sigma_\varepsilon^2$).


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

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

105    GLMM, because the overdispersed Poisson distribution can be considered a re-parameterization of

106    the negative binomial distribution (Gelman and Hill 2007; Ver Hoef and Boveng 2007). Imagine

107    count data repeatedly measured from a number of individuals with associated data on $k$ covariates.

108    We fit an overdispersed Poisson (OP) GLMM with the log link function (model 2):

109    $y_{ij} \sim \text{OP}(\lambda_{ij},\ \omega),$                                      eqn 9

110    $\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$                   eqn 10

111    $\alpha_i \sim \text{Gaussian}(0,\ \sigma_\alpha^2),$                           eqn 11

112    where $y_{ij}$ is the $j$th observation of the $i$th individual and $y_{ij}$ follows an overdispersed Poisson

113    distribution with two parameters, $\lambda_{ij}$ and $\omega$, $\ln(\lambda_{ij})$ is the latent value for the $j$th observation of the $i$th

114    individual, $\omega$ is the overdispersion parameter (when the multiplicative dispersion parameter $\omega$ is 1,

115    the model becomes a standard Poisson GLMM), $\alpha_i$ is an individual-specific effect, assumed to be

116    normally distributed in the population with the mean and variance of 0 and $\sigma_\alpha^2$, respectively (as in

6

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

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

119    $$R^2_{\text{OP}-\ln(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \ln(1 + \omega / \lambda)},$$        eqn 12

120    $$\text{ICC}_{\text{OP}-\ln} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \ln(1 + \omega / \lambda)},$$        eqn 13

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

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

123    variance $\sigma^2_\varepsilon$ (Table 1, for derivation see Appendix S1), $\omega$ is the overdispersion parameter, and $\lambda$ is

124    the mean value of $\lambda_{ij}$. We discuss how to obtain $\lambda$ below.

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

126    $$y_{ij} \sim \text{NB}(\lambda_{ij}, \ \theta),$$        eqn 14

127    $$\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$$        eqn 15

128    $$\alpha_i \sim \text{Gaussian}(0, \ \sigma^2_\alpha),$$        eqn 16

129    where $y_{ij}$ is the $j$th observation of the $i$th individual and $y_{ij}$ follows a negative binomial distribution

130    with two parameters, $\lambda_{ij}$ and $\theta$, where $\theta$ is the shape parameter of the negative binomial distribution

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

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

133    $$R^2_{\text{NB}-\ln(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \ln(1 + 1/\lambda + 1/\theta)},$$        eqn 17

134    $$\text{ICC}_{\text{NB}-\ln} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \ln(1 + 1/\lambda + 1/\theta)},$$        eqn 18

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

136    $$y_{ij} \sim \text{gamma}(\lambda_{ij}, \ \nu),$$        eqn 19

7

137     $\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i,$        eqn 20

138     $\alpha_i \sim \text{Gaussian}(0, \ \sigma_\alpha^2),$        eqn 21

139     where $y_{ij}$ is the $j$th observation of the $i$th individual and $y_{ij}$ follows a gamma distribution with two

140     parameters, $\lambda_{ij}$ and $v$, where $v$ is the shape parameter of the gamma distribution (sometimes

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

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

143     as:

144     $R^2_{\text{gamma}-\ln(m)} = \dfrac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \ln(1 + 1/v)},$        eqn 22

145     $\text{ICC}_{\text{gamma}-\ln} = \dfrac{\sigma_\alpha^2}{\sigma_\alpha^2 + \ln(1 + 1/v)},$        eqn 23

## 146    Obtaining the observation-level variance by the 'first' delta method

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

148     level variance $\sigma_\varepsilon^2$ can be obtained via the variance of the log-normal distribution, as described

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

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

151     discussed in some detail below.

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

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

154     of a variable $x$ when the (error) variance of $x$ itself is known (see Ver Hoff 2012; for an accessible

155     reference for biologists, Powell 2007). A simple case of the delta method for variance

156     approximation can be written as:

157     $\text{var}[f(x)] \approx \text{var}[x] \left( \dfrac{d}{dx} f(x) \right)^2,$        eqn 24

8

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

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

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

161    found in the Appendices). It is the delta method that Foulley *et al.* (1987) used to derive the

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

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

164    distributions without overdispersion, $\sigma_d^2$ is equal to $\sigma_\varepsilon^2$ because $\sigma_e^2 = 0$). One clear advantage of

165    the delta method is its flexibility, and we can easily obtain the observation-level variance $\sigma_\varepsilon^2$ for all

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

167    obtain $\sigma_\varepsilon^2$ for the Tweedie (compound Poisson-gamma) distribution, which has been used to model

168    non-negative real numbers in ecology (e.g., Foster & Bravington 2013; Zhang 2013). For the

169    Tweedie distribution, the variance on the observed scale has the relationship $\text{var}[y] = \varphi\mu^p$ where $\mu$

170    is the mean on the observed scale and $\varphi$ is the dispersion parameter (comparable to $\lambda$ and $\omega$ in

171    Equation 9), and $p$ is a positive constant called an index parameter. Therefore, when used with the

172    log-link function, an approximated $\sigma_\varepsilon^2$ value can be obtained by $\varphi\mu^{(p-2)}$ according to Equation 24.

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

174    The use of the trigamma function $\psi_1$ is limited to distributions with log link, but it should provide

175    the most accurate estimate of the observation level variance $\sigma_\varepsilon^2$. This is because the variance of a

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

177    gamma distribution (Tempelman and Gianola 1999) and hence $\sigma_\varepsilon^2$ is $\psi_1(\nu)$. At the level of the

178    statistical parameters (Table 1; on the 'expected data' scale; *sensu* deVillemereuil et al. 2016; see

179    their Figure 1), Poisson and negative binomial distributions can be both seen as re-

180    parameterizations of gamma distributions (Tempelman and Gianola 1999) and $\sigma_\varepsilon^2$ can be obtained

181    using the trigamma function (Table 1). For example, $\sigma_\varepsilon^2$ for the Poisson distribution is $\psi_1(\lambda)$ with

182    the speciality that in the case of Poisson distributions $\sigma_\varepsilon^2 = \sigma_d^2$. As we show in Appendix S2,

183    $\ln(1+1/\lambda)$ (log-normal approximation), $1/\lambda$ (delta method approximation) and $\psi_1(\lambda)$ (trigamma

184    function) are similar if $\lambda$ is greater than 2. Nonetheless, our recommendation is to use the trigamma

185    function for obtaining $\sigma_\varepsilon^2$ whenever this is possible.

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

187    sense, it is not; see de Villemereuil et al. 2016) using negative binomial GLMMs, the trigamma

188    function has been previously used to obtain observation-level variance (Matos et al. 1997;

189    Tempelman and Gianola 1999; cf. de Villemereuil et al. 2016). Table 1 summarises observation-

190    level variance $\sigma_\varepsilon^2$ for overdispersed Poisson, negative binomial and gamma distributions for

191    commonly used link functions.

## How to estimate $\lambda$ from data

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

194    random effect (model 5):

195    $y_{ij} \sim \text{Poisson}(\lambda_{ij}),$ 　　　　　　　　　　　　　　　　　　eqn 25

196    $\ln(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{p} \beta_h x_{hij} + \alpha_i + e_{ij},$ 　　　　　　　　　　eqn 26

197    $\alpha_i \sim \text{Gaussian}(0,\ \sigma_\alpha^2),$ 　　　　　　　　　　　　　　　eqn 27

198    $e_{ij} \sim \text{Gaussian}(0,\ \sigma_e^2),$ 　　　　　　　　　　　　　　　eqn 28

199    where $y_{ij}$ is the $j$th observation of the $i$th individual, and follows a Poisson distribution with the

200    parameter $\lambda_{ij}$, $e_{ij}$ is an additive overdispersion term for $j$th observation of the $i$th individual, and the

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

202    $\text{ICC}_{\text{GLMM}}$ can be calculated as:

10

203 $$R^2_{\text{P-ln}(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_\alpha + \sigma^2_e + \ln(1+1/\lambda)},$$ eqn 29

204 $$\text{ICC}_{\text{P-ln}} = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_e + \ln(1+1/\lambda)},$$ eqn 30

205 where, as metioned above, the term $\ln(1+1/\lambda)$ is $\sigma^2_\varepsilon$ (or $\sigma^2_d$) for Poisson distributions with the log

206 link (Table 1).

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

208 model or models with centred fixed factors) $\exp(\beta_0)$ as an estimator of $\lambda$ (Nakagawa and Schielzeth

209 2010, 2013). We also suggested that it is possible to use the mean of observed values $y_{ij}$.

210 Unfortunately, these two recommendations are often inconsistent with each other. This is because,

211 given the model 5 (and all the models in the previous section), the following relationships hold:

212 $$\exp(\beta_0) \le \text{E}[y_{ij}]$$ eqn 31

213 $$\text{E}[\lambda_{ij}] = \exp(\beta_0 + 0.5\sigma^2_\tau)$$ eqn 32

214 $$\text{E}[y_{ij}] = \text{E}[\lambda_{ij}]$$ eqn 33

215 where E represents the expected value (i.e., mean) on the observed scale, $\beta_0$ is the mean value on

216 the latent scale (i.e. $\beta_0$ from the intercept-only model), $\sigma^2_\tau$ is the total variance on the latent scale

217 (e.g., $\sigma^2_\alpha + \sigma^2_e$ in the models 1and 5, and $\sigma^2_\alpha$ in models 2-4; Nakagawa and Schielzeth 2010; see

218 also Carrasco 2010). In fact, $\exp(\beta_0)$ gives the median value of $y_{ij}$ rather than the mean of $y_{ij}$,

219 assuming a Poisson distribution. Thus, the use of $\exp(\beta_0)$ will often overestimate $\sigma^2_d$, providing

220 conservative (smaller) estimates of $R^2$ and ICC, than when using averaged $y_{ij}$, which is a better

221 estimate of $\text{E}[y_{ij}]$. Quantitative differences between the two approaches may often be negligible, but

222 when $\lambda$ is small, the difference can be substantial so the choice of the method needs to be reported

223 for reproducibility (Appendix S2). Our new recommendation is to obtain $\lambda$ via Equation 32. When

224 sampling is balanced (i.e. observations are equally distributed across individuals and covariates),

225     Equation 32 and the mean of the observed values will give similar values, but when unbalanced,

226     method Equation 32 is preferable. This recommendation for obtaining $\lambda$ also applies to negative

227     binomial GLMMs (see Table 1).


228     ## Jensen's inequality and the 'second' delta method

229     A general form of Equation 31 is known as Jensen's inequality, $g(\bar{x}) \leq \overline{g(x)}$ where $g$ is a convex

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

231     transformed values (the opposite is true for a concave function; that is, $g(\bar{x}) \geq \overline{g(x)}$; Rao 2002). In

232     fact, whenever the function is not strictly linear, simple application of the inverse link function (or

233     back-transformation) cannot be used to translate the mean on the latent scale into the mean value on

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

235     GLMMs (and also generalized linear models GLMs and linear models with transformed response

236     variables).

237     Although log-link GLMMs (e.g., model 5) have an analytical formula (Equation 32), this is not

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

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

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

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

242     (Oehlert 1992; Ver Hoef 2012). A simple case of the delta method for bias correction can be written

243     as:

244     $$\mathrm{E}[f(x)] \approx f(x) + 0.5\sigma_\tau^2 \frac{d^2}{dx^2} f(x),$$         eqn 34

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

246     Equations 24 and 32. By employing this bias correction delta method (with

247    $d^2 \exp(x)/dx^2 = \exp(x))$, we can approximate Equation 32 using the same symbols as in Equations

248    31-33:

249    $$E[\lambda_{ij}] = E[\exp(\beta_0)] \approx \exp(\beta_0) + 0.5\sigma_\tau^2 \exp(\beta_0)$$    eqn 35

250    The comparison between Equation 32 (exact) and Equation 35 (approximate) is shown in Appendix

251    S3. The approximation is most useful when the exact formula is not available as in the case of a

252    binomial GLMM with logit link (model 6):

253    $$y_{ij} \sim \text{binomial}(p_{ij}, n_{ij}),$$    eqn 36

254    $$\text{logit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij},$$    eqn 37

255    $$\alpha_i \sim \text{Gaussian}(0, \sigma_\alpha^2),$$    eqn 38

256    $$e_{ij} \sim \text{Gaussian}(0, \sigma_e^2),$$    eqn 39

257    where $y_{ij}$ is the number of 'success' in $n_{ij}$ trials by the $i$th indivdiual at the $j$th occasion (for binary

258    data, $n_{ij}$ is always 1), $p_{ij}$ is the underlying probability of success, and the other symboles are the

259    same as above.

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

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

262    For example, the overall intercept, $\beta_0$ on the latent scale could be transformed not just with the

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

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

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

266    $$E[y_{ij}] = E[\text{logit}^{-1}(\beta_0)] \approx \frac{\exp(\beta_0)}{1+\exp(\beta_0)} + 0.5\sigma_\tau^2 \frac{\exp(\beta_0)(1-\exp(\beta_0))}{(1+\exp(\beta_0))^3}.$$    eqn 40

13

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

268     Another approximation proposed by Zeger et al. (1988) produces similar (but slightly better)

269     estimates than Equation 40. Using our notation, this approximation can be written as:

270     $$\mathrm{E}[p_{ij}] \approx \mathrm{logit}^{-1}\left( \beta_0 \sqrt{1 + \left(\frac{16\sqrt{3}}{15\pi}\right)^2 \sigma_\tau^2}^{\,-1} \right).$$        eqn 41

271     A comparison between Equations 40 and 41 is also shown in Appendix S3. This approximation

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

273     6 but using the probit link (i.e., $\mathrm{probit}(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + \alpha_i + e_{ij}$ in place of Equation 37):

274     $$\mathrm{E}[p_{ij}] = \mathrm{probit}^{-1}\left( \beta_0 \sqrt{1 + \sigma_\tau^2}^{\,-1} \right).$$        eqn 42

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

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

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

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

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

280     limitation of the delta method are described in Oehlert (1992).


## Special considerations for binomial GLMMs

282     The observation-level variance $\sigma_\varepsilon^2$ can be thought of as being added to the latent scale on which

283     other variance components are also estimated in a GLMM (Equations 10, 15, 20, 26, 37 for models

284     2-6). Since the proposed $R^2_{\mathrm{GLMM}}$ and ICC $_{\mathrm{GLMM}}$ are ratios between variance components and their

285     additive combinations, we can show using the delta method that $R^2_{\mathrm{GLMM}}$ and ICC $_{\mathrm{GLMM}}$ calculated

286     via $\sigma_\varepsilon^2$ approximate to those of $R^2$ and ICC on the observation (original) scale (shown in Appendix

14

287    S4 ). In some cases, there exist specific formulas for ICC on the observation scale (Nakagawa and

288    Schielzeth 2010). In the past, we distinguished between ICC on the latent scale and on the

289    observation scale (Nakagawa and Schielzeth 2010). Such a distinction turns out to be strictly

290    appropriate only for binomial distributions but not for Poisson distributions (and probably also not

291    for other non-Gaussian distributions). This is because the property of what we have called the

292    distribution-specific variance $\sigma_d^2$ for binomial distributions (e.g. $\pi^2/3$ for binomial error distribution

293    with the logit link function) is quite different from what we have discussed as the observation-level

294    variance $\sigma_\varepsilon^2$ although these two types of variance are related conceptually (i.e., both represents

295    variance due to non-Gaussian distributions with specific link functions). Let us explain this further.

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

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

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

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

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

301    variance; Nakagawa & Schielzeth 2010, 2013), the observation-level variance of the binomial

302    distribution changes as $p$ changes (see Appendix S5), suggesting these two observation-level

303    variances ($1/\lambda$ and $1/(p(1–p))$) are analogous while the distribution-specific variance $\pi^2/3$ is not.

304    Further, the minimum value of $1/(p(1–p))$ is 4, which is larger than $\pi^2/3 \approx 3.29$, meaning that the

305    use of $1/p(1–p)$ in $R^2$ and ICC will always produce larger values than those using $\pi^2/3$.

306    Consequently, Browne et al. (2005) showed that ICC values (or variance partition coefficients,

307    VPCs) estimated using $\pi^2/3$ were higher than corresponding ICC values on the observation

308    (original) scale using logistic-binomial GLMMs (see also Goldstein et al. 2002; Nakagawa and

309    Schielzeth 2010). Then, what is $\pi^2/3$?

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

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

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

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

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

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

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

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

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

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

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

321 (*sensu* Nakagawa and Schielzeth 2010) while, as mentioned above, $R^2$ and ICC values using the

322 observation-level variance estimate the counterparts on the observation (original) scale (cf. de

323 Villemereuil et al. 2016). The use of the theoretical distribution-specific variance will almost

324 always provide different values of $R^2_{\text{GLMM}}$ and ICC $_{\text{GLMM}}$ from those using the observation-level

325 obtained via the delta method (see Appendix S5). In any case, we should be aware that binomial

326 GLMMS are special cases for obtaining $R^2_{\text{GLMM}}$ and ICC $_{\text{GLMM}}$ from binomial GLMMs.


### Worked examples: revisting the beetles

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

329 for Nakagawa and Schielzeth (2013). In brief, the dataset represents a hypothetical species of beetle

330 that has the following life cycle: larvae hatch and grow in the soil until they pupate, and then adult

331 beetles feed and mate on plants. Larvae are sampled from 12 different populations ('Population';

332 see Fig. 1). Within each population, larvae are collected at two different microhabitats ('Habitat'):

333 dry and wet areas as determined by soil moisture. Larvae are exposed to two different dietary

334 treatments ('Treatment'): nutrient rich and control. The species is sexually dimorphic and can be

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

336 the other reddish brown ('Morph', labeled as A and B in Fig 1). Sexed pupae are housed in standard

16

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

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

339    960).

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

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

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

343    dispersion), (ii) the incidence of endoparasitic infections that we generated as being negative

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

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

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

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

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

349    All data generation and analyses were conducted in R 3.3.1 (R Development Core Team). We used

350    functions to fit GLMMs from the three R packages: 1) the *glmmadmb* function from glmmADMB

351    (Fournier et al. 2012), 2) the *glmmPQL* function from MASS (Venables and Ripley 2002) and 3)

352    the *glmer* and *glmer.nb* functions from lme4 (Bates et al. 2015). In Table 1, we only report results

353    from *glmmADMB* because this is the only function that can fit models with all relevant

354    distributional families. All scripts and results are provided as an electronic supplement (Appendix

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

356    which was fitted by the *cpglmm* function from the cplm package (Zhang 2013). Notably, our

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

358    piecewiseSEM (Lefcheck 2016). Another important note is that we often find less congruence in

359    GLMM results from the different packages than those of linear mixed-effects models, LMM. Thus,

360    it is recommended to run GLMMs in more than one package to check robustness of the results

361    although this may not always be possible.

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

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

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

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

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

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

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

369    models (implying that the denominator of Equation 10 shrinks).  However, the numerator also

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

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

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

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

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

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


## Alternatives and a cautionary note

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

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

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

380    key rate parameter $\lambda$ for Poisson and negative binomial distributions. We discussed important

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

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

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

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

385    not variance components of GLMMs (Schielzeth and Nakagawa 2013).

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

387     et al. (2016) have proposed $R^2$ for fixed effects in GLMMs, which they referred to as $R^2_{\beta*}$ (an

388     extension of an $R^2$ for fixed effects in linear mixed models or $R^2_{\beta}$ by Edwards et al. 2008). They

389     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

390     distribution (error structure) with any link function. Jaeger et al. (2016) highlight that in the

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

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

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

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

395     semi-partial $R^2$ via commonality analysis (see Ray-Mukherjee et al. 2014; note that unique variance

396     for each predictor in commonality analysis corresponds to semi-partial $R^{2;}$; Nimon and Oswald

397     2013).

398     Second, de Villemereuil et al. (2016) provided a framework with which one can estimate exact

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

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

401     the same as, our ICC $_{GLMM}$ using the observation-level variance, $\sigma^2_{\varepsilon}$ described above. Further, this

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

403     potential difficulty is that the method of de Villemereuli et al. is exact but that a numerical method

404     is used to solve relevant equations so one will require a software package (e.g., the QGglmm

405     package; de Villemereuil et al. 2016).

406     Finally, we finish by repeating what we said at the end of our original $R^2$ paper (Nakagawa and

407     Schielzeth 2013). Both $R^2$ and ICC are indices that are likely to reflect only one or a few aspects of

408     a model fit to the data and should not be used for gauging the quality of a model. We encourage

409     biologists use $R^2$ and ICC in conjunctions with other indices like information criteria (e.g. AIC, BIC

410    and DIC), and more importantly, with model diagnostics such as checking for model assumptions,

411    heteroscedasticity and sensitivity to outliers.

## Acknowledgements

417

# References

Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* **67,** 1-48.

Bolker, B.M. (2008) *Ecological models and data in R.* Princeton University Presss, Princeton, NJ.

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 in Ecology & Evolution,* **24,** 127-135.

Browne, W.J., Subramanian, S.V., Jones, K. & Goldstein, H. (2005) Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society Series a-Statistics in Society,* **168,** 599-613.

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.

de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. (in press) General methods for evolutionary quantitative genetic inference from generalised mixed models. *Genetics.* doi: 10.1534/genetics.115.186536

Foster, S.D. & Bravington, M.V. (2013) A Poisson-Gamma model for analysis of ecological non-negative continuous data. *Environmental and Ecological Statistics,* **20,** 533-552.

Foulley, J.L., Gianola, D. & Im, S. (1987) Genetic Evaluation of Traits Distributed as Poisson-Binomial with Reference to Reproductive Characters. *Theoretical and Applied Genetics,* **73,** 870-877.

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. *Optimization Methods & Software,* **27,** 233-249.

443 Gelman, A. & Hill, J. (2006) *Data analysis using regression and multilevel/hierarchical models*

444        Cambridge University Press, Cambridge.

445 Goldstein, H., Browne, W. & Rasbash, J. (2002) Partitioning variation in multilevel models.

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

447 Hoef, J.M.V. (2012) Who Invented the Delta Method? *American Statistician,* **66,** 124-127.

448 Hox, J. (2010) *Multilevel analysis*. Routledg, New York.

449 Jaeger, B.C., Edwards, L.J., Das, K. & Sen, P.K. (2016) An R2 statistic for fixed effects in the

450        generalized linear mixed model. *Journal of Applied Statistics***,** doi:

451        10.1080/02664763.02662016.01193725.

452 Johnson, P.C.D. (2014) Extension of Nakagawa & Schielzeth's R-GLMM(2) to random slopes

453        models. *Methods in Ecology and Evolution,* **5,** 944-946.

454 LaHuis, D.M., Hartman, M.J., Hakoyama, S. & Clark, P.C. (2014) Explained Variance Measures

455        for Multilevel Models. *Organizational Research Methods,* **17,** 433-451.

456 Lefcheck, J.S. (2016) PIECEWISESEM: Piecewise structural equation modelling in R for ecology,

457        evolution, and systematics. *Methods in Ecology and Evolution,* **7,** 573-579.

458 Lessells, C.M. & Boag, P.T. (1987) Unrepeatable repeatabilities - a common mistake. *Auk,* **104,**

459        116-121.

460 Matos, C.A.P., Thomas, D.L., Gianola, D., Tempelman, R.J. & Young, L.D. (1997) Genetic

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

462        Estimation of genetic parameters. *Journal of Animal Science,* **75,** 76-87.

463 Morrissey, M.B., de Villemereuil, P., Doligez, B. & Gimenez, O. (2014) Bayesian approaches to

464        the quantitative genetic analysis of natural populations. *Quantitative genetics in the wild*

465        (eds A. Charmantier, D. Garant & L.E.B. Kruuk), pp. 228-253. Oxford University Press,

466        Oxford.

467 Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a

468        practical guide for biologists. *Biological Reviews,* **85,** 935-956.

469   Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from

470        generalized linear mixed-effects models. *Methods in Ecology and Evolution,* **4,** 133-142.

471   Nimon, K.F. & Oswald, F.L. (2013) Understanding the Results of Multiple Linear Regression:

472        Beyond Standardized Regression Coefficients. *Organizational Research Methods,* **16,** 650-

473        674.

474   Oehlert, G.W. (1992) A note on the delta method. *American Statistician,* **46,** 27-29.

475   Powell, L.A. (2007) Approximating variance of demographic parameters using the delta method: A

476        reference for avian biologists. *Condor,* **109,** 949-954.

477   R Development Core Team (2016) R: A language and environment for statistical computing. R

478        Foundation for Statistical Computing, Vienna, Austria.

479   Rao, C.R. (2002) *Linear statistical inference and its applications,* 2nd ed. edn. John Wiley & Sons,

480        New York.

481   Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. (2014)

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

483        in the face of multicollinearity. *Methods in Ecology and Evolution,* **5,** 320-328.

484   Schielzeth, H. & Nakagawa, S. (2013) Nested by design: model fitting and interpretation in a mixed

485        model era. *Methods in Ecology and Evolution,* **4,** 14-24.

486   Snijders, T. & Bosker, R. (1999) *Multilevel Analysis: an Introduction to basic and advanced*

487        *multilevel modeling*. Sage, London.

488   Snijders, T. & Bosker, R. (2011) *Multilevel Analysis: an Introduction to basic and advanced*

489        *multilevel modeling,* 2nd edn. Sage, London.

490   Tempelman, R.J. & Gianola, D. (1999) Genetic analysis of fertility in dairy cattle using negative

491        binomial mixed models. *Journal of Dairy Science,* **82,** 1834-1847.

492   Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistics with S,* 4 edn. Springer, New

493        York.

494    Ver Hoef, J.M. & Boveng, P.L. (2007) Quasi-Poisson vs. negative binomial regression: how should

495        we model overdispersed count data? *Ecology,* **88,** 2766-2772.

496    Zhang, Y.W. (2013) Likelihood-based and Bayesian methods for Tweedie compound Poisson linear

497        mixed models. *Statistics and Computing,* **23,** 743-757.

498

499

Table 1. The observation-level variance $\sigma_\varepsilon^2$ for the three distributional families: quasi-Poisson (overdispersed Poisson), negative binomial and gamma with the three different methods for deriving $\sigma_\varepsilon^2$: the delta method, long-normal approximation and the trigamma function, $\psi_1$.

| 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 $\omega=1$) | $\lambda>0$ $\omega>0$ | $var[y]=\lambda\omega$ | square-root | $0.25\,\omega$ | - | |
| Negative binomial (NB) | $NB(\lambda,\ \theta)$ | $E[y]=\lambda$ | log | $\dfrac{1}{\lambda}+\dfrac{1}{\theta}$ | $\ln\left(1+\dfrac{1}{\lambda}+\dfrac{1}{\theta}\right)$ | $\psi_1\left(\left[\dfrac{1}{\lambda}+\dfrac{1}{\theta}\right]^{-1}\right)$ |
| | $\lambda>0$ $\theta>0$ | $var[y]=\lambda+\dfrac{\lambda^2}{\theta}$ | square-root | $0.25\left(1+\dfrac{\lambda}{\theta}\right)$ | - | |
| Gamma | $gmma(\lambda,\ \nu)$ | $E[y]=\lambda$ | log | $\dfrac{1}{\nu}$ | $\ln\left(1+\dfrac{1}{\nu}\right)$ | $\psi_1(\nu)$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| $\lambda > 0$ $v > 0$ | | $\text{var}[y] = \dfrac{\lambda^2}{v}$ | inverse (reciprocal) | $\dfrac{1}{v\lambda^2}$ | - | |
| Gamma (alternative parameterization) | gamma($v$, $\kappa$) | $E[y] = \dfrac{v}{\kappa}$ | log | $\dfrac{1}{v}$ | $\ln\left(1 + \dfrac{1}{v}\right)$ | $\psi_1(v)$ |
| | $v > 0$ $\kappa > 0$ | $\text{var}[y] = \dfrac{v}{\kappa^2}$ | inverse (reciprocal) | $\dfrac{\kappa^2}{v^3}$ | - | |

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

Table 2. The distribution-specific variance $\sigma_d^2$ and observation-level variance $\sigma_\varepsilon^2$ 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) | $\ln(-\ln(1-p))$ | $\dfrac{\pi^2}{6} \sim 1.65$ (Gumbel distribution) | $\dfrac{p}{\left(\ln(1-p)\right)^2(1-p)}$ (min $\sim$ 1.54; $p \sim 0.8$; $\sim$ 2.08; $p = 0.5$) |

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

27

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 | - | - |
| | | | | [-2.511, -1.884] | | [-0.230, -0.196] | | [-1.256, -0.955] | | - |
| **Random effects** | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ |
| 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 |
| $R^2_{\text{GLMM}(m)}$ | - | 10.01% | - | 48.83% | - | 49.54% | - | 78.34% | - | 3.98% (4.57%) |
| $R^2_{\text{GLMM}(c)}$ | - | 47.19% | - | 86.91% | - | 72.52% | - | 93.34% | 3.98% (4.57%) | 27.46% (31.55%) |
| ICC$_{\text{[Population]}}$ | 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$_{\text{[Container]}}$ | 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$_{\text{[Container]}}$ 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).