

## A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models

Shinichi Nakagawa<sup>1,2\*</sup> and Holger Schielzeth<sup>3</sup>

<sup>1</sup>National Centre for Growth and Development, Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9054, New Zealand; <sup>2</sup>Department of Behavioral Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner-Straße, 82319 Seewiesen, Germany; and <sup>3</sup>Department of Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33615, Bielefeld, Germany

### Summary

1. The use of both linear and generalized linear mixed-effects models (LMMs and GLMMs) has become popular not only in social and medical sciences, but also in biological sciences, especially in the field of ecology and evolution. Information criteria, such as Akaike Information Criterion (AIC), are usually presented as model comparison tools for mixed-effects models.

2. The presentation of ‘variance explained’ ( $R^2$ ) as a relevant summarizing statistic of mixed-effects models, however, is rare, even though  $R^2$  is routinely reported for linear models (LMs) and also generalized linear models (GLMs).  $R^2$  has the extremely useful property of providing an absolute value for the goodness-of-fit of a model, which cannot be given by the information criteria. As a summary statistic that describes the amount of variance explained,  $R^2$  can also be a quantity of biological interest.

3. One reason for the under-appreciation of  $R^2$  for mixed-effects models lies in the fact that  $R^2$  can be defined in a number of ways. Furthermore, most definitions of  $R^2$  for mixed-effects have theoretical problems (e.g. decreased or negative  $R^2$  values in larger models) and/or their use is hindered by practical difficulties (e.g. implementation).

4. Here, we make a case for the importance of reporting  $R^2$  for mixed-effects models. We first provide the common definitions of  $R^2$  for LMs and GLMs and discuss the key problems associated with calculating  $R^2$  for mixed-effects models. We then recommend a general and simple method for calculating two types of  $R^2$  (marginal and conditional  $R^2$ ) for both LMMs and GLMMs, which are less susceptible to common problems.

5. This method is illustrated by examples and can be widely employed by researchers in any fields of research, regardless of software packages used for fitting mixed-effects models. The proposed method has the potential to facilitate the presentation of  $R^2$  for a wide range of circumstances.

**Key-words:** coefficient of determination, goodness-of-fit, heritability, information criteria, intra-class correlation, linear models, model fit, repeatability, variance explained

### Introduction

Many biological datasets have multiple strata due to the hierarchical nature of the biological world, for example, cells within individuals, individuals within populations, populations within species and species within communities. Therefore, we need statistical methods that explicitly model the hierarchical structure of real data. Linear mixed-effects models (LMMs; also referred to as multilevel/hierarchical models) and their extension, generalized linear mixed-effects models (GLMMs) form a class of models that incorporate multilevel hierarchies in data. Indeed, LMMs and GLMMs are becoming a part of standard methodological tool kits in biological sciences (Bolker *et al.* 2009), as well as in social and medical sciences (Gelman & Hill 2007; Congdon 2010; Snijders & Bosker 2011). The widespread use of GLMMs demonstrates

that a statistic that summarizes the goodness-of-fit of mixed-effects model to the data would be of great importance. There seems currently no such summary statistic that is widely accepted for mixed-effects models.

Many scientists have traditionally used the coefficient of determination,  $R^2$  (ranging from 0 to 1), as a summary statistic to quantify the goodness-of-fit of fixed effects models such as multiple linear regressions, ANOVA, ANCOVA and generalized linear models (GLMs). The concept of  $R^2$  as ‘variance explained’ is intuitive. Because  $R^2$  is unitless, it is extremely useful as a summary index for statistical models because one can objectively evaluate the fit of models and compare  $R^2$  values across studies in a similar manner as standardized effect size statistics under some circumstances (e.g. models with the same responses and similar set of predictors or in other words, it can be utilized for meta-analysis; Nakagawa & Cuthill 2007).

In Table 1, we briefly summarize 12 properties of  $R^2$  (based on Kvålseth 1985 and Cameron & Windmeijer 1996; compilation adopted from Orelie & Edwards 2008) that will provide

\*Correspondence author. E-mail: shinichi.nakagawa@otago.ac.nz

**Table 1.** Twelve properties of 'traditional'  $R^2$  for regression models; adopted from Orelie & Edwards (2008)

Property	References
$R^2$ must represent a goodness-of-fit and have intuitive interpretation	Kvålseth (1985)
$R^2$ must be unit free; that is, dimensionless	Kvålseth (1985)
$R^2$ should range from 0 to 1 where 1 represents a perfect fit	Kvålseth (1985)
$R^2$ should be general enough to apply to any type of statistical model	Kvålseth (1985)
$R^2$ values should not be affected by different model fitting techniques	Kvålseth (1985)
$R^2$ values from different models fitted to the same data should be directly comparable	Kvålseth (1985)
Relative $R^2$ values should be comparable to other accepted goodness-of-fit measures	Kvålseth (1985)
All residuals (positive and negative) should be weighted equally by $R^2$	Kvålseth (1985)
$R^2$ values should always increase as more predictors are added (without degrees-of-freedom correction)	Cameron & Windmeijer (1996)
$R^2$ values based on residual sum of squares and those based on explained sum of squares should match	Cameron & Windmeijer (1996)
$R^2$ values and statistical significance of slope parameters should show correspondence	Cameron & Windmeijer (1996)
$R^2$ should be interpretable in terms of the information content of the data	Cameron & Windmeijer (1996)

the reader with a good sense of what a 'traditional'  $R^2$  statistic should be and also provide a benchmark for generalizing  $R^2$  to mixed-effects models. Generalizing  $R^2$  from linear models (LMs) to LMMs and GLMMs turns out to be a difficult task. A number of ways of obtaining  $R^2$  for mixed models have been proposed (e.g. Snijders & Bosker 1994; Xu 2003; Liu, Zheng & Shen 2008; Orelie & Edwards 2008). These proposed methods, however, share some theoretical problems or practical difficulties (discussed in detail below), and consequently, no consensus for a definition of  $R^2$  for mixed-effects models has emerged in the statistical literature. Therefore, it is not surprising that  $R^2$  is rarely reported as a model summary statistic when mixed models are used.

In the absence of  $R^2$ , information criteria are often used and reported as comparison tools for mixed models. Information criteria are based on the likelihood of the data given a fitted model (the 'likelihood') penalized by the number of estimated parameters of the model. Commonly used information criteria include Akaike Information Criterion (AIC) (Akaike 1973), Bayesian information criterion (BIC), (Schwarz 1978) and the more recently proposed deviance information criterion (DIC), (Spiegelhalter *et al.* 2002; reviewed in Claeskens & Hjort 2009; Grueber *et al.* 2011; Hamaker *et al.* 2011). Information criteria are used to select the 'best' or 'better' models, and they are indeed useful for selecting the most parsimonious models from a candidate model set (Burnham & Anderson 2002). There are, however, at least three important limitations to the use of information criteria in relation to  $R^2$ : (i) while information criteria provide an estimate of the relative fit of alternative models, they do not tell us anything about the absolute model fit (cf. evidence ratio; Burnham & Anderson 2002), (ii) information criteria do not provide any information on variance explained by a model (Orelie & Edwards 2008), and (iii) information criteria are not comparable across different datasets under any circumstances, because they are highly dataset specific (in other words, they are not standardized effect statistics which can be used for meta-analysis; Nakagawa & Cuthill 2007).

In this paper, we start by providing the most common definitions of  $R^2$  in LMs and GLMs. We then review previously

proposed definitions of  $R^2$  measures for mixed-effects models and discuss the problems and difficulties associated with these measures. Finally, we explain a general and simple method for calculating variance explained by LMMs and GLMMs and illustrate its use by simulated ecological datasets.

## Definitions of $R^2$

In this section, we first describe some of the existing methods for estimating a coefficient of determination,  $R^2$ , for LMs. A standard (general) linear model (LM) can be written as:

$$y_i = \beta_0 + \sum_{h=1}^p \beta_h x_{hi} + \varepsilon_i, \quad \text{eqn 1}$$

$$\varepsilon_i \sim \text{Gaussian}(0, \sigma_\varepsilon^2), \quad \text{eqn 2}$$

where  $y_i$  is the  $i$ th response value,  $x_{hi}$  is the  $i$ th value for the  $h$ th predictor,  $\beta_0$  is the intercept,  $\beta_h$  is the slope (regression coefficient) of the  $h$ th predictor,  $\varepsilon_i$  is the  $i$ th residual value and residual errors are normally (Gaussian) distributed with a variance of  $\sigma_\varepsilon^2$ . Such regression models are fitted by ordinary least squares (OLS) methods that minimize the sum of squared distances between observed and fitted responses (i.e. minimizing the residual sum of squares). The residual sum of squares appears in the formulation of the most common definition for the coefficient of determination,  $R^2$  (Kvålseth 1985; Draper & Smith 1998).

$$R_{\text{O}}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{eqn 3}$$

$$\hat{y}_i = \hat{\beta}_0 + \sum_{h=1}^p \hat{\beta}_h x_{hi}, \quad \text{eqn 4}$$

where  $n$  is the number of observations (i.e. the total sample size),  $\bar{y}$  is the mean of the response,  $\hat{y}_i$  is the  $i$ th fitted response value,  $\hat{\beta}_0$  and  $\hat{\beta}_h$  are estimates of  $\beta_0$  and  $\beta_h$ , respectively, and the subscript 'O' in  $R_{\text{O}}^2$  signifies OLS regression. An interesting and important feature to note here is that the definition of

‘variance explained’ is rather indirectly composed of 1 minus the ‘variance unexplained’ (we revisit this very point later). An equivalent yet perhaps more intuitive formulation of  $R^2_{\text{O}}$  can also be written as:

$$R^2_{\text{O}} = 1 - \frac{\text{var}(y_i - \hat{y}_i)}{\text{var}(y_i)}, \quad \text{eqn 5}$$

$$R^2_{\text{O}} = 1 - \frac{\sigma^2_{\epsilon}}{\text{var}(y_i)}, \quad \text{eqn 6}$$

where ‘var’ indicates the variance of what is in the following parentheses. Equation 6 can also be expressed as the ratio between the residual variance of the model of interest and the residual variance of the null model (also referred to as the empty model or the intercept model):

$$R^2_{\text{O}} = 1 - \frac{\sigma^2_{\epsilon}}{\sigma^2_{\epsilon_0}}, \quad \text{eqn 7}$$

where  $\sigma^2_{\epsilon_0}$  is the residual variance of the null model.

There are two difficulties with generalizing this definition of  $R^2_{\text{O}}$  to the GLMM context. When generalizing to non-Gaussian response variables (i.e. GLMs), it is not straightforward to get an appropriate estimate of the residual variance. Also, when generalizing to mixed-effects models that consist of error terms at different hierarchical levels (see below), it is not immediately obvious which estimate should be used as the unexplained variance. For GLMs,  $R^2$  can be defined using the maximum likelihood (ML) of the full and null models (Maddala 1983). Perhaps, the best-known and most popular definition is:

$$R^2_{\text{g}} = 1 - \left( \frac{L_0}{L_{\beta}} \right)^{\frac{2}{n}}, \quad \text{eqn 8}$$

where  $L_{\beta}$  is the likelihood of the data given the fitted model of interest and  $L_0$  is the likelihood of the data given the null model,  $n$  is the total sample size, the subscript ‘g’ in  $R^2_{\text{g}}$  signifies ‘general’ (this formulation is based on the geometric mean squared improvement; see Menard 2000). Because  $R^2_{\text{g}}$  cannot become 1 even when the model of interest fits data perfectly, Nagelkerke (1991) proposed an adjustment to Equation 8:

$$R^2_{\text{G}} = \left[ 1 - \left( \frac{L_0}{L_{\beta}} \right)^{\frac{2}{n}} \right] / \left[ 1 - (L_0)^{\frac{2}{n}} \right], \quad \text{eqn 9}$$

where the denominator term can be interpreted as the maximum possible value of  $R^2_{\text{g}}$  and the subscript ‘G’ in  $R^2_{\text{G}}$  signifies ‘General’. A definition of  $R^2$ , which is comparable to  $R^2_{\text{G}}$ , is:

$$R^2_{\text{D}} = 1 - \frac{-2 \ln(L_{\beta})}{-2 \ln(L_0)} \quad \text{eqn 10}$$

We have deliberately left  $-2$  in the denominator and numerator so that  $R^2_{\text{D}}$  (‘D’ signifies ‘deviance’) can be compared with Equation 3. For a LM (Equation 1), the  $-2$  log-likelihood statistic (sometimes referred to as deviance) is equal to the residual sum of squares based on OLS of this model (Menard

2000; see a series of  $R^2_{\text{D}}$  formulas for non-Gaussian responses in Table 1 of Cameron & Windmeijer 1997). There are several other likelihood-based definitions of  $R^2$  (reviewed in Cameron & Windmeijer 1997; Menard 2000), but we do not review these definitions, as they are less relevant to our approach below. We will instead discuss the generalization of  $R^2$  to LMMs and GLMMs, and associated problems in this process, in the next section.

### Common problems when generalizing $R^2$

First, let us imagine an experimental design where we sample repeatedly from the same set of individuals. Extending the LM shown in Equations 1–2, we can fit a LMM with one random factor (‘individuals’ in our example) defined as:

$$y_{ij} = \beta_0 + \sum_{h=1}^p \beta_h x_{hij} + \alpha_j + \epsilon_{ij}, \quad \text{eqn 11}$$

$$\alpha_j \sim \text{Gaussian}(0, \sigma^2_{\alpha}), \quad \text{eqn 12}$$

$$\epsilon_{ij} \sim \text{Gaussian}(0, \sigma^2_{\epsilon}), \quad \text{eqn 13}$$

where  $y_{ij}$  is the  $i$ th response of the  $j$ th individual,  $x_{hij}$  is the  $i$ th value of the  $j$ th individual for the  $h$ th predictor,  $\beta_0$  is the intercept,  $\beta_h$  is the slope (regression coefficient) of the  $h$ th predictor,  $\alpha_j$  is the individual-specific effect from a normal distribution of individual-specific effects with mean of zero and variance of  $\sigma^2_{\alpha}$  (between-individual variance) and  $\epsilon_{ij}$  is the residual associated with the  $i$ th value of the  $j$ th individual from a normal distribution of residuals with mean of zero and variance of  $\sigma^2_{\epsilon}$  (within-individual variance). As seen in the previous equations, LMMs have by definition more than one variance component (in this case two:  $\sigma^2_{\alpha}$  and  $\sigma^2_{\epsilon}$ ), while LMs have only one (Equations 1 and 2).

One of the earliest definitions of  $R^2$  for mixed-effects models is based on the reduction of each variance component when including fixed-effect predictors separately; in other words, separate  $R^2$  for each random effect and the residual variance (Raudenbush & Bryk 1986; Bryk & Raudenbush 1992; we detail this measurement in the section ‘Related issues’). This approach is analogous to Equation 7. As pointed out by Snijders & Bosker (1994), however, it is not uncommon that some predictors can reduce  $\sigma^2_{\epsilon}$  while simultaneously increasing  $\sigma^2_{\alpha}$ , and *vice versa* even though the total sum of variance components ( $\sigma^2_{\epsilon} + \sigma^2_{\alpha}$ ) is usually reduced (for an example, see Table 1 in Snijders & Bosker 1994). Such behaviour of variance components can sometimes result in negative  $R^2$  because  $\sigma^2_{\epsilon}$  and  $\sigma^2_{\alpha}$  can be larger than  $\sigma^2_{\epsilon_0}$  and  $\sigma^2_{\alpha_0}$ , respectively (i.e. the corresponding variance components in the intercept model).

To avoid this problem, Snijders & Bosker (1994) proposed what they refer to as  $R^2_1$  and  $R^2_2$  for LMMs with one random factor (as in Equation 11): one  $R^2$  value is calculated for each level of a LMM (i.e. the unit level and the grouping/individual level).  $R^2_1$  can be expressed in two forms (analogous to Equations 5 and 7):

$$R_1^2 = 1 - \frac{\text{var}(y_{ij} - \hat{y}_{ij})}{\text{var}(y_{ij})}, \quad \text{eqn 14}$$

$$\hat{y}_{ij} = \hat{\beta}_0 + \sum_{h=1}^p \hat{\beta}_h x_{hij}, \quad \text{eqn 15}$$

$$R_1^2 = 1 - \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2}{\sigma_{\varepsilon 0}^2 + \sigma_{\alpha 0}^2}, \quad \text{eqn 16}$$

where  $R_1^2$  is variance explained at the unit of analysis (i.e. level 1; within-individual variance explained),  $\hat{y}_{ij}$  is the  $i$ th fitted value for  $j$ th individual and other notations are as above. In a similar manner,  $R_2^2$  can be written as:

$$R_2^2 = 1 - \frac{\text{var}(\bar{y}_j - \hat{\bar{y}}_j)}{\text{var}(\bar{y}_j)}, \quad \text{eqn 17}$$

$$R_2^2 = 1 - \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2/k}{\sigma_{\varepsilon 0}^2 + \sigma_{\alpha 0}^2/k}, \quad \text{eqn 18}$$

$$k = \frac{M}{\sum_{j=1}^M \frac{1}{m_j}}, \quad \text{eqn 19}$$

where  $R_2^2$  is variance explained at the individual level (i.e. level 2; between-individual variance explained),  $\bar{y}_j$  is the mean observed value for the  $j$ th individual,  $\hat{\bar{y}}_j$  is the fitted value for  $j$ th individual,  $k$  is the harmonic mean of the number of replicates per individuals,  $m_j$  is the number of replicates for the  $i$ th individual,  $M$  is the total number of individuals, and other notations are as above. An advantage of using  $R_1^2$  and  $R_2^2$  is that we can evaluate how much variance is explained at each level of the analysis. However, there are at least three problems with this approach: (i) it turns out that  $R_1^2$  and  $R_2^2$  can decrease in larger models (note that  $R_0^2$  can only increase when more predictors are added without the degrees of freedom adjustment; see Table 1), (ii) it is not clear how  $R_1^2$  and  $R_2^2$  can be extended to more than two levels (i.e. more than one random factor) and (iii) it is also not obvious how  $R_1^2$  and  $R_2^2$  are to be generalized to GLMMs.

The first problem means that because  $(\sigma_\varepsilon^2 + \sigma_\alpha^2)$  of a model with more predictors can be larger than that of a model of fewer predictors,  $R_1^2$  and  $R_2^2$  could also take negative values (Snijders & Bosker 1994). In other words, the estimate of  $(\sigma_\varepsilon^2 + \sigma_\alpha^2)$  can be larger than that of  $(\sigma_{\varepsilon 0}^2 + \sigma_{\alpha 0}^2)$ . Snijders & Bosker (1999) offer two explanations for decreases in  $R^2$  and/or negative  $R^2$  in a larger model: (i) chance fluctuation (or sampling variance) that is most prominent when the sample size is small or (ii) misspecification of the model, when the new predictor is redundant in relation to one or more other predictors in the model. Snijders & Bosker (1999) suggest that decreases in  $R_1^2$  and  $R_2^2$  (changes in the ‘wrong’ direction) can be used as a diagnostic in model selection. However, such misspecification does not need to be the cause of an increase in  $(\sigma_\varepsilon^2 + \sigma_\alpha^2)$  (and consequently decreases in  $R_1^2$  and  $R_2^2$ ).

The second problem of extending  $R_1^2$  and  $R_2^2$  to models with more than two levels was addressed by Gelman & Pardoe (2006), who provide a solution to extend  $R_1^2$  and  $R_2^2$  to any arbitrary numbers of levels (or random factors) in a Bayesian

framework. However, its general implementation is rather difficult, and we therefore refer to the original publication for those interested in this method.

The third problem of generalizing  $R_1^2$  and  $R_2^2$  is particularly profound because the residual variance,  $\sigma_\varepsilon^2$ , cannot be easily defined for non-Gaussian responses (see also below). At first glance, adopting likelihood-based  $R^2$  measures such as in Equations 8–10 could resolve this problem although such a method only provides  $R^2$  at the unit level (i.e. level 1); indeed, this type of solution has been recommended before (Edwards *et al.* 2008). Unfortunately, there are three obstacles to using a likelihood-based  $R^2$  like  $R_D^2$  for generalized models: (i) the likelihoods cannot be compared when models are fitted by restricted maximum likelihood (REML) (the standard way to estimate variance components in LMMs; Pinheiro & Bates 2000), (ii) it is not clear whether we should use the likelihood from the null model such as  $y_{ij} = \beta_0 + \varepsilon_{ij}$  (excluding random factors) or from the null model such as  $y_{ij} = \beta_0 + \alpha_j + \varepsilon_{ij}$  (including random factors; see Equation 10) and (iii) likelihood-based  $R^2$  measures applied to LMMs and GLMMs are also subject to the problem of decreased or even negative  $R^2$  with the introduction of additional predictors. We are not aware of a solution to this latter obstacle, but partial solutions to obstacles (i) and (ii) have been suggested and need separate discussion.

The first obstacle of fitting models with REML only applies to LMMs, and this can be resolved by using the ML estimates instead of REML. However, it is well known that variance components will be biased when models are fitted by ML (e.g. Pinheiro & Bates 2000).

With respect to the second obstacle regarding the choice of null models, it seems that both are permitted and accepted in the literature (e.g. Xu 2003; Orelie & Edwards 2008). Inclusion of random factors in the intercept model, however, can certainly change the likelihood of the null model that is used as a reference, and thus, it changes  $R^2$  values. This relates to an important matter. For mixed-effects models,  $R^2$  can be categorized loosely into two types: marginal  $R^2$  and conditional  $R^2$  (Vonesh, Chinchilli & Pu 1996). Marginal  $R^2$  is concerned with variance explained by fixed factors, and conditional  $R^2$  is concerned with variance explained by both fixed and random factors. So far, we only concentrated on the former, marginal  $R^2$ , but we will expand more on the distinction between the two types in the next section.

Although we do not review all proposed definitions of  $R^2$  for mixed-effects models here (see Menard 2000; Xu 2003; Orelie & Edwards 2008; Roberts *et al.* 2011), it appears that all alternative definitions of  $R^2$  suffer from one or more aforementioned problems and their implementations may not be straightforward. In the next section, we introduce a definition of  $R^2$ , which is simple and common to both LMMs and GLMMs and probably less prone to the aforementioned problems than previously proposed definitions.

## General and simple $R^2$ for GLMMs

We first revisit the point that variance explained ( $R_0^2$ ) is actually defined via the variance unexplained by the model, and



now we redefine  $R^2_{\text{O}}$  more directly in terms of variance explained:

$$R^2_{\text{O}} = \frac{\sum_{i=1}^n (\bar{y} - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{eqn 20}$$

$$R^2_{\text{O}} = \frac{\text{var}(\hat{y}_i)}{\text{var}(y_i)}, \quad \text{eqn 21}$$

where the notations are as in Equations 3–6. Below, we extend this more direct formulation first to LMMs and then to GLMMs. For simplicity, we use a LMM with two random factors as an example. For the sake of illustration, assume that the two random effects are ‘groups’ (with individuals uniquely assigned to groups) and ‘individuals’ (with multiple observations per individual) (c.f. Equations 11–13). Observations are thus clustered in individuals, and individuals are nested within groups (see Schielzeth & Nakagawa 2012 for a discussion of nesting in mixed models). The model can be written as:

$$y_{ijk} = \beta_0 + \sum_{h=1}^p \beta_h x_{hijk} + \gamma_k + \alpha_{jk} + \varepsilon_{ijk}, \quad \text{eqn 22}$$

$$\gamma_k \sim \text{Gaussian}(0, \sigma_\gamma^2), \quad \text{eqn 23}$$

$$\alpha_{jk} \sim \text{Gaussian}(0, \sigma_\alpha^2), \quad \text{eqn 24}$$

$$\varepsilon_{ijk} \sim \text{Gaussian}(0, \sigma_\varepsilon^2), \quad \text{eqn 25}$$

where  $y_{ijk}$  is the  $i$ th response of the  $j$ th individual, belonging to the  $k$ th group,  $x_{hijk}$  is the  $i$ th value of the  $j$ th individual in the  $k$ th group for the  $h$ th predictor,  $\gamma_k$  is the group-specific effect from a normal distribution of group-specific effects with mean of zero and variance of  $\sigma_\gamma^2$ ,  $\alpha_{jk}$  is the individual-specific effect from a normal distribution of individual-specific effects with mean of zero and variance of  $\sigma_\alpha^2$  and  $\varepsilon_{ijk}$  is the residual from a normal distribution of group-specific effects with mean of zero and variance of  $\sigma_\varepsilon^2$ . An  $R^2$  for LMM given by Equation 22 can be defined as:

$$R^2_{\text{LMM}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \quad \text{eqn 26}$$

$$\sigma_f^2 = \text{var}\left(\sum_{h=1}^p \beta_h x_{hijk}\right), \quad \text{eqn 27}$$

where  $\sigma_f^2$  is the variance calculated from the fixed effect components of the LMM (c.f. Snijders & Bosker 1999),  $m$  in the parentheses indicates marginal  $R^2$  (i.e. variance explained by fixed factors; see below for conditional  $R^2$ ). Estimating  $\sigma_f^2$  can, in principle, be carried out by predicting fitted values based on the fixed effects alone (equivalent to multiplying the design matrix of the fixed effects with the vector of fixed effect estimates) followed by calculating the variance of these fitted

values (Snijders & Bosker 1999). Note that  $\sigma_f^2$  should be estimated without degrees-of-freedom correction.

An obvious advantage of this formulation is that  $R^2_{\text{LMM}(m)}$  will never be negative. It is possible that  $R^2_{\text{LMM}(m)}$  can decrease by the addition of predictors (remember that  $R^2_{\text{O}}$  never decrease with more predictors), but this is unlikely, because  $\sigma_f^2$  should always increase when predictors are added to the model (compare Equations 16 and 26).

We now generalize  $R^2_{\text{LMM}(m)}$  to GLMMs. We have mentioned already that for non-Gaussian responses, it is difficult to define the residual variance,  $\sigma_\varepsilon^2$ . However, it is possible to define the residual variance on the latent (or link) scale, although this definition of the residual variance is specific to the error distribution and the link function used in the analysis. In GLMMs,  $\sigma_\varepsilon^2$  can be expressed as three components: (i) multiplicative dispersion ( $\omega$ ), (ii) additive dispersion ( $\sigma_e^2$ ) and (iii) distribution-specific variance ( $\sigma_d^2$ ) (detailed in Nakagawa & Schielzeth 2010). GLMMs can be implemented in two distinct ways, either by multiplicative or additive dispersion; dispersion is fitted to account for variance that exceeds or falls short of the distribution-specific variance (e.g. from binomial or Poisson distributions). In this paper, we only consider additive dispersion implementation of GLMMs although the formulae that we present below can be easily modified for the use with GLMMs that apply to multiplicative dispersion. For more details and also for a review of intra-class correlation (also known as repeatability) and heritability, both of which are closely connected to  $R^2$  (see Nakagawa & Schielzeth 2010). When additive dispersion is used,  $\sigma_\varepsilon^2$  is equal to the sum of the additive dispersion component and the distribution-specific variance ( $\sigma_e^2 + \sigma_d^2$ ), and thus,  $R^2$  for GLMMs can be defined as:

$$R^2_{\text{GLMM}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + \sigma_d^2}, \quad \text{eqn 28}$$

where  $R^2_{\text{GLMM}(m)}$  is variance explained on the latent (or link) scale rather than original scale. This can be easily generalized to multiple levels:

$$R^2_{\text{GLMM}(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 + \sigma_d^2}, \quad \text{eqn 29}$$

where  $u$  is the number of random factors in GLMMs (or LMMs) and  $\sigma_l^2$  is the variance component of the  $l$ th random factor. Equation 29 can be modified to express conditional  $R^2$  (i.e. variance explained by fixed and random factors).

$$R^2_{\text{GLMM}(c)} = \frac{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2}{\sigma_f^2 + \sum_{l=1}^u \sigma_l^2 + \sigma_e^2 + \sigma_d^2}. \quad \text{eqn 30}$$

As one can see in Equation 30, conditional  $R^2$  ( $R^2_{\text{GLMM}(c)}$ ) despite its somewhat confusing name can be interpreted as the variance explained by the entire model. Both marginal and conditional  $R^2_{\text{GLMM}}$  convey unique and interesting information, and we recommend they both be presented in publications.

In the case of a Gaussian response and an identity link (as used in LMMs), the linked scale variance and the original scale variance are the same and the distribution-specific variance is zero. Thus,  $(\sigma_e^2 + \sigma_d^2)$  reduces to  $\sigma_e^2$  in Equations 29 and 30. For other GLMMs, the link-scale variance will differ from the original scale variance. We here present  $R^2$  calculated on the link scale because of its generality: Equations 29 and 30 can be applied to different families of GLMMs, given the knowledge of distribution-specific variance  $\sigma_d^2$  and a model that fits additive overdispersion (e.g. MCMCglmm; Hadfield 2010). Importantly, when the denominators of Equations 29 and 30 include  $\sigma_d^2$  (i.e. for GLMM), both types of  $R_{\text{GLMM}}^2$  will never become 1 in contrast to traditional  $R^2$  (see also Table 1). Table 2 summarizes the specifications for binary/proportion data and count data, which are equivalent to Equations 22–25. The GLMM formulations presented in Table 2 for binomial GLMMs were first presented by Snijders & Bosker (1999). They also show that this approach can be extended to multinomial GLMMs where the response is categorical with more than two levels (Snijders & Bosker 1999; see also Dean, Nakagawa & Pizzari 2011). However, to our knowledge, equivalent formulas for Poisson GLMMs (i.e. count data) have not been previously described (for derivation, see Appendix 1).

As a technical note, we mention that for binary data the additive overdispersion is usually fixed to 1 for computational reasons, as additive dispersion is not identifiable (see Goldstein, Browne & Rasbash 2002). Furthermore, some of the  $R^2$  formulae include the intercept  $\beta_0$  (like in the case Poisson models for count data). In such cases,  $R^2$  values will be more easily interpreted when fixed effects are centred or otherwise have meaningful zero values (see Schielzeth 2010; see also Appendix 1). We further note that for Poisson models with square-root link and a mean of  $Y_{ijk} < 5$ , the given formula is likely to be inaccurate because the variance of square-root transformation of count data substantially exceeds 0.25 (Table 2; Nakagawa & Schielzeth 2010).

## Related issues

While an obvious advantage of using  $R_{\text{GLMM}}^2$  is its simplicity, one drawback is that  $R_{\text{GLMM}}^2$  does not provide information regarding variance explained at each level in a manner that  $R_1^2$  and  $R_2^2$  do. This shortcoming may be remedied by providing the proportion change in variance (PCV; Merlo *et al.* 2005a,b) as Supporting information in publications. Using Equations 22–25, PCV at three different levels can be expressed as:

$$C_\gamma = 1 - \frac{\sigma_\gamma^2}{\sigma_{\gamma 0}^2}, \quad \text{eqn 31}$$

$$C_\alpha = 1 - \frac{\sigma_\alpha^2}{\sigma_{\alpha 0}^2}, \quad \text{eqn 32}$$

$$C_\varepsilon = 1 - \frac{\sigma_\varepsilon^2}{\sigma_{\varepsilon 0}^2}, \quad \text{eqn 33}$$

where  $C_\gamma$ ,  $C_\alpha$  and  $C_\varepsilon$  are PCV at the level of groups, individuals and units (observations), respectively, and  $\sigma_{\gamma 0}^2$ ,  $\sigma_{\alpha 0}^2$  and

$\sigma_{\varepsilon 0}^2$  are variance components from the intercept model (i.e. Equation 22; PCV for additive dispersion,  $\sigma_{\varepsilon 0}^2$  can also be calculated by replacing  $\sigma_\varepsilon^2$  with  $\sigma_\varepsilon^2 + \sigma_d^2$ ). Proportion change in variance is in fact one of earliest proposed  $R^2$  measures for LMMs (Raudenbush & Bryk 1986; Bryk & Raudenbush 1992), although it can take negative values (Snijders & Bosker 1994). We think, however, that presenting PCV along with  $R_{\text{GLMM}}^2$  will turn out to be very useful, because PCV monitors changes specific to each variance component, that is, how the inclusion of additional predictor(s) has reduced (or increased) variance component at different levels. For example, if  $C_\gamma = 0.12$ ,  $C_\alpha = -0.05$  and  $C_\varepsilon = 0.23$ , the negative estimate shows that variance at the individual level has increased (i.e.  $\sigma_\alpha^2 > \sigma_{\alpha 0}^2$ ). Additionally, we refer the reader to Hössjer (2008) who describes an alternative approach for quantifying variance explained at different levels using variance components from a single model.

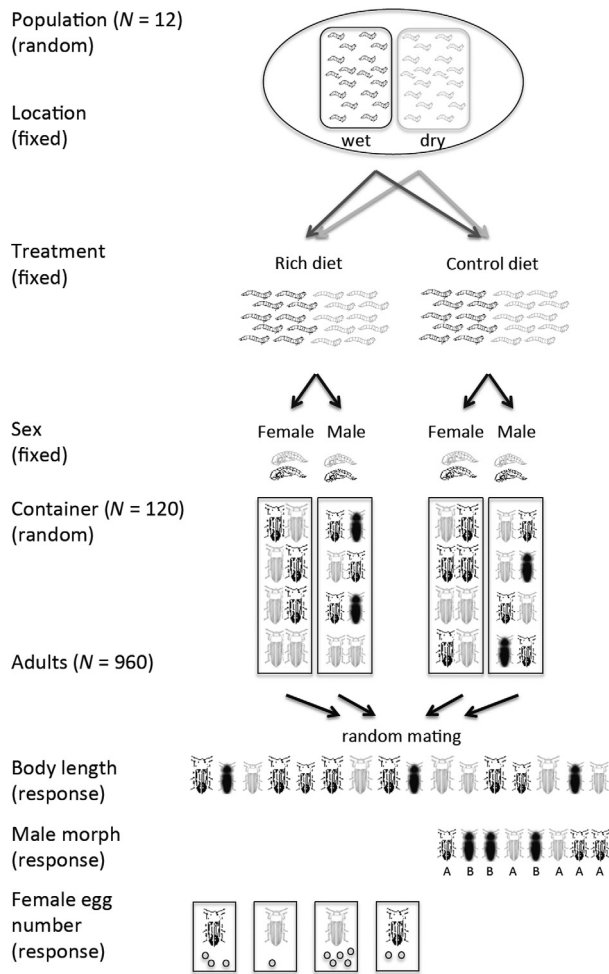
So far, we have only discussed random intercept models (e.g. Equations 22) not random-slope models where slopes are fitted for each group (usually along with random intercepts at each level; see Schielzeth & Forstmeier (2009) highlighting the necessity to fit random-slope models when the main interest is on data-level fixed effect predictors). Snijders & Bosker (1999) point out that calculating  $R^2$  like  $R_1^2$  and  $R_2^2$ , it is easy to do so for random intercept models, but for random-slope models is tedious (as variance components for slopes cannot be easily integrated with other variance components, e.g. Schielzeth & Forstmeier 2009). Snijders & Bosker (1999) mention that  $R_1^2$  and  $R_2^2$  obtained from random-slope models are usually very similar to those obtained from random intercept models, where the same fixed effects are fitted. Therefore, we recommend calculating  $R_{\text{GLMM}}^2$  (both marginal and conditional) from corresponding random intercept models for random-slope models, although PCV should be calculated for the random-slope models of interest.

## Worked examples

We will illustrate how the calculation of  $R_{\text{GLMM}}^2$  along with PCV using simulated datasets. Consider a hypothetical species of beetle that has the following life cycle: larvae hatch and grow in the soil until they pupate, and then adult beetles feed and mate on plants. They are a generalist species and so are widely distributed. We are interested in the effect of extra nutrients during the larval stage on subsequent morphology and reproductive success. Larvae are sampled from 12 different populations ('Population'; see Fig. 1). Within each population, larvae are collected at two different microhabitats ('Habitat'): dry and wet areas as determined by soil moisture. Larvae are exposed to two different dietary treatments ('Treatment'): nutrient rich and control. The species is sexually dimorphic and can be easily sexed at the pupa stage ('Sex'). Male beetles have two different colour morphs: one dark and the other reddish brown ('Morph', labelled A and B in Fig 1), and morphs are supposedly subject to sexual selection. Sexed pupae are housed in standard containers until they mature ('Container'). Each container holds eight same-sex animals from a single

**Table 2.** Examples of generalized linear mixed models (GLMMs) with binomial and Poisson errors (two random factors) and corresponding marginal and conditional  $R^2$

	Binary and proportion data	Count data	
Link function	Logit link	Probit link	Square-root link
Distribution-specific variance	$\pi^2/3$	1	0.25
Model specification	$Y_{ijk} = \text{Binomial}(m_{ijk}, p_{ijk}),$ $\text{link}(p_{ijk}) = \beta_0 + \sum_{h=1}^p \beta_h x_{ihjk} + \gamma_k + \alpha_{jk} + e_{ijk}$ $\gamma_k \sim \text{Gaussian}(0, \sigma_\gamma^2)$ $\alpha_{jk} \sim \text{Gaussian}(0, \sigma_\alpha^2)$ $e_{ijk} \sim \text{Gaussian}(0, \sigma_e^2)$	$Y_{ijk} = \text{Poisson}(\mu_{ijk}),$ $\text{link}(\mu_{ijk}) = \beta_0 + \sum_{h=1}^p \beta_h x_{ihjk} + \gamma_k + \alpha_{jk} + e_{ijk}$ $\gamma_k \sim \text{Gaussian}(0, \sigma_\gamma^2)$ $\alpha_{jk} \sim \text{Gaussian}(0, \sigma_\alpha^2)$ $e_{ijk} \sim \text{Gaussian}(0, \sigma_e^2)$	
Description	$Y_{ijk}$ is the number of ‘successes’ in $m_{ijk}$ trials by the $j$ th individual in the $k$ th group at the $i$ th occasion (for binary data, $m_{ijk}$ is 1), $p_{ijk}$ is the underlying (latent) probability of success for the $j$ th individual in the $k$ th group at the $i$ th occasion (for binary data, $\sigma_\alpha^2$ is 0).	$Y_{ijk}$ is the observed count for the $j$ th individual in the $k$ th group at the $i$ th occasion, $\mu_{ijk}$ is the underlying (latent) mean for the $j$ th individual in the $k$ th group at the $i$ th occasion.	
Marginal $R^2$ ( $R_{\text{GLMM}(m)}^2$ )	$R_{\text{logit}(m)}^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + \pi^2/3}$	$R_{\text{probit}(m)}^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + 1}$	$R_{\text{sqr}(m)}^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + 0.25}$
Conditional $R^2$ ( $R_{\text{GLMM}(c)}^2$ )	$R_{\text{logit}(c)}^2 = \frac{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + \pi^2/3}$	$R_{\text{probit}(c)}^2 = \frac{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + 1}$	$R_{\text{sqr}(c)}^2 = \frac{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\gamma^2 + \sigma_\alpha^2 + \sigma_e^2 + 0.25}$



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

population, but with a mix of individuals from the two habitats ( $N_{\text{container}} = 120$ ;  $N_{\text{animal}} = 960$ ). Three traits are measured after maturation: (i) body length of adult beetles (Gaussian distribution), (ii) frequencies of the two distinct male colour morphs (binomial or Bernoulli distribution) and (iii) the number of eggs laid by each female (Poisson distribution) after random mating (Fig. 1).

Data for this hypothetical example were created in R 2.15.0 (R Development Core Team 2012). We used the function *lmer* in the R package *lme4* (version 0.999375-42; Bates, Maechler & Bolker 2011) for fitting LMMs and GLMMs. We modelled three response variables (see also Table 3): (i) the body length with a Gaussian error ('Size models'), (ii) the two male morphs with the binomial error (logit-link function; 'Morph models') and (iii) the female egg numbers with the Poisson error (log-link function; 'Fecundity models'). For each dataset, we fitted the null (intercept/empty) model and the 'full' model; all models contained 'Population' and 'Container' as random factors; we included an additive dispersion term (see Table 2) in Fecundity models. The full models all included 'Treatment' and 'Habitat' as fixed factors; 'Sex' was added as a fixed factor to the body size model. Two kinds of  $R^2_{\text{GLMM}}$  and PCV for the

three variance components were calculated as explained above. The results of modelling the three different datasets are summarized in Table 3; all datasets and an R script are provided as online supplements (Data S1-4).

In all the three model sets, some variance components in the full models were larger than corresponding variance components in the null models (e.g.  $\sigma^2_{\alpha} < \sigma^2_{\epsilon_0}$ ). In Morph models, the sum of all the random effect variance components in the full model was greater than the total variance in the null model (c.f.  $(\sigma^2_{\epsilon} + \sigma^2_{\alpha}) < (\sigma^2_{\epsilon_0} + \sigma^2_{\alpha_0})$ ; see above; Snijders & Bosker 1994). All these patterns result in negative PCV values (see Table 3), while  $R^2_{\text{GLMM}}$  values never become negative. In Morph and Fecundity models,  $R^2_{\text{GLMM}(m)}$  values are relatively minor (8–10%) compared with  $R^2_{\text{GLMM}(c)}$  values. In Size models, on the other hand,  $R^2_{\text{GLMM}(m)}$  was nearly 40%. This was due to a very large effect of 'Sex' in body size model; in this model, the 'Treatment' and 'Habitat' effects together accounted for only c. 1% of the variance (not shown in Table 3). The variance among containers in the null Size model was conflated with the variance caused by differences between the sexes in the null model, as 'Sex' and 'Container' are confounded by the experimental design (single sex in each container; Fig. 1). A part of the variation assigned to 'Container' in the null model was explained by the fixed effect 'Sex' in the full model. Finally, it is important to note that both 'Treatment' and 'Habitat' effects were statistically significant in all the datasets in most cases (five out of six). Much of data variability, however, resided in the random effects along with residuals (additive dispersion) and in the distribution-specific variance. Note that differences between corresponding  $R^2_{\text{GLMM}(m)}$  and  $R^2_{\text{GLMM}(c)}$  values reflect how much variability is in random effects. Importantly, comparing the different variance components including that of the fixed factors within as well as between models, we believe, could help researchers gaining extra insights into their datasets (Merlo *et al.* 2005a,b). We also note that in some cases, calculating a variance component for each fixed factor may prove useful.

## Final remarks

Here, we have provided a general measure of  $R^2$  that we label  $R^2_{\text{GLMM}}$ . Both marginal and conditional  $R^2_{\text{GLMM}}$  can be easily calculated, regardless of the statistical package used to fit the models. While we do not claim that  $R^2_{\text{GLMM}}$  is a perfect summary statistic, it is less susceptible to the common problems that plague alternative measures of  $R^2$ . We further believe that  $R^2_{\text{GLMM}}$  can be used as a quantity of biological interest and hence  $R^2_{\text{GLMM}}$  might be thought of as being estimated from the data rather than calculated for a particular dataset. The empirical usefulness of  $R^2_{\text{GLMM}}$  as an estimator of the explained variance should still be tested in future studies. As with every estimator of biological interest, it is desirable to quantify the uncertainty around this estimate (e.g. 95% confidence interval, which could be approximated by parametric bootstrapping or MCMC sampling). As far as we are aware, such uncertainty estimates have not been considered for traditional



**Table 3.** Hypothetical mixed-effects modelling of the effects of nutrient manipulations on body length (mm) (Size models), male morphology (Morph models) and female eggs (Fecundity models);  $N_{\text{population}} = 12$ ,  $N_{\text{container}} = 120$  and  $N_{\text{animal}} = 960$ 

Model name	Size models Gaussian mixed models		Morph models Binary mixed models (logit link)		Fecundity models Poisson mixed models (log link)	
	Null Model	Full Model	Null Model	Full Model	Null Model	Full Model
Fixed effects	<i>b</i> [95% CI]	<i>b</i> [95% CI]	<i>b</i> [95% CI]	<i>b</i> [95% CI]	<i>b</i> [95% CI]	<i>b</i> [95% CI]
Intercept	14.08 [13.41, 14.76]	15.22 [14.53, 15.91]	-0.38 [-0.96, 0.21]	-1.25 [-1.96, -0.54]	1.54 [1.22, 1.86]	1.23 [0.91, 1.56]
Treatment (experiment)	-	0.31 [0.18, 0.45]	-	1.01 [0.60, 1.43]	-	0.51 [0.41, 0.26]
Habitat (wet)	-	0.09 [-0.05, 0.23]	-	0.68 [0.27, 1.09]	-	0.10 [0.001, 0.20]
Sex (male)	-	-2.66 [-2.89, -2.45]	-	-	-	-
Random effects	VC	VC	VC	VC	VC	VC
Population	1.181	1.379	0.946	1.110	0.303	0.304
Container	2.206	0.235	< 0.0001	0.006	0.012	0.023
Residuals (additive dispersion)	1.224	1.197	-	-	0.171	0.100
Fixed factors	-	1.809	-	0.371	-	0.067
PCV <sub>[Population]</sub>	-	-16.77%	-	-17.34%	-	-0.54%
PCV <sub>[Container]</sub>	-	89.37%	-	<-100%	-	-84.32%
PCV <sub>[Residuals]</sub>	-	2.21%	-	-	-	41.54%
$R^2_{\text{GLMM}(m)}$	-	39.16%	-	7.77%	-	9.76%
$R^2_{\text{GLMM}(c)}$	-	74.09%	-	31.13%	-	57.23%
AIC	3275	3063	602.4	573.1	902.7	811.9
BIC	3295	3097	614.9	594.0	920.4	836.9

CI, confidence interval; PCV, proportion change in variance; NA, not applicable/available; AIC, Akaike Information Criterion; BIC, Bayesian information criterion; ML, maximum likelihood; REML, restricted maximum likelihood; VC, variance components.

For full models, the intercept represents control, dry and female. 95% CI was estimated by assuming an infinitely large degree of freedom (i.e.  $t = 1.96$ ). For Size models, AIC and BIC values were calculated using ML but other parameters were from REML estimations (see the text for the reason).

$R^2$ . Perhaps, future studies can also investigate the usefulness of uncertainty estimates for  $R^2_{\text{GLMM}}$  and other  $R^2$  measurements.

We finish with a cautionary note that  $R^2$  should not replace model assessments such as diagnostic checks for heteroscedasticity, validating assumptions on the distribution of random effects and outlier analyses. Above, we presented  $R^2$  with the motivation of summarizing the amount of variance explained in a model that is suitable for the specific research questions and datasets. It should only be used on models that have been checked for quality by other means. It is also important to realize that the  $R^2$  can be large due to predictors that are not of direct interest in a particular study (Tjur 2009) such as the sex effect on body size in our example. Despite these limitations, when used along with other statistics such as AIC and PCV,  $R^2_{\text{GLMM}}$  will be a useful summary statistic of mixed-effects models for both biologists and other scientists alike.

## Acknowledgements

We thank S. English, C. Grueber, F. Korner-Nievergelt, E. Santos, A. Senior and T. Uller for comments on earlier versions and M. Lagisz for help in preparing Fig. 1. We are also grateful to the Editor R. O'Hara and two anonymous referees, whose comments improved this paper. T. Snijders provided guidance on how to calculate variance for fixed effects. H.S. was supported by an Emmy-Noether fellowship of the German Research Foundation (SCHI 1188/1-1).

## References

- Akaike, H. (1973) Information theory as an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* (eds B.N. Petrov & F. Csaki), pp. 267–281. Akademiai Kiado, Budapest.
- Bates, D., Maechler, M. & Bolker, B. (2011) lme4: linear mixed-effects models. R package, version 0.999375-42. <http://CRAN.R-project.org/package=lme4>.
- 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.
- Bryk, A.S. & Raudenbush, S. (1992) *Hierarchical Linear Models*. Sage, Newbury Park, CA.
- Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, Berlin.
- Cameron, A.C. & Windmeijer, F.A.G. (1996)  $R$ -squared measures for count data regression models with applications to health-care utilization. *Journal of Business & Economic Statistics*, **14**, 209–220.
- Cameron, A.C. & Windmeijer, F.A.G. (1997) An  $R$ -squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, **77**, 329–342.
- Claeskens, G. & Hjort, N.L. (2009) *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.
- Congdon, P.D. (2010) *Applied Bayesian Hierarchical Methods*. CRC, Boca Raton, FL.
- Dean, R., Nakagawa, S. & Pizzari, T. (2011) The risk and intensity of sperm ejection in female birds. *American Naturalist*, **178**, 343–354.
- Draper, N.R. & Smith, H. (1998) *Applied Regression Analysis*, 3rd edn. Wiley, New York.
- Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F. & Schabenberger, O. (2008) An  $R^2$  statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, **27**, 6137–6157.
- Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

- Gelman, A. & Pardoe, L. (2006) Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, **48**, 241–251.
- Goldstein, H., Browne, W. & Rasbash, J. (2002) Partitioning variation in multilevel models. *Understanding Statistics*, **1**, 223–231.
- Grueber, C.E., Nakagawa, S., Laws, R.J. & Jamieson, I.G. (2011) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, **24**, 699–711.
- Hadfield, J.D. (2010) MCMC methods for multi-response Generalised Linear Mixed Models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.
- Hamaker, E.L., van Hattum, P., Kuiper, R.M. & Hoijtink, H. (2011) Model selection based on information criteria in multilevel modeling. *Handbook of Advanced Multilevel Analysis* (eds J. Hox & J.K. Roberts), pp. 231–255. Routledge, New York.
- Hösjer, O. (2008) On the coefficient of determination for mixed regression models. *Journal of Statistical Planning and Inference*, **138**, 3022–3038.
- Kvålseth, T.O. (1985) Cautionary note about  $R^2$ . *American Statistician*, **39**, 279–285.
- Liu, H.H., Zheng, Y. & Shen, J. (2008) Goodness-of-fit measures of R(2) for repeated measures mixed effect models. *Journal of Applied Statistics*, **35**, 1081–1092.
- Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.
- Menard, S. (2000) Coefficients of determination for multiple logistic regression analysis. *American Statistician*, **54**, 17–24.
- Merlo, J., Chaix, B., Yang, M., Lynch, J. & Rastam, L. (2005a) A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology and Community Health*, **59**, 1022–1028.
- Merlo, J., Yang, M., Chaix, B., Lynch, J. & Rastam, L. (2005b) A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*, **59**, 729–736.
- Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.
- Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.
- Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, **85**, 935–956.
- Orelien, J.G. & Edwards, L.J. (2008) Fixed-effect variable selection in linear mixed models using  $R^2$  statistics. *Computational Statistics & Data Analysis*, **52**, 1896–1907.
- Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-effects Models in S and S-Plus*. Springer, New York.
- R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raudenbush, S. & Bryk, A.S. (1986) A hierarchical model for studying school effects. *Sociology of Education*, **59**, 1–17.
- Roberts, J.K., Monaco, J.P., Stovall, H. & Foster, V. (2011) Explained variance in multilevel models. *Handbook of Advanced Multilevel Analysis* (eds J. Hox & J.K. Roberts), pp. 219–230. Routledge, New York.
- Schielzeth, H. (2010) Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, **1**, 103–113.
- Schielzeth, H. & Forstmeier, W. (2009) Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, **20**, 416–420.
- Schielzeth, H. & Nakagawa, S. (2012) Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, doi: 10.1111/j.2041-210x.2012.00251.x.
- Schwarz, G.E. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Snijders, T.A. & Bosker, R.J. (1994) Modeled variance in two-level models. *Sociological Methods & Research*, **22**, 342–363.
- Snijders, T. & Bosker, R. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage, London.
- Snijders, T. & Bosker, R. (2011) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edn. Sage, London.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.R. & van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 583–616.
- Tjur, T. (2009) Coefficients of determination in logistic regression models – a new proposal: the coefficient of discrimination. *American Statistician*, **63**, 366–372.
- Vonesh, E.F., Chinchilli, V.P. & Pu, K.W. (1996) Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, **52**, 572–587.
- Xu, R.H. (2003) Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, **22**, 3527–3541.

Received 10 July 2012; accepted 12 September 2012  
Handling Editor: Robert B. O'Hara

## Appendix 1

### DERIVATION OF DISTRIBUTION-SPECIFIC VARIANCE ( $\sigma_a^2$ ) FOR POISSON DISTRIBUTIONS

When a random variable  $x$  is Poisson-distributed, the mean and variance of  $x$  is respectively:

$$E(x) = \lambda, \quad (\text{A1})$$

$$\text{var}(x) = \lambda. \quad (\text{A2})$$

The distribution of  $\ln(x)$  can be approximated by the natural logarithm of a log-normal distribution. Then, the variance of  $\ln(x)$  can be approximated as:

$$\text{var}(\ln(x)) = \ln\left(1 + \frac{\text{var}(x)}{E(x)^2}\right). \quad (\text{A3})$$

By substituting Equations A1 and A2 into Equation A3, we obtain:

$$\text{var}(\ln(x)) = \ln\left(1 + \frac{\lambda}{\lambda^2}\right). \quad (\text{A4})$$

Therefore,

$$\text{var}(\ln(x)) = \ln\left(1 + \frac{1}{\lambda}\right). \quad (\text{A5})$$

When we replace  $\text{var}(\ln(x))$  with  $\sigma_a^2$  and  $\lambda$  with  $\exp(\beta_0)$ , we obtain:

$$\sigma_a^2 = \ln\left(1 + \frac{1}{\exp(\beta_0)}\right). \quad (\text{A6})$$

Simulations (unpublished data, the authors) show that as  $E(x)$  approaches 0, this approximation becomes unreliable. Also,  $\exp(\beta_0)$  should be obtained either from a model with centred or scaled variables (*sense* Schielzeth 2010), or an intercept-only model while including all random effects. Note that the former approach may be limited when a model includes categorical variables.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Data S1.** Data file for Size model (BeetlesBody.csv).

**Data S2.** Data file for Morph models (BeetlesMale.csv).

**Data S3.** Data file for Fecundity models (BeetlesFemale.csv).

**Data S4.** R code for examples (R\_code\_examples.R).