# VARIANCE COMPONENTS ESTIMATION FOR CONTINUOUS AND DISCRETE DATA, WITH EMPHASIS ON CROSS-CLASSIFIED SAMPLING DESIGNS

*Brian R. Gray*

*Upper Midwest Environmental Sciences Center, U.S. Geological Survey, La Crosse, WI 54603 USA*

## INTRODUCTION

Variance components may play multiple roles (cf., Cox and Solomon 2003). First, magnitudes and relative magnitudes of the variances of random factors may have important scientific and management value in their own right. For example, variation in levels of invasive vegetation among and within lakes may suggest causal agents that operate at both spatial scales—a finding that may be important for scientific and management reasons.

Second, variance components may also be of interest when they affect precision of means and covariate coefficients. For example, variation in the effect of water depth on the probability of aquatic plant presence in a study of multiple lakes may vary by lake. This variation will affect the precision of the average depth-presence association.

Third, variance component estimates may be used when designing studies, including monitoring programs. For example, to estimate the numbers of years and of samples per year required to meet long-term monitoring goals, investigators need estimates of within- and among-year variances. Other chapters in this volume (Chapters 7 and 10) as well as extensive external literature outline a framework for applying estimates of variance components to the design of monitoring efforts. In particular, a series of papers with an ecological monitoring theme examined the relative importance of multiple sources of variation, including of variation in

Gray, B.R. 2012. Variance components estimation for continuous and discrete data, with emphasis on cross-classified sampling designs. In Gitzen R.A., J.J. Millspaugh, A.B. Cooper, and D.S. Licht, Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, pp. 200-227.

means among sites, years and site-years, for the purposes of temporal trend detection and estimation (Larsen et al. 2004, and references therein).

Due to the scientific and management value of variance components and their role in study design, the statistical properties of variance component estimators should be investigated. Specifically, investigators may explore bias, precision and other properties of estimators using variance component estimates derived from, for example, simulated datasets. Such evaluations may be undertaken to compare performance of alternative estimators in general, or to assess the performance of one or more estimators under the constraints of a specific study or monitoring design. Note that understanding the properties of variance component estimators represents an important aspect of wisely using variance component estimates for designing monitoring programs.

What, specifically, are variance components? Consider each of multiple measurements, $y_{ij}$, $i = 1, \ldots, n$, from each of multiple units ('sites'), $j = 1, \ldots, m$, and presume that $y_{ij}$ derives from the model $y_{ij} = \beta_{00} + u_{0j} + e_{ij}$, where $\beta_{00}$ denotes the expected value (grand mean) of $y_{ij}$, $\mathrm{E}[y_{ij}]$, $u_{0j}$ a site-specific effect on $y_{ij}$ and $e_{ij}$ residual variation. Presume further that the $u_{0j}$'s are independent and identically distributed (iid) with mean zero and variance $\sigma^2_{\text{site}}$, that the $e_{ij}$'s are iid (given the $u_{0j}$'s) with mean zero and variance $\sigma^2_e$, and that the $e_{ij}$'s are uncorrelated with the $u_{0j}$'s. Then it may be shown that the variance of $y_{ij}$, $\sigma^2_y$, is $\sigma^2_{\text{site}} + \sigma^2_e$. Hence, we see that $\sigma^2_{\text{site}}$ and $\sigma^2_e$ are components of $\sigma^2_y$, the variance of $y$ (Searle et al. 1992).

This chapter aims to help readers estimate variances associated with monitoring designs, to think critically about the value of variance estimates from small samples, and to emphasize that estimator performance may affect study design conclusions. The previous chapter in this volume (Chapter 8) provides a general introduction to defining and estimating components of variation relevant to monitoring and other ecological studies. In this chapter, I focus more specifically on estimation of variance components, using not only analysis of variance (ANOVA) but also maximum likelihood (ML)—whether full (FML) or restricted (REML), partial or quasilikelihood—and Bayesian approaches. More attention is paid to ML estimation because ANOVA is traditionally restricted to use with linear models, use of ANOVA is addressed in Chapter 8, and variance component estimation has been addressed in the literature more extensively using likelihood than Bayesian methods.

This chapter addresses variance components estimation from continuous, categorical, and count data. Variance-component estimation for categorical and count data involves considerations and approaches that may seem highly technical to some of this volume's readers. However, advanced readers may find this coverage useful, given that categorical and count data are commonly encountered in ecological studies. Throughout the chapter, general comments from the literature are augmented by results from simulations. Foci include estimation from small samples and from data from cross-classified sampling designs.

**A cross-classified model**

The ecological and temporal foci of this volume suggest the consideration of two-way cross-classified random effects sampling designs. Such designs permit multiple observations from, for example, multiple sites in each of multiple years. A candidate model equation may be represented by

$$y_{ijk} = \beta_{00} + u_{0j} + u_{0k} + u_{0jk} + \varepsilon_{ijk} \qquad (1)$$

where $y_{ijk}$ denotes the $i^{\text{th}}$ observation, $i = 1, \ldots, n_{jk}$, within the $jk^{\text{th}}$ site-year, $j = 1, \ldots, m_{\text{site}}$, $k = 1, \ldots, m_{\text{year}}$; $\beta_{00}$ denotes the expected value (grand mean) of $y_{ijk}$; $u_{0j}$, $u_{0k}$ and $u_{0jk}$ denote random site, year and site×year effects (respectively), and $\varepsilon_{ijk}$ denotes residual error. The random terms are presumed distributed with means zero, and variances $\sigma^2_{\text{site}}$, $\sigma^2_{\text{year}}$, $\sigma^2_{\text{site-year}}$ and $\sigma^2_{\varepsilon}$, respectively. Note that the model is a random effects model (covariates are absent) and that, from a multilevel (hierarchical) modeling perspective, the three group-level terms are at the same level while the observations occur at a lower level, the observation level. For further information about modeling outcomes with cross-classified random group effects, see Meyers and Beretvas (2006) and the multilevel modeling texts by Raudenbush and Bryk (2002), Goldstein (2003) and Hox (2010).

**Simulation models**

The above cross-classified model and the specified variance components provide a general framework relevant to many long-term monitoring efforts and other ecological studies (see also Chapter 7). However, the majority of studies in the literature assess estimator performance for designs less relevant to the focus of this volume. Further, many of those studies presume sample sizes that exceed those typically associated with ecological studies. Specifically, pilot studies designed to quantify variance components for use in designing monitoring studies will often be associated with cost and time constraints that yield both few observations per site-year group ($n$) and few years ($m_{\text{year}}$), and often relatively few sites ($m_{\text{site}}$). For these reasons, I augmented findings from the literature with those from data simulated under (*1*) with relatively small sample size assumptions.

For most simulation scenarios conducted for this chapter, site, year, site×year and residual effects were distributed as normal random variates, with means 0 and variances as defined under (*1*). For a few scenarios, however, random effects were generated from non-normal distributions to explore the effects of violating normal distributional assumptions common to most estimation methods. In other scenarios, I altered the relative magnitudes of variance components, sample sizes ($m_{\text{site}}$, $m_{\text{year}}$, or $n$) and other factors relevant to specific data types (e.g., a naïve Poisson distributional assumption for conditional, negative binomial-distributed count data). With the exception of those scenarios specified as unbalanced, numbers of observations per site-year were constant (i.e., $n_{jk} = n$ for all $jk$) and small ($n$ = 5 or 10).

In all simulations, site and year effects were independent, and site-year ($u_{0jk}$) effects and residual errors ($\varepsilon_{ijk}$) conditionally so. In practice, real data may be more complex—site effects may be spatially correlated, annual effects may be temporally correlated and conditional site-year effects may be spatio-temporally correlated. Given small sample sizes, however, such concerns will not easily be addressed. The appropriateness of these simplifying assumptions does not affect the conclusion that small sample sizes may jeopardize accurate and precise estimation of variance components. Owing to the temporal focus of this volume and because variation in year effects may most directly affect temporal inferences (Larsen et al. 2004), I provide results primarily for $\sigma^2_{\text{year}}$ estimators; results for other estimators may be obtained using code supplied at the volume's web site.

## VARIANCE COMPONENTS ESTIMATION AND LINEAR MODELS

Variance component estimation using linear models with random effects is a relatively developed topic, dating to at least 1861 (Airy 1861, Scheffé 1956) and represents the focus of many texts (e.g., Searle et al. 1992, Rao 1997, Cox and Solomon 2003). Variance components have traditionally been estimated using ANOVA and, beginning in the 1960s and 1970s, increasingly with ML and REML (Searle et al. 1992). More recently, computational advances have stimulated interest in Bayesian methods of estimating variance components (Draper 2008). See Appendix 1 for description of ANOVA, ML and Bayesian methods of estimating variance components using linear models.

The maturity of variance components estimation using linear models, combined with the common problem for ecological studies of small sample sizes, motivated this chapter's focus on issues associated with sample size. For variance component models, the most relevant sample size is typically defined relative to the variable of interest—years if among-year variance is of interest, sites if among-site variance is of interest, and so on. This perspective often leads to greater attention being paid to the estimation of group-level than to observation-level components: Sample sizes at group levels are typically much smaller than sample sizes at the observation level, although exceptions may occur when $\sigma_e^2$ is estimated by group or by other data subset.

## Bias of variance component estimators

Bias is commonly considered when evaluating the performance of estimators. However, the importance of bias as a criterion for evaluating estimators is often treated as less important for variance component than for fixed effect estimators. For ANOVA, this reflects that the favorable property of the unbiasedness of ANOVA variance estimators comes at the expense of accepting negative variance estimates while the common solution of setting those negative estimates to zero yields positive bias. However, FML and REML estimators of group-level variance components are often biased (Searle et al. 1992). Another concern is that bias reflects the idea of multiple repetitions of the same study (Searle et al. 1992, section 2.3; see also Chapter 3). Of course, variance components are often estimated from observational data—where replication may be empirically unattainable.

Despite the above issues, bias of variance component estimators remains an important, albeit qualified, concern. For example, Browne and Draper (2000, 2006) documented negligible bias with REML for group-level variance components but relatively large negative biases with FML at 6 and 12 groups (~-21% and -11%, respectively); biases associated with Bayesian estimation using Gibbs sampling depended on prior and posterior summary (mean, median, mode) but were generally greater in magnitude than those documented for ML. Biases decreased in magnitude for both ML and Bayesian methods when numbers of groups increased to 24 and 48. Raudenbush (2008) argued that, in a likelihood setting, small numbers of observations per group may yield group-level variance component estimates that are biased low; by contrast, the residual error variance, $\sigma_{\varepsilon}^{2}$, which is estimated from all observations from all groups, will generally be accurately estimated (and, for this reason, estimation of $\sigma_{\varepsilon}^{2}$ will represent a minor focus of this chapter).

The above findings from the literature are congruent with those from the simulations conducted for this chapter using small sample sizes and cross-classified random effects designs (Table 1). Relative biases associated with ANOVA and REML estimators of $\sigma_{\text{year}}^{2}$ were ignorable when $m_{\text{year}}$ = 10 and 20, and modest when $m_{\text{year}}$ = 3 and 5 (scenarios 1 – 4; ANOVA, REML). The unexpected finding of lower bias for the ANOVA and REML estimators at years = 3 is an artifact of the treatment of negative values: Negative ANOVA estimates were, as is commonly done, set to zero while REML (and FML) estimation ensures nonnegative estimates by setting negative solutions to zero. (For all estimation methods, the proportion of nonpositive variance estimates under the years = 3 scenario was ≥10%.) In contrast to the ANOVA and REML cases, biases for FML were severe at small sample sizes (scenarios 1 – 2). As noted in Appendix 1, the bias of the FML estimator is approximately -*k* / sample size (provided the proportion of zero estimates is small), where *k* denotes the number of linearly independent fixed covariate terms. For this and other reasons, REML is typically preferred over FML for estimating variance components. See McCulloch and Searle (2001) for a fuller treatment of this topic.

The effects of manipulating attributes unrelated to year on bias of $\widehat{\sigma^2_{\text{year}}}$ were typically minor. Relative to that of the $m_{\text{year}} = 5$ reference (scenario 2), bias improved modestly when $n$ was doubled but was not improved by doubling the number of sites (scenarios 5, 6). Bias improved trivially when $\sigma^2_\varepsilon$ was substantially decreased (scenario 9); such decreases may occur with sampling protocol refinement or adjustment for covariates that vary at the observation scale. As also described by Cools et al. (2009), the effects on bias were not severe when datasets were unbalanced (scenarios 7-8).

**Precision**

The topics of precision and of precision comparisons across estimators of group-level variances, particularly when numbers of groups are few, have received limited attention. Browne and Draper (2006) documented improvements in the precision of group-level variance component estimators as numbers of groups increased, when datasets were balanced, and as the among-group variance : within-group variance ratio (i.e., the intra-class correlation coefficient) increased. While Monte Carlo standard deviations for scenarios with few groups ($m$ = 6, 12) were typically smaller for FML than for REML and for both relative to Bayesian estimators, among-method differences narrowed or disappeared when number of groups reached 24. Coverage by Bayesian credible intervals approximated nominal coverage with as few as 12 groups. See Browne and Draper (2000) for a related study, and Gelman and Hill (2007, section 19.6) for comments on selection of priors for estimating group-level variances in studies with few groups.

The precisions of $\sigma^2_{\text{year}}$ estimators under the small sample-cross-classified design simulation study of this chapter were poor, with Monte Carlo standard deviations approaching or exceeding in value those of the corresponding variance component point estimates (Table 1). Substantial improvements in precision occurred only with increased numbers of years, chiefly when years reached 10 and 20 (scenarios 3, 4). As may be expected, precision was largely unaffected by increases in numbers of sites sampled, numbers of observations sampled per site-year, and changes in $\sigma^2_{\text{site}}$ and $\sigma^2_{\varepsilon}$ (scenarios 5, 6, 9, 10). Precision deteriorated modestly when datasets were unbalanced, with less deterioration when cells contained one rather than no observations (scenario 7, 8; cf., Cools et al. 2009).

Precision concerns when numbers of groups are small may be addressed by treating inferences as provisional pending more years of monitoring (presuming variance component magnitudes remain approximately constant) and by decreasing variances by adjusting for covariates. When planning for new studies, investigators should consider that available or pilot variance component estimates derived from few groups may not only be biased but also be estimated imprecisely. The importance of these concerns may be addressed using simulations (Appendix 4).

**Mean squared error**

Bias and precision may be quantified in combination using a number of methods, including that of root mean squared error (RMSE = $\sqrt{\text{bias}^2 + \text{variance}}$). For the simulation study and as may be inferred from the preceding discussion, relative RMSEs for $\widehat{\sigma^2_{\text{year}}}$ decreased over the 3, 5, 10 and 20 year series (0.35, 0.24, 0.16 and 0.10 units, respectively), with largest decreases occurring when numbers of years were relatively small (ANOVA/REML, Table 1).

**Distributional assumptions**

Group-level random effects such as those associated with equation (1) are typically assumed to be normally distributed. Maas and Hox (2004) addressed the reasonableness of this assumption by intentionally treating skewed (chi-squared with one degree of freedom), heavy-tailed (uniform) and light-tailed (Laplace) group effects as normally distributed. The effects of these distributional violations on variance component point estimates from REML were modest, with largest reported relative bias (12%) associated with their study's smallest sample size scenario (30 groups with 5 observations per group).

Modest effects of distribution violations were also seen when group effects in this chapter's simulation study were random uniform but were naively modeled under a normal distributional assumption (scenarios 12, 13, Table 1). At years = 5 and relative to the balanced case, mean variance point estimates were larger while Monte Carlo standard deviations decreased (scenarios 2, 12). Increasing the number of years to 20 narrowed differences among point estimates but widened relative differences among precision estimates (scenarios 3, 13).

The possibility that distributional assumptions have not been essentially met should lead investigators to treat variance component estimates from small samples with additional caution.

Snijders and Berkhof (2008) address diagnostic checks for model residuals at group and observation scales. Unfortunately, the reasonable satisfaction of distribution assumptions may be unclear when numbers of groups are small; in these cases, the practical importance of distributional-assumption failure may be addressed using simulations.

**Variance partition coefficients**

Group-level variance components are often reported as proportions of total variation. Such proportions are termed variance partition coefficients (VPCs), and may be used for both study planning and scientific purposes (Kincaid et al. 2004; Browne et al. 2005; Hox 2010, chapter 12). For linear models without covariates with random coefficients, VPCs are equivalent to the intra-cluster or intra-class correlation coefficient (ICC) familiar from survey statistics (see also Chapter 8). For these models (and given a single group-level variance term), ICC = VPC = 0 denotes independent observations while ICC = 1 denotes an absence of variation within groups; most studies with grouped data yield ICC estimates between 0 and 1.

The two-way cross-classified random effects model associated with (*1*) yields VPCs that correspond to correlation between two outcomes from the same site and year, same site but different year, and same year but different site, respectively (cf., Raudenbush and Bryk 2002):

$$\mathrm{VPC}_{\text{site-year}} = \frac{\sigma^2_{\text{site}} + \sigma^2_{\text{year}} + \sigma^2_{\text{site-year}}}{\sigma^2_{\text{site}} + \sigma^2_{\text{year}} + \sigma^2_{\text{site-year}} + \sigma^2_{\varepsilon}}$$

$$\mathrm{VPC}_{\text{site}} = \frac{\sigma^2_{\text{site}}}{\sigma^2_{\text{site}} + \sigma^2_{\text{year}} + \sigma^2_{\text{site-year}} + \sigma^2_{\varepsilon}} \qquad (2)$$

$$\mathrm{VPC}_{\mathrm{year}} = \frac{\sigma^2_{\mathrm{year}}}{\sigma^2_{\mathrm{site}} + \sigma^2_{\mathrm{year}} + \sigma^2_{\mathrm{site\text{-}year}} + \sigma^2_{\varepsilon}}$$

The effects of sample size and balance on VPC estimation are illustrated using $\widehat{\mathrm{VPC}}_{\mathrm{year}}$, as derived from simulation study estimates. For scenarios 1 – 8, biases in mean $\widehat{\mathrm{VPC}}_{\mathrm{year}}$ from ANOVA and REML estimation were <15% (Table 2). However, biases in the medians of the typically right-skewed $\widehat{\mathrm{VPC}}_{\mathrm{year}}$ exceeded |20%| when years ≤5 and, for ML, reached -32% and -47% when years = 5 and 3, respectively. The properties of VPC estimators for linear models with multiple random effects is poorly addressed in the literature; the results from the current study suggest caution when estimating VPCs using ML estimates from few groups.

## VARIANCE COMPONENT ESTIMATION AND CATEGORICAL DATA

### Binomial data

The binomial distribution derives from the idea of a Bernoulli trial, which is defined as an experiment with only two exclusive outcomes (e.g., female or male, hatched or failed to hatch). Let *p* denote the probability of success (e.g., female, hatched). A binomial experiment is defined as a sequence of identical and independent Bernoulli trials, each with success probability *p*; the binomial outcome, a count, represents the number of successes. Data that might be treated as binomial include counts of the number of females in a litter or of the number of hatched eggs within nests.

Models of binomial data typically address fundamental characteristics of those data, including that expected responses (probabilities) are nonnegative and don’t exceed one and that sampling variances are functions of the expected responses. These characteristics are typically viewed as precluding the use of standard linear models with binomial data. Instead, binomial data are typically modeled via the inverses of cumulative distribution functions, including those of the standard logistic, standard normal and standard extreme value distributions. Use of these functions yield logistic (logit), probit and complementary log-log models, respectively (McCullagh and Nelder 1989).

Binomial data are commonly modeled using generalized linear models (GLMs). GLMs generalize linear regression by incorporating a linear predictor while also permitting linear and nonlinear associations between response means and predictors. The combination of a linear predictor and model nonlinearity is permitted by use of a function that links the expected value of the response with the linear predictor. Link functions commonly used for binomial data include the logit, probit and cumulative log-log functions mentioned above. The addition of random effects to a GLM yields a generalized linear mixed model (GLMM). Further descriptions of GLMs and GLMMs are provided by Lindsey (1997) and McCullough and Searle (2001).

A two-way random effects cross-classified GLMM equation for the expected value of a binomial probability, $p_{jk}$, with logit or log odds link is:

$$\text{logit}(p_{jk}) = \log\left(\frac{p_{jk}}{1-p_{jk}}\right) = \beta_{00} + u_{0j} + u_{0k} + u_{0jk} \qquad (3)$$

The terms in the linear component of (*3*) correspond to those in (*1*): $\beta_{00}$ denotes a grand intercept, and $u_{0j}$, $u_{0k}$ and $u_{0jk}$ denote, respectively, random site, year and site×year effects. Analogous to the case with (*1*), the random terms are presumed distributed (albeit now on the logit scale) with means zero and variances $\sigma^2_{\text{site}}$, $\sigma^2_{\text{year}}$, and $\sigma^2_{\text{site-year}}$, respectively. The omission of index "*i*" [cf., (*1*)] arises from treating grouped Bernoulli outcomes as binomial counts, with binomial index *n*; the error term, denoted $\varepsilon_{ijk}$ in (*1*), is subsumed in the nominal distributional assumption (e.g., logistic).

An important change from the linear models case is that variances for binomial data may be estimated on both link and measurement scales, whereas for linear models these scales are equivalent. As with linear models, variance estimates on the link scale may be important for evaluating properties of variance component estimators and for study planning purposes. On the other hand, scientific interest will often focus on the scale at which the response occurs and is measured. For standard models of binomial data, we may then speak about variances on both the link (e.g., logit) and measurement (or probability) scales; both are addressed in the following sections. Also, I focus principally on logit normal models. The use of link functions other than the logit for GLMMs of binomial data are comparatively rare. From a variance components estimation perspective, Callens and Croux (2005) evaluated the complementary log-log function while Browne et al. (2007) provided variance component models using mixed effects probit models. Although I here focus on binomial data from hierarchical designs, the analysis of multinomial data—whether ordered or nominal—from hierarchical designs may be treated similarly to the case with binomial data; this topic is addressed in detail by Fielding (2003), Hedeker (2008) and Hox (2010).

**Variance component estimation on the link scale**

GLMMs of binomial data may be fitted using multiple methods, including linear approximations.

Linearization methods employ Taylor series expansions and a linear model of the resulting data approximations. These methods are broadly divided into those that use Taylor series expansions about the expected marginal value of random effects (i.e., zero) and those that expand about subject-specific predictions; these methods are often defined as marginal quasi-likelihood (MQL) and penalized or predictive quasi-likelihood (PQL), respectively. Linearizations may be first- or second-order (MQL1 and MQL2, and PQL1 and PQL2, respectively); linearized "data" may also be evaluated using residual maximum likelihood (RMQL, RPQL).

While linearization methods are commonly used, they may yield negatively biased variance component estimates when used with binomial outcomes with small numbers of observations per group, *n*, or when among-cluster variances are large (>0.5 on the logit scale; Goldstein and Rasbash 1996, Pinheiro and Chao 2006). Other studies confirm these concerns, and also demonstrate superior variance component estimation by PQL1 relative to that of MQL1 and of PQL2 and MQL2 relative to PQL1 and MQL1, respectively (Rodriguez and Goldman 1995, 2001; Goldstein and Rasbash 1996; Guo and Zhao 2000; Callens and Croux 2005; Brown and Draper 2006; Diaz 2007).

GLMMs may also be fitted by approximating the marginal likelihood using numerical approximations. These methods, which include Laplace estimation and Gaussian quadrature, have yielded variance component biases that are smaller in magnitude than those seen with MQL and PQL (see below; Pinheiro and Chao 2006, Diaz 2007, Moineddin et al. 2007). Further, these methods yield likelihood estimates and so permit the calculation of likelihood-based information criteria and likelihood ratio tests. Downsides are that these methods are more computationally intensive than their MQL and PQL counterparts, and may not be suitable for designs with multiple random components. In particular, Gaussian quadrature is precluded for models with crossed random factors (McCulloch and Searle 2001), including (*3*). Another consideration is that, due to improved relative precision of the PQL estimator, PQL may outperform Laplace estimation of variance components under a mean squared error criterion (Callens and Croux 2005, Diaz 2007).

The relative performance of Bayesian estimators of variance components from hierarchical models of binomial data has seen relatively little study. Findings to date suggest the potential comparability or superiority of Bayesian estimators relative to Gaussian quadrature estimators, and the superiority of both relative to MQL and PQL estimators (Rodriguez and Goldman 2001, Browne and Draper 2006). Fitting hierarchical models to binomial data using Bayesian methods is described by Gelman and Hill (2007).

We earlier saw with linear models that violation of distributional assumptions for group-level effects may be associated with biased variance component estimators (Table 1). Whether such biases might be nonignorable under reasonable assumptions for GLMMs has been poorly studied. Moineddin et al. (2007) described distribution-related biases that were resolved with increased *n* when group-level effects were random uniform but not when those effects were $t_{(df=3)}$-distributed. Unfortunately, the Moineddin et al. study employed models that were relatively complex given many of the considered sample sizes. The topic of distributional assumptions in the context of variance component estimation from binomial data requires further research.

Results from the simulation scenarios conducted with binomial data for this chapter confirm the expectation of bias for variance components estimated from few groups (scenarios 1 – 3; Table 3). At years = 5, biases were typically substantial in relative magnitude (range: -43% to -17%) and also varied substantially among estimation methods. At years = 20, however, biases were relatively modest (≤|17%|) and varied less among methods (-17% to -7%). Variation in the magnitude of bias among estimation methods was generally in the order MQL1 > PQL1 > Laplace > RPQL1 > Markov chain Monte Carlo (MCMC). Second-order linearization methods were not evaluated in this simulation study.

As with the linear model simulations, improvements in $\widehat{\sigma^2_{\text{year}}}$ biases were typically no more than modest when attributes other than number of years were varied (scenarios 4-8). In particular, relative biases were similar regardless of whether $\sigma^2_{\text{year}}$ was moderate (0.3 units) or large (1 unit; scenario 8).

Precisions of $\sigma^2_{\text{year}}$ estimators were generally poor, with most improvement seen when numbers of years were increased (compare Monte Carlo standard deviations, Table 3). For example, Monte Carlo standard deviations decreased by ~40% as numbers of years increased from 5 to 20 (scenarios 1, 3) but by roughly half that amount when median $p_{ij}$ increased from 0.12 ($\beta_{00}$ = -2; scenario 1) to 0.5 ($\beta_{00}$ = 0; scenario 6). Note that ranking of precisions by estimation method generally followed $\widehat{\sigma^2_{\text{year}}}$ and, hence, typically appeared best for MQL1 and PQL1, the estimators with the greatest negative biases. For PQL1 and from an MSE perspective, the improved relative precision typically trumped the importance of increased bias—leading to lower MSEs for this method than for other likelihood-based methods. This latter finding is consistent with those reported by Callens and Croux (2005) and Diaz (2007).

The above evidence suggests that investigators should treat estimates of group-level variances from binomial outcomes with few groups as imprecise and, for likelihood-based estimators, as negatively biased. The practical importance of these concerns may be addressed using simulation studies prior to the use of those estimates for study planning.

**Variance Component Estimation on the Probability Scale**

As already mentioned, variance components from GLMMs will typically be more meaningful to scientists and managers when reported on measurement scales. For example, consider a hypothetical example where hatch success (eggs hatched / eggs laid) of a songbird is measured at each of *j* sites during each of *k* years. For ecological interpretation, probability-scale estimates (e.g., of year-to-year variation in mean probability of success) would typically be more useful than logit-scale variance estimates. Methods for estimating variance components on measurement scales include simulations and, for categorical outcomes, a latent variable method (Goldstein et al. 2002). See Appendix 2 for an overview of these approaches. As variance component estimation methods on measurement scales have seen little study and appear to have rarely been employed with ecological data (see Li et al. 2008 for an exception), comments will center on results from this chapter's simulations.

Concerns associated with estimation of variance components on the probability scale include that expected values may be method dependent, and that investigators have not settled on a protocol for method selection in all cases. For cross-classified models, a further concern is that methods have not been evaluated in the peer-reviewed literature. For these reasons, I infer bias in $\widehat{\sigma^2_{\text{year}}}$ primarily by comparison with estimates from the model with the largest number of years.

From the simulation study, we see that $\widehat{\sigma^2_{\text{year}}}$ on the probability scale increased with number of years (scenarios 1 – 3, Table 4). As explained above, this finding is concordant with an assumption of declining bias in $\widehat{\sigma^2_{\text{year}}}$ with increasing number of years, a finding that also parallels that seen with the logit scale estimates (Table 3). As also seen with the logit scale and linear model estimates, biases in $\widehat{\sigma^2_{\text{year}}}$ were largely unaffected by changes in *n*, sites and $\sigma^2_{\text{site}}$

(scenarios 4, 5, 7; Table 4).

Variance component estimates on the probability scale for the simulation study were generally small (Table 4). This partly arose from the constraint on variability associated with the small median probability [i.e., median $p_{jk}$ = antilogit(-2) = 0.12]. As evidence, note that, while holding variance components on the logit scale constant, the $\sigma^2_{\text{year}}$ estimate on the probability scale more than doubled when the median probability increased to ~0.50 (scenario 6). As this example demonstrates, comparing variance components on the probability scale will be challenging unless those components have means or medians that are comparable (i.e., that are roughly equal distances from $p$ = 0.5).

Estimating variances on inverse link scales becomes more challenging when, as may often be the case for ecologists, covariates are present (Goldstein et al. 2002, Li et al. 2008). For example, Browne et al. (2005) used multi-covariate, multi-level logistic models of literacy within states and districts within states in India to infer the importance of addressing literacy needs of females in rural areas and of whole states with low literacy rates (rather than districts with low rates). This seemingly promising approach has seen essentially no use in the ecological literature.

**Variance partition coefficients for binomial models**

While VPCs from GLMMs may be estimated on link and measurement scales, VPCs on the link scale seem most useful for study planning purposes (cf., Gray and Burlew 2007). A challenge to estimating VPCs on the link scale is that of defining the variation of an observation on that scale. Gray and Burlew (2007) addressed this issue for count models using the delta method while, for categorical outcomes, the latent variable approach mentioned above may also be used.

VPC estimates on the probability scale may vary by estimation approach, with VPCs calculated using the simulation approach often yielding smaller estimates than those estimated using the latent variable approach (Goldstein et al. 2002, Browne et al. 2005, Li et al. 2008). These findings were also seen with this chapter's simulation study (Table 5). A major difference between the simulation and latent variable approaches is made clear by the behavior of VPC estimates as $\beta_{00}$ increased from -2 to 0 (scenarios 1, 6). VPCs estimated using the latent variable approach varied little among $\beta_{00}$ values because $\widehat{\sigma^2_{\text{year}}}$ varied little *on the logit scale* with $\beta_{00}$. By contrast, VPCs estimated using the simulation approach varied substantially with $\beta_{00}$ because $\widehat{\sigma^2_{\text{year}}}$ varied substantially *on the probability scale* with $\beta_{00}$. The simulation approach will often be preferred over the latent variable approach (Goldstein et al. 2002, Li et al. 2008).

**Categorical data and classification errors**

Ecological data that are categorical often incorporate classification errors, a concern that is relevant to variance components estimation when classification error probabilities are heterogeneous. For species detection / nondetection data, a misclassification occurs when a species that is present is not detected. Classification errors may also occur for multicategory outcomes such as frog calling indices. Failure to address classification errors may yield erroneous inferences for both dichotomous and ordered multinomial outcomes (Royle and Link 2005, Mackenzie et al. 2006, Holland et al. 2010, Holland and Gray 2011).

Addressing classification errors becomes challenging when the probabilities of those errors vary among sampling units. For dichotomous outcomes, failure to address such variation will typically yield biased estimators of both the probability of detection and the probability of the true state (Royle and Dorazio 2008). The analogue of this problem has also been demonstrated for multinomial models (Holland and Gray 2010).

In many settings, variation in a parameter would typically be addressed by allowing the parameter to vary according to a mixing distribution (e.g., the logit-normal distribution employed above). For two-category models like those commonly used to estimate occupancy, however, the probability of the state variable of interest (e.g., site occupancy) is not identifiable across mixing distributions, with the practical importance of this concern increasing as the mean detection probability decreases and as unexplained among-site variation in the classification error increases (Royle 2006). It seems reasonable to expect that the same problem will apply to multinomial abundance models. The result is that, while variation in misclassification parameters is estimable, such estimation may not always qualitatively improve inferences on the variable of interest.

## VARIANCE COMPONENT ESTIMATION AND COUNT DATA

As with their categorical counterparts, count data are typically modeled using GLMs and, given random group effects, using GLMMs. GLMs and GLMMs of counts typically presume counts are distributed as Poisson or negative binomial (NB) random variates (conditional on any fixed or random effects). An NB distributional assumption allows for the common case where sampling variation exceeds that expected under a Poisson assumption. GLMMs of counts are typically formulated using a log link, a link that conveys the advantage of ensuring predicted means are nonnegative.

The estimation problems described above for modeling clustered categorical data have not generally been addressed for count models. Consequently, few studies have compared variance component estimators for count models as a function of estimation method. A modest exception is found in an adaptive Gaussian quadrature study by Pinheiro and Chao (2006). As part of that study, the authors evaluated RPQL, Laplace and adaptive Gaussian quadrature estimators of fixed effects and variance components under a single simulation scenario. Data arose from many groups (300) of two conditional Poisson outcomes each (i.e., $n = 2$); the median group mean was large (~17) while the among-cluster variance on the log scale was low (0.09 units). Given this setting, fixed effects and variance components were, as might be expected, estimated with ignorable bias (<1.7%) by all three methods. For comparisons among GLMMs and a GLMM elaboration see Lee and Nelder (2001) and references therein.

Simulations of count data for this chapter suggest wide differences among estimators in performance when outcomes were NB distributed and, to lesser degree, when outcomes were Poisson distributed (Table 6). Simulations were conducted under the log-link analogue of model equation (*3*):

$$\log(\mu_{jk}) = \beta_{00} + u_{0j} + u_{0k} + u_{0\,jk} \tag{4}$$

Convergence rates for PQL1 and RPQL1 models were poor when outcomes were NB-distributed (Table 6: upper panel) and often little better when outcomes were Poisson distributed (Table 6: lower panel). Restricting subsequent attention to MQL1 and Laplace estimators, bias and precision was typically best under Laplace estimation. Note that the MQL1 estimators of the NB dispersion parameter and of $\sigma^2_{\text{site-year}}$ were typically positively biased and estimated imprecisely. For scenario NB3, for example, the MQL1 estimator (with Monte Carlo standard deviations) yielded NB dispersion parameter and $\sigma^2_{\text{site-year}}$ estimates of 1.00 (0.24) and 0.37 (0.17)—with true values 0.5 and 0.15, respectively. By contrast, the corresponding Laplace estimates were 0.50 (0.03) and 0.15 (0.02), respectively. Evidence of positive bias in $\sigma^2_{\text{site-year}}$ was also seen with MQL1 models of Poisson outcomes. As also seen with the binomial models, variation in attributes other than number of years typically led to only minor effects on the bias and precision of $\sigma^2_{\text{year}}$ (Table 6).

Naïvely treating NB-distributed outcomes as Poisson-distributed yielded generally minor changes in $\sigma^2_{\text{year}}$ estimates (compare scenarios NB1, NB10; Table 6). Instead, the effects of misspecifying the conditional generating distribution was seen in inflated $\sigma^2_{\text{site-year}}$ estimates; these estimates were, on average, biased high by 59% and 68% for MQL and Laplace estimation, respectively (relative to those derived under the conditional NB assumption). These findings reflect the wisdom of addressing extra-Poisson variation in models of count outcomes.

While the results described in this section provide modest evidence in favor of Laplace estimation, readers interested in VC estimation from count data should also consider MCMC and, for fully nested models, adaptive Gaussian quadrature.

The contamination of count data with structural zeroes has been a common concern for modelers and ecologists (Cameron and Trivedi 1998, Gray 2005). This topic of zero inflated count data has been addressed from a model specification rather than estimation method perspective for grouped count data by Min and Agresti (2005), Lee et al. (2006), Moghimbeigi et al. (2008) and Gray et al. (2010).

**Variance components from count models on the measurement scale**

Interest in variance components (and in VPCs in particular) on the measurement scale for count outcomes has lagged that for binomial outcomes. An implicit exception is Goldstein et al. (2002). While Goldstein et al. focused on binomial outcomes, those authors noted that three of the four methods they proposed—model linearizations using Taylor series approximations, a normal distributional approximation, and simulations—might be considered for use with other nonlinear models. Stryhn et al. (2008) considered the above three methods from a count perspective, and also considered exact formulae (calculated using integration formulae for exponential functions). They found the simulation and integration methods yielded results that were not only similar but also superior to those derived by the linearization and normal approximation methods. The Stryhn et al. paper is sparse on methodology and results (but see presentation at http://people.upei.ca/hstryhn/iccpoisson.ppt); a more detailed follow-up paper is expected (H. Stryhn, personal communication, 25 Aug 2009).

Estimating VPCs on the measurement scale for cross-classified random effects count models has not apparently been addressed in the literature. One approach that seems promising but which has apparently not been evaluated in the published literature would adapt the provisional method supplied in Appendix 2 for counts by substituting the exponential for the logistic function in step 1 and by modifying $\mathrm{var}(y_{ijk})$ in step 5 to reflect the assumed count distribution.

## SOFTWARE FOR STUDY DESIGN

Given the focus of this book, many readers may be interested in variance component estimation for the purposes of designing new monitoring efforts and other ecological studies. Other chapters in this volume address study design at length, but here I briefly mention a few tools useful for study design for the types of data considered in this chapter.

The design of such studies may be explored using study-specific simulations or using a number of specialized freeware packages.  For example, MLPowSim estimates statistical power using either R or the multilevel software package MLwiN, and using ML, REML or MCMC; MLPowSim may be used with continuous, binary and count data (Browne et al. 2009).  Another package, PINT, calculates approximate standard errors for estimates of fixed effects, as well as optimal sample sizes, for linear models with two levels (Snijders and Bosker 1993; PINT is available at http://stat.gamma.rug.nl/multilevel.htm). The program Optimal Design calculates sample size, statistical power, and optimal allocation of resources for multi-level studies with continuous and binary outcomes (Spybrook et al. 2009).  Note that Wang and Gelfand (2002) address sample size determination from a Bayesian perspective.

Methods for addressing study design for clustered count outcomes are underdeveloped. Models of trends among grouped counts with a design focus have presumed linearity (Gibbs et al. 1998, Urquhardt et al. 1998, Kincaid et al. 2004) or that counts were lognormally distributed (Gerrodette 1987, 1991).  Exceptions include Purcell et al. (2005) and Gray and Burlew (2007). Purcell et al. estimated statistical power to detect trends in a hierarchical count setting using Monte Carlo methods under an assumption that data were Poisson, conditional on random observer and/or route effects.  Gray and Burlew provided algorithms for estimating precision of and power to detect trends in a single population using GLMMs with conditional Poisson and NB distributional assumptions.

## FUTURE RESEARCH AND DEVELOPMENT

Sampling designs often incorporate variable selection probabilities.  For example, sampling units associated with heterogeneous habitats may be oversampled relative to those from homogeneous habitats. Variation in sampling probabilities may be addressed using survey statistical methods through the use of sampling weights. Unfortunately, the improper use of sampling weights using models (as distinct from design-based methods) may lead to biased estimates of among-group variation. Addressing this shortcoming through so-called design-adjusted models represents an area of ongoing research. Statistical modeling packages that accommodate sampling weights include Mplus, MLwiNand Stata (Muthén and Muthén 2010, Rasbash et al. 2009, StataCorp 2009). Further discussion of this topic is provided by Rabe-Hesketh and Skrondal (2006) and Carle (2009).

**SUMMARY**

Variance components may be estimated for scientific, management and study planning purposes. Scientific purposes include, for example, whether nitrate concentrations in lakes vary more among than within lakes, and whether either might be associated with putatively causal agents (e.g., agricultural runoff). Variance component estimates are used for study planning when, for example, an investigator wishes to select the number of groups (e.g., lakes) and the number of observations within each group for a future study. A major concern is that variance component estimators may be biased and yield imprecise estimates when the number of groups is small. This concern appears especially relevant for ecologists who, for logistic or cost reasons, may design studies with few sites and/or few years.

This chapter reviews the estimation of variance components and variance partition coefficients (VPCs) for continuous, categorical and count data that are clustered, and with emphases on studies with small sample sizes and crossed random effects. Variance components estimated from few groups using linear models of continuous outcomes may exhibit only modest bias when estimated using ANOVA or REML but may be substantially biased when estimated using FML. For all three estimation methods, however, precision is expected to be poor unless the number of groups is modest to large (e.g., more than 10, and possibly as many as 100).

Variance components estimated from GLMMs of categorical data may be expected to be both biased and imprecise when number of groups are few (e.g., <20 to as high as <100, depending on estimation method). The performance of Bayesian estimators of variance components from categorical data appears promising but has received relatively little attention in the literature.

GLMM estimators of variance components from count data have received less attention than have their categorical counterparts. Information supplied in this chapter suggests that, for cross-classified random effects models of count data, the Laplace estimator should be preferred over first-order quasilikelihood (QL) variance component estimators; the QL estimators suffered from poor convergence rates (PQL and RPQL) or substantial bias associated with $\sigma^2_{\text{site-year}}$ (MQL). Readers interested in VC estimation from count data should also consider Markov chain Monte Carlo and, for fully nested models, adaptive Gaussian quadrature.

The estimation of VPCs has received relatively little attention in the ecological literature. This is particularly the case for VPCs from categorical and count data (for which methods appear to have first been published in 2002 and 2008, respectively; Goldstein et al. 2002, Stryhn et al. 2008). For these discrete outcomes, VPCs may be estimated on both measurement and modeling or link scales. A method for estimating VPCs for binary outcomes on the measurement scale from two-way cross-classified random effects designs is proposed in Appendix 2.

**ACKNOWLEDGMENTS**

I thank Sherwin Toribio for providing the Bayesian estimates in Table 3, David Afshartous, Jialiang Li and Chuck Rose for helpful reviews, Bob Gitzen for helpful editorial suggestions, and Bethany Bell for reviewing the SAS code used for many of the simulations. This study was partially funded by the Upper Mississippi River's Long Term Resource Monitoring Program. References to proprietary software do not imply endorsement by the US government.

# APPENDIX 1. METHODS OF VARIANCE COMPONENTS ESTIMATION USING LINEAR MODELS

## Analysis of variance

Analysis of variance (ANOVA) may be used to estimate variance components by equating sums of squares to expected values. Chapter 8 in this volume provides an overview and example of ANOVA, while Chapter 7 illustrates another application of ANOVA estimation of variance components for a cross-classified model very similar to that associated with (1). Here, I briefly review general properties of ANOVA estimators of variance components.

The ANOVA estimators of variance components possess a number of advantages. One is that the ANOVA estimators make no distributional assumptions (other than that random effects have means zero and finite variances). Another is that they are unbiased. (However, note that Searle et al. 1992, section 2.3, question the importance of unbiasedness for variance component estimators.) A third is that, given balanced data sets and normality, ANOVA variance component estimators are "best unbiased" in the sense that, among unbiased estimators of variance components, they have uniformly smallest variance. A final and heuristic advantage is that variance components may often be defined in closed form—an advantage seen when variance component concepts are motivated using ANOVA arguments even when alternatives to ANOVA are recommended (Searle et al. 1992, Snijders and Bosker 1999).

However, ANOVA estimators of variance components also possess a number of negative attributes. First, their unbiasedness comes at the cost of allowing group-level variance estimates that are negative (variances, by definition, are nonnegative). Unfortunately, the usual solution of setting negative variance estimates to zero eliminates the unbiasedness property of these estimators. Second, the minimum variance properties mentioned above as a positive attribute for balanced data does not apply under unbalanced data assumptions. Third, unbalanced data—the common case for observational data—lead to estimators that are not only more complex but also which are not unique. Further details associated with ANOVA estimation of variance components for one-way and multi-way classifications are provided by Searle et al. (1992), Cox and Solomon (2003) and, with a less technical approach, Muller (2009).

**Maximum likelihood**

Maximum likelihood (ML) requires assuming an underlying probability distribution for a given set of data. ML may then be used to estimate values of the parameters associated with that distribution. For ML, the estimates are those that are considered most likely—given the data and the selected distribution.

ML estimators (MLEs) possess a number of favorable properties. These include properties those that are asymptotic—that are approached as sample size goes to infinity. These latter properties include: consistency (under fairly weak assumptions, MLEs converge to the value being estimated), normality, and minimum variance (among asymptotically unbiased estimators, and given commonly attained conditions). Another favorable property is that MLEs must be within the range of the given parameter (thereby eliminating the negative variance estimates that were possible under ANOVA). Likelihood-based methods also accommodate data that are missing at random. Last, the distributional assumption may be tailored to the process in question. For example, count data may be presumed to follow one or more of a number of potential count distributions.

Disadvantages associated with using ML include some of those associated with the favorable properties listed above. First, the favorable properties that are associated with large samples may not be reasonable for small samples. For a given study, investigators may infer that a sample is “large enough” based on experience, including that associated with Monte Carlo simulations (e.g., Table 1). Another concern relates to the possibility of unreasonable distributional assumptions. Given small sample sizes, for example, it may be difficult to determine on statistical grounds which distributional assumption is most reasonable. A rejoinder is that scientific theory and information from previous studies may (and arguably should) be used to select distributional assumptions.

An important concern related to the use of full maximum likelihood (FML—as distinct from REML, described below) to estimate variance components is that FML does not adjust for degrees of freedom associated with fixed effects. For data without clustering, for example, the FML variance estimator is $\sum (y_i - \mu)^2 / n$ rather than $\sum (y_i - \mu)^2 / (n-k)$, where $k$ denotes the number of linear independent predictors. Therefore, the bias of the FML estimator is approximately -$k$ / sample size (provided the proportion of zero estimates is small). See McCulloch and Searle (2001) for a fuller treatment of this topic. Further information about MLEs is provided by Casella and Berger (1990) and Searle et al. (1992).

### Restricted maximum likelihood

Restricted or residual maximum likelihood (REML) represents ML on a function of the data, specifically that function of the data from which fixed effects have been removed. REML confers the twin advantages of yielding estimators that are invariant to fixed effects, and of eliminating the variance component bias related to degrees of freedom described for FML estimators in the previous paragraph. For balanced data, REML estimators of variance components equal the expected value of the ANOVA estimates—provided that negative ANOVA estimates are set to zero. REML estimators of variance components are typically preferred over their ML counterparts (McCulloch and Searle 2001).

**Bayesian estimation**

In Bayesian statistics, each parameter is treated as a random variable, with variation described by a probability distribution. This distribution, which is assigned without reference to the data in question, is termed a prior distribution. The prior distribution is then updated using information from sample data, thereby yielding a posterior distribution for the parameter in question. This updating of the prior distribution occurs via Bayes theorem. The updated, posterior distribution is used for making inferences on the parameter in question. Introductions to Bayesian analysis with an ecological flavor are provided by Link et al. (2002) and Link and Barker (2009); more detailed treatments are provided by, for example, Gelman and Hill (2007) and Draper (2008).

Bayesian methods are valid with small samples, will not yield negative variance estimates and don’t require normality assumptions when variances are estimated. On the other hand, Bayesian estimators of group-level variances are not unbiased and have received relatively little attention in the literature; estimating variances using Bayesian methods may be more computationally intensive than is the case under classical methods. Fitting models of grouped data using Bayesian methods is addressed in Goldstein (2003, section 2.13), Browne and Draper 2006, Gelman and Hill (2007), Draper (2008) and in Hox (2010, section 11.4); the Goldstein and Hox references are less technical.

# APPENDIX 2. ESTIMATING VARIANCE COMPONENTS AND VARIANCE PARTITION COEFFICIENTS FOR TWO-WAY CROSS-CLASSIFIED RANDOM EFFECTS DESIGNS ON THE PROBABILITY SCALE USING SIMULATIONS

## Overview

Variance components on the inverse link scale may best be estimated using simulations (Li et al. 2008). In this case, estimation begins by treating parameter and variance estimates on the link scale as population values. The user then generates a large number of means (say, $m$ = 5000) under the population assumptions, and then transforms those means using an inverse link transformation. The variance of the means on the inverse link scale is then estimated by method of moments; variance at the measurement scale (given the model) is estimated under the sampling distribution assumption and across all simulated means.

For example, consider a random effects logistic regression model with a single group-level random effect, and with estimated grand intercept and variance on the logit scale of -1 and 0.7 units, respectively. Presume group effects on the logit scale are, as is typically assumed, normally distributed. Then generate a large number (e.g., 5000) of random normal variates and treat these (after adding a grand intercept) as the population of group means. Trivially, presume two random normal variates of 0.5 and -0.5 are generated. Given the grand intercept of -1, the corresponding means on the logit scale are then (-1 + 0.5) = -0.5 and (-1 – 0.5) = -1.5, respectively. These means may be transformed to probabilities using the inverse logit transformation

$$p_j = \frac{\exp(z_j)}{1+\exp(z_j)} = \frac{1}{1+\exp(-z_j)}$$

where $z_j$ and $p_j$ denote means on logit and probability scales, respectively. By this transform the two means on the measurement scale yield $p_j$ = 0.38 and 0.18, respectively. The variance of these means, by method of moments, is

$$\frac{1}{n-1}\sum_{j=1}^{2}(p_j - \hat{\bar{p}})^2 = 0.02$$

where $\hat{\bar{p}}$ denotes the sample mean. The variance at the measurement scale is estimated by the mean variance of the putative Bernoulli observations

$$\frac{1}{n}\sum_{j=1}^{2} p_j(1-p_j) = [0.38(1-0.38)+0.18(1-0.18)]/2 = 0.1916$$

The variance at the measurement scale for binomial models is typically presumed that of a Bernoulli outcome—because the number of trials per binomial count may vary and because covariates may vary across binary observations (Goldstein et al. 2002). Of course and as already emphasized, the choice of $m$ = 2 groups for this example represented a heuristic device; calculations should routinely be performed with much larger $m$. An elaboration of this method for two-way cross-classified random effects models is provided below.

Other methods for estimating variance components on the inverse link scale include Taylor series linearizations and, for categorical outcomes, a latent variable method (Goldstein et al. 2002). The latter method is commonly employed but appears appropriate only when the outcome of interest might reasonably derive from a continuous distribution (Snijders and Bosker 1999, ch. 14; Goldstein et al. 2002). For example, the probability that an organism may succumb to a toxicant may be postulated to derive from a standard logistic or standard normal cumulative distribution function. In these cases, the outcome may be treated as arising from a threshold model, and with variance that of a standard logistic or Gaussian outcome (i.e., $\pi^2/3$ or 1, respectively). For the example above, variance components at the group and measurement scales using the latent variable method would be the group-level variance (0.7 units) and $\pi^2/3$, respectively.

As with variance components, VPCs appear more informative from scientific and management perspectives when expressed on measurement scales. For the example above, VPCs calculated using the simulation and latent variable approaches are $0.02 / (0.02 + 0.1916) = 0.09$ and $0.07 / (0.07 + \pi^2/3) = 0.18$, respectively. Note that the latent variable approach yields the same VPC estimates for both link and measurement scales.

**Simulation Method for the Cross-classified Model**

As described above, the simulation method of estimating variance components and VPCs on the measurement or probability scale represents a reconstruction of the data generation process by computer simulations paired with a recording of the observed variation. Here, I propose a simulation method for use with the two-way cross-classified random effects model associated with (*3*). This method was adapted from the corresponding methodology for fully nested models (Goldstein et al. 2002, Browne et al. 2005, Li et al. 2008); I thank Bill Browne for reviewing an early draft of the proposed method. Note that the method simplifies when, as may often be the case, variation in random interaction effects, var($p_{jk}$), is treated as either inestimable or as identically zero. The method is as follows:

1. From the model [e.g., (*3*)], simulate a large number M1 (say 5,000) of main effects $u_{0j}$, $j$ = 1, …, M1, using the corresponding sample estimate of variance (e.g., $\widehat{\sigma^2_{\text{site}}}$). For each $j$, simulate M2 (say 5,000) of main effects $u_{0k}$, $k$ = 1, …, M2, using the corresponding sample estimate of variance (e.g., $\widehat{\sigma^2_{\text{year}}}$). For each unique combination, $jk$, simulate M3 (say, 30) interaction effects $u_{0jkl}$, $l$ = 1, …, M3, using the corresponding sample estimate of variance (e.g., $\widehat{\sigma^2_{\text{site-year}}}$). Calculate the M = M1×M2×M3 $p_{jkl}$'s as

$$p_{jkl} = \frac{1}{1+\exp(-(\widehat{\beta_{00}} + u_{0j} + u_{0k} + u_{0jkl}))}$$

   where $\widehat{\beta_{00}}$ denotes the grand intercept estimate.

2. Calculate the uncorrected (marginal) variation of the $p_j$'s within $k$ by method of moments. It may be convenient to use a single $l$ replicate per $jk$:

$$\text{var}(p_j)_{\text{m}} = \frac{1}{\text{M2}}\sum_{k=1}^{\text{M2}} \text{var}(p_{jkl} \mid k, l=1) \approx \text{var}(p_j) + \text{var}(p_{jk})$$

3. Similarly, calculate the uncorrected (marginal) variation of the $p_k$'s within $j$ by method of moments. It may be convenient to use a single $l$ replicate per $jk$:

$$\operatorname{var}(p_k)_{\mathrm{m}} = \frac{1}{\mathrm{M1}}\sum_{j=1}^{\mathrm{M1}} \operatorname{var}(p_{jkl} \mid j, l=1) \approx \operatorname{var}(p_k) + \operatorname{var}(p_{jk})$$

4. Calculate the variation of the $p_{jk}$'s by method of moments:

$$\operatorname{var}(p_{jk}) = \frac{1}{\mathrm{M1M2}}\sum_{j=1}^{\mathrm{M1}}\sum_{k=1}^{\mathrm{M2}} \operatorname{var}(p_{jkl})$$

5. Then,

$$\operatorname{var}(p_j) \approx \operatorname{var}(p_j)_{\mathrm{m}} - \operatorname{var}(p_{jk})$$

$$\operatorname{var}(p_k) \approx \operatorname{var}(p_k)_{\mathrm{m}} - \operatorname{var}(p_{jk})$$

$$\operatorname{var}(y_{ijk}) = \frac{1}{\mathrm{M1M2}}\sum_{j=1}^{\mathrm{M1}}\sum_{k=1}^{\mathrm{M2}} p_{jkl}(1 - p_{jkl}) \mid l=1 \tag{B1}$$

6. VPCs may be estimated from (*B1*) using (*2*), above.

The method ignores sampling variation in $\widehat{\beta_{00}}$ and in the variance estimates, and presumes linearity on the probability scale. The first concern may be addressed by nesting the method within a larger Monte Carlo simulation. The second concern remains unaddressed (Li et al. 2008). This method may be adapted to accommodate covariates after Goldstein et al. (2002), Browne et al. (2005) and Li et al. (2008).

The above method is somewhat demanding computationally. An approximation to this method which is less demanding but which elaborates the linearity assumption relies on differencing rather than on replicating on *jk* to estimate var($p_{jk}$). Under the model defined by eqn. (3) and the Table 3 legend, this alternative "differencing" method yields $\sigma^2_{\text{year}}$ estimates on the probability scale that are smaller than those estimated using the above-described method by 17%, 6%, 28% and 15% (for scenarios 1 to 5, 6, 7 and 8, respectively). Corresponding differences for $\text{VPC}_{\text{year}}$ estimates were similar: -16%, -5%, -28% and -14%, respectively. For this alternative method, steps 1 – 3 above are followed with the caveat that the interaction effects are not replicated (i.e., M3 = 1). Steps 4 and 5 become:

Alternate step 4. Calculate the uncorrected (marginal) variation of the $p_{jk}$'s by method of moments

$$\text{var}(p_{jk})_{\text{m}} \approx \text{var}(p_j) + \text{var}(p_k) + \text{var}(p_{jk})$$

Alternate step 5. Then,

$$\text{var}(p_j) \approx \text{var}(p_{jk})_{\text{m}} - \text{var}(p_k)_{\text{m}}$$

$$\mathrm{var}(p_k) \approx \mathrm{var}(p_{jk})_{\mathrm{m}} - \mathrm{var}(p_j)_{\mathrm{m}}$$

(*B1* (alt))

$$\mathrm{var}(p_{jk}) \approx \mathrm{var}(p_j)_{\mathrm{m}} + \mathrm{var}(p_k)_{\mathrm{m}} - \mathrm{var}(p_{jk})_{\mathrm{m}}$$

$$\mathrm{var}(y_{ijk}) = \frac{1}{\mathrm{M1M2}} \sum_{j=1}^{\mathrm{M1}} \sum_{k=1}^{\mathrm{M2}} p_{jk}(1 - p_{jk})$$

The properties of these two methods have not been rigorously investigated (cf., Li et al. 2008).

## APPENDIX 3. QUALITATIVE SUMMARY FOR PROGRAM MANAGERS AND ADMINISTRATORS

Managers often rely on estimates from grouped or clustered data. Examples include estimates of mean animal abundance from multiple observations from each of multiple lakes, streams or years. Such data should typically be presumed correlated within clusters, with the correlation arising because data from one cluster will typically be more like other data from that same cluster (and less like data from other clusters). Failure to address correlation in clustered data may yield invalid conclusions.

Correlation *within* clusters may be viewed as the flip side of variation *among* clusters. For example, the average or mean abundance of a species may vary from year to year, a finding that implies that data from a single year are relatively similar. As discussed in the previous paragraph, this similarity ensures abundance observations are correlated within years. Consequently, variation of means among years implies correlation within years.

This information—correlation and variation among year or cluster means—will often lead, from a planning perspective, to treating the number of clusters as more important than the number of data points within clusters. Since data from clusters are typically correlated, the information associated with those data is partially shared with other observations from the same cluster. This sharing of information ensures that data from the same cluster contain less information than is contained in an otherwise equivalent set of uncorrelated observations. Therefore, the contribution of observations within clusters to *effective* sample size is smaller than their contribution to the stated or nominal sample size. For example, 100 observations derived as 20 clusters of 5 observations might contain information equivalent to only 70 independent observations.

For designing future studies, investigators need estimates of variation among and within clusters. Given study goals, these estimates would then be used to select combinations of numbers of clusters and of observations per cluster. Determining the ratio between numbers of clusters and observations per cluster will typically also take into account the costs of sampling within clusters and of traveling between clusters. In the context of long-term monitoring, where years may often be viewed as clusters, there may be logistical/budgetary benefits associated with sampling less frequently but more intensively (e.g., every $k^{th}$ year).

Estimates of variation among clusters may also be used for scientific purposes. For example, variation in levels of invasive vegetation among lakes may suggest a causal agent that operates for entire lakes (e.g. lake substrate or lake management practice).

This chapter reviews variance estimation for types of data and situations that are commonly encountered in ecological studies. Major points include that among-cluster variances will often be both biased and estimated imprecisely unless the number of clusters is moderate to large (20 to possibly as many as 100). The accuracy and precision concerns described in this chapter should be considered when designing monitoring programs and when interpreting variance components for scientific and management purposes.

**APPENDIX 4. COMMON PROBLEMS AND DIFFICULT GRAY AREAS (with Robert A. Gitzen)**

As this chapter documents, variance components estimated from small sample sizes will often be characterized by low precision and – depending on the data type and estimator – substantial levels of bias. Investigators need to consider these concerns when using variance estimates from small samples for study design. Often, a simulation approach is useful for evaluating the potential magnitude of available estimates, following the same approach as demonstrated in this chapter but with sample sizes and other details tailored to the specific situation. The bias/precision observed in such simulations can be useful for determining a range of plausible values for each variance component; study design options may then be evaluated using multiple plausible values rather than that of a single point estimate (see also Chapter 8). In many cases, although absolute results of such design studies may change greatly across the range of plausible values, the comparative trade-offs among alternative design options may be relatively robust to uncertainties in the variance estimates. Absent such a finding, further data collection may be required.

The concern with poorly estimated variance components is especially problematic when monitoring program personnel wish to estimate the number of years until power to detect a given trend reaches a given level. The value of $\sigma^2_{\text{year}}$ will often have an overriding influence on power to detect temporal trends (e.g., Urquhart et al. 1998), and bias and imprecision associated with estimators of $\sigma^2_{\text{year}}$ may lead to biased and imprecise estimates of power to detect trends. In these cases, the simulation approach described in the previous paragraph will be useful. Also, in some cases, suitable long-term data sets conducive to meaningful estimation of $\sigma^2_{\text{year}}$ will be available, either from the system under investigation or from systems with patterns of variation hypothesized to be similar.

**REFERENCES**

Airy, G. B. 1861. On the algebraical and numerical theory of errors of observations and the combination of observations. Macmillan and Co., Cambridge and London.

Browne, W. J., and D. Draper. 2000. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. Computational Statistics 15:391-420.

Browne, W. J., and D. Draper. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models, Bayesian Analysis 1:673–514.

Browne, W. J., R. H. McCleery, B. C. Sheldon, and R.A. Pettifor. 2007. Using cross-classified multivariate mixed response models with application to life history traits in great tits (*Parus major*). Statistical Modelling 7:217-238.

Browne, W. J., M. Golalizadeh, and R. M. A. Parker. 2009. A guide to sample size calculations for random effect models via simulation and the MLPowSim software package. ISBN: 0-903024-96-9. University of Bristol, UK. <http://seis.bris.ac.uk/~frwjb/esrc/MLPOWSIMmanual.pdf>. Accessed 19 Aug 2010.

Browne, W. J., S. V. Subramanian, K. Jones, and H. Goldstein. 2005. Variance partitioning in multilevel models that exhibit overdispersion. Journal of the Royal Statistical Society, Series A 168:599-614.

Callens, M., and C. Croux. 2005. Performance of likelihood-based estimation methods for multilevel binary regression models. Journal of Statistical Computation and Simulation 75:1003-1017.

Cameron, A. C., and P. K. Trivedi. 1998. Regression analysis of count data. Cambridge University Press, Cambridge, UK.

Carle, A. C. 2009. Fitting multilevel models in complex survey data with design weights:

Recommendations. BMC Medical Research Methodology 9:49. doi:10.1186/1471-2288-9-49.

Casella, G., and R. L. Berger. 1990. Statistical inference. Duxbury Press, Belmont, California, USA.

Cools, W., W. Van Den Noortgate, and P. Onghena. 2009. Design efficiency for imbalanced multilevel data. Behavior Research Methods 41:192-203.

Cox, D. R., and P. J. Solomon. 2003. Components of variance. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Diaz, R. 2007. Comparison of PQL and Laplace 6 estimates of hierarchical linear models when comparing groups of small incident rates in cluster randomised trials. Computational Statistics & Data Analysis 51:2871-2888.

Draper, D. 2008. Bayesian multilevel analysis and MCMC. Pages 77-140 *in* J. de Leeuw and E. Meijer, editors. Handbook of multilevel analysis. Springer, New York, USA.

Fielding, A. 2003. Ordered category responses and random effects in multilevel and other complex structures. Pages 181-208 *in* S. P. Reise, and N. Duan, editors. Multilevel modeling—methodological advances, issues, and applications. Lawrence Erlbaum, Mahwah, New Jersey, USA.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge, New York, USA.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364–1372.

Gerrodette, T. 1991. Models for power of detecting trends—a reply to Link and Hatfield. Ecology 72:1889-1892.

Gibbs, J. P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals.

BioScience 48:935–940.

Goldstein, H. 2003. Multilevel statistical models. Third edition. Arnold, London, UK.

Goldstein, H. and J. Rasbash. 1996. Improved approximations for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 159:505-513.

Goldstein, H., W. J. Browne, and J. Rasbash. 2002. Partitioning variation in multilevel models. Understanding Statistics 1:223-232.

Gray, B. R. 2005. Selecting a distributional assumption for modelling relative abundances of benthic macroinvertebrates. Ecological Modeling 50:715–729.

Gray, B. R., and M. M. Burlew. 2007. Estimating trend precision and power to detect trends across grouped count data. Ecology 88:2364–2372.

Gray, B. R., R. J. Haro, and J. T. Rogala. 2010. Using random slopes models to estimate among-group variability in covariate effects: A case study using the association between mayfly counts and particle size. Environmental and Ecological Statistics 17:573-591.

Guo, G., and H. Zhao. 2000. Multilevel modeling for binary data. Annual Review of Sociology 26:441-462.

Hedeker, D. 2008. Multilevel models for ordinal and nominal variables. Pages 237-274 *in* J. de Leeuw, and E. Meijer, editors. Handbook of multilevel analysis. Springer, New York, USA.

Holland, M. D., and B. R. Gray. 2011. Multinomial mixture model with heterogeneous classification probabilities. Environmental and Ecological Statistics 18:257-270.

Holland, M. D., G. Meeden and B. R. Gray. 2010. A finite population Bayes procedure for censored categorical abundance data. Journal of the Indian Society of Agricultural Statistics 64:171-175.

Hox, J. J. 2010. Multilevel analysis—techniques and applications. Second edition. Routledge, New York, USA.

Kincaid, T. M., D. P. Larsen, and N. S. Urquhart. 2004. The structure of variation and its influence on the estimation of status: indicators of condition of lakes in the Northeast, U.S.A. Environmental Monitoring and Assessment 98:1–21.

Larsen, D. P., P. R. Kaufmann, T. M. Kincaid, and N. S. Urquhart. 2004. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. Canadian Journal of Fisheries and Aquatic Sciences 61:283–291.

Lee, A. H., K. Wang, J. A. Scott, K. K. W. Yau, and G. J. Mclachlan. 2006. Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. Statistical Methods in Medical Research 15:47-61.

Lee, Y., and J. A. Nelder. 2001. Hierarchical generalised linear models: a synthesis of generalised linear models, random effect models and structured dispersions. Biometrika 88:987–1006.

Li, J., B. R. Gray, and D. M. Bates. 2008. An empirical study of statistical properties of variance partition coefficients for multi-level logistic regression models. Communications in Statistics - Simulation and Computation 37:2010 – 2026.

Lindsey, J. K. 1997. Applying generalized linear models. Springer-Verlag, New York, USA.

Link, W. A., and R. J. Barker. 2009. Bayesian inference—with ecological applications. Academic Press, London, UK.

Link, W. A., E. Cam, J. D. Nichols, and E. G. Cooch. 2002. Of bugs and birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66:277-291.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. Second edition. SAS Institute, Cary, North Carolina, USA.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325-337.

Maas, C. J. M., and J. J. Hox. 2004. Robustness issues in multilevel regression analysis. Statistica Neerlandica 58:127-137.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, San Diego, USA.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, London, UK.

McCulloch, C. E., and S. R. Searle. 2001. Generalized, linear, and mixed models. John Wiley and Sons, New York, USA.

Meyers, J. L., and S. N. Beretvas. 2006. The impact of inappropriate modeling of cross-classified data structures. Multivariate Behavioral Research 41:473–497.

Min, Y., and A. Agresti. 2005. Random effect models for repeated measures of zero-inflated count data. Statistical Modelling 5:1-19.

Moghimbeigi, A., M. R. Eshraghian, K. Mohammad, and B. McArdle. 2008. Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. Journal of Applied Statistics 35:1193-1202.

Moineddin, R., F. I. Matheson, and R. H. Glazier. 2007. A simulation study of sample size for multilevel logistic regression models. BMC Medical Research Methodology 7:34.

Muller, K. E. 2009. Analysis of variance concepts and computations. Wiley Interdisciplinary

Reviews: Computational Statistics 1:279-282.

Muthén, L.K., and B. O. Muthén. 2010. Mplus user's guide. Sixth edition. Muthén & Muthén, Los Angeles, USA. <http://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v6.pdf>. Accessed 19 Aug 2010.

Pinheiro, J. C., and E. C. Chao. 2006. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. Journal of Computational and Graphical Statistics 15:58–81.

Purcell, K. L., S. R. Mori, and M. K. Chase. 2005. Design considerations for examining trends in avian abundance using point counts: examples from oak woodlands. Condor 107:305–320.

R Development Core Team. 2006. R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rabe-Hesketh, S., and A. Skrondral. 2006. Multilevel modeling of complex survey data. Journal of the Royal Statistical Society, Series A 169:805-827.

Rao, P. S. R. S. 1997. Variance components estimation—mixed models, methodologies and applications. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Rasbash, J., F. Steele, W. J. Browne, and H. Goldstein. 2009. A user's guide to MLwiN, v2.10. Centre for Multilevel Modelling, University of Bristol, UK. <http://www.cmm.bristol.ac.uk/MLwiN/download/manuals.shtml>. Accessed 19 Aug 2010.

Raudenbush, S. W. 2008. Many small groups. Pages 207-236 *in* J. de Leeuw, and E. Meijer, editors. Handbook of multilevel analysis. Springer, New York, USA.

Raudenbush, S. W., and A. S. Bryk. 2002. Hierarchical linear models: applications and data analysis methods. Second edition. Sage, Thousand Oaks, California, USA.

Rodriguez, G., and N. Goldman. 1995. An assessment of estimation procedures for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A 158:73-89.

Rodriguez, G., and N. Goldman. 2001. Improved estimation procedures for multilevel models with binary response: a case-study. Journal of the Royal Statistical Society, Series A 164:339-355.

Royle, J. A. 2006. Site occupancy models with heterogeneous detection probabilities. Biometrics 62:97-102.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, San Diego.

Royle, J. A., and W. A. Link. 2005. A general class of multinomial mixture models for anuran calling survey data. Ecology 86:2505-2512.

SAS Institute [SAS]. 2009. SAS OnlineDoc 9.2. SAS Institute Inc., Cary, North Carolina, USA.

Scheffé, H. 1956. Alternative models for the analysis of variance. Annals of Mathematical Statistics 27:251-271.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley, New York, USA.

Snijders, T. A. B., and J. Berkhof. 2008. Diagnostic checks for multilevel models. Pages 141-175 *in* J. de Leeuw, and E. Meijer, editors. Handbook of multilevel analysis. Springer, New York, USA.

Snijders, T. A. B., and R. J. Bosker. 1993. Standard errors and sample sizes for two-level research. Journal of Educational Statistics 18:237–259.

Snijders, T. A. B., and R. J. Bosker. 1999. Multilevel analysis. Sage Publications, London, UK.

Spybrook, J., S. W.Raudenbush, R. Congdon, and A. Martínez. 2009. Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software. <http://www.wtgrantfoundation.org/resources/overview/research_tools>. Accessed 19 Aug 2010.

StataCorp. 2009. Stata Statistical Software: Release 11. StataCorp LP, College Station, Texas, USA.

Stryhn, H., J. Sanchez, P. Morley, C. Booker, and I. R. Dohoo. 2008. Interpretation of variance parameters in multilevel Poisson regression models. Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006.

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8:246–257.

Wang, F., and Gelfand, A.E. 2002. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science 17:193-208.

Table 1**.** Estimates of among-year variances, $\sigma^2_{\text{year}}$, from random effects linear models by estimation method. Unless otherwise indicated, numbers of sites, years and observations per site-year ($n$) = 10, 5 and 5, respectively, $\sigma^2_{\varepsilon}$ = 1, $\sigma^2_{\text{site}} = \sigma^2_{\text{year}}$ = 0.3, and $\sigma^2_{\text{site-year}}$ = 0.15.[a]

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates by estimation method (Monte Carlo standard deviation), % zero estimates[b] | | |
|---|---|---|---|---|
| | | ANOVA | FML | REML |
| | | Sample size modifications | | |
| 1 | Years = 3 | 0.303 (0.345), 0.10 | 0.207 (0.235), 0.13 | 0.303 (0.345), 0.10 |
| 2 | Years = 5 (reference) | 0.285 (0.239), 0.02 | 0.236 (0.195), 0.02 | 0.285 (0.239), 0.02 |
| 3 | Years = 10 | 0.298 (0.159), 0.00 | 0.278 (0.146), 0.00 | 0.298 (0.159), 0.00 |
| 4 | Years = 20 | 0.297 (0.103), 0.00 | 0.290 (0.100), 0.00 | 0.297 (0.103), 0.00 |
| 5 | $n$ = 10 | 0.292 (0.236), 0.01 | 0.244 (0.193), 0.02 | 0.292 (0.236), 0.01 |
| 6 | sites = 20 | 0.284 (0.221), 0.01 | 0.233 (0.179), 0.01 | 0.284 (0.221), 0.01 |
| | | Unbalanced data sets | | |
| 7 | Unbalanced-1[c] | 0.284 (0.308), 0.06 | 0.228 (0.205), 0.06 | 0.279 (0.251), 0.05 |
| 8 | Unbalanced-2[d] | 0.300 (0.322), 0.06 | 0.231 (0.221), 0.11 | 0.293 (0.267), 0.08 |
| | | Variance component modifications | | |
| 9 | $\sigma^2_{\varepsilon}$ = 0.5 | 0.286 (0.233), 0.02 | 0.239 (0.191), 0.02 | 0.286 (0.233), 0.02 |

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates by estimation method (Monte Carlo standard deviation), % zero estimates[b] | | |
|---|---|---|---|---|
| | | ANOVA | FML | REML |
| 10 | $\sigma^2_{\text{site}} = 1$ | 0.285 (0.239), 0.02 | 0.251 (0.204), 0.02 | 0.285 (0.239), 0.02 |
| 11 | $\sigma^2_{\text{year}} = 1$ | 0.954 (0.740), 0.00 | 0.777 (0.595), 0.00 | 0.954 (0.740), 0.00 |
| | | Distributional violations | | |
| 12 | Uniform[e] | 0.302 (0.201), 0.02 | 0.251 (0.165), 0.02 | 0.302 (0.201), 0.02 |
| 13 | Uniform, years = 20 | 0.302 (0.079), 0.00 | 0.295 (0.077), 0.00 | 0.302 (0.079), 0.00 |

[a]Estimates represent means of simulation-specific estimates; simulations per scenario = 500; ‘ANOVA,’ ‘FML’ and ‘REML’ denote analysis of variance (Type III estimation), full maximum likelihood and residual maximum likelihood, respectively (see Littell et al. 2006 for computational details); negative variance estimates (ANOVA only) were set to zero; model convergence proportions for ML and REML = 1.00. Results were generated using SAS’ linear mixed modeling procedure (PROC MIXED; SAS 2009).

[b]≤1E-4 units; includes negative estimates (ANOVA only).

[c]All cells filled (in 3 of 5 years, 70% of sites contain only 1 observation).

[d]Some cells empty (in 3 of 5 years, 70% of sites contain no observations).

[e]Random effects generated as random uniform random variates with nominal variance.

Table 2. Among-year variance partition coefficient ($VPC_{year}$) estimates from random effects linear models by estimation method. Unless otherwise indicated, numbers of sites, years and observations per site-year (*n*) = 10, 5 and 5, respectively, $\sigma^2_{\varepsilon}$ = 1, $\sigma^2_{site} = \sigma^2_{year}$ = 0.3, and $\sigma^2_{site\text{-}year}$ = 0.15.[a]

| Scenario | Attribute Varied | $\widehat{VPC}_{year}$ by estimation method (VPC = 0.17) [Mean (Monte Carlo standard deviation), median] | | |
|---|---|---|---|---|
| | | ANOVA | FML | REML |
| | | Sample size modifications | | |
| 1 | Years = 3 | 0.15 (0.13), 0.12 | 0.11 (0.10), 0.09 | 0.15 (0.13), 0.12 |
| 2 | Years = 5 (reference) | 0.17 (0.09), 0.13 | 0.16 (0.08), 0.12 | 0.17 (0.09), 0.13 |
| 3 | Years = 10 | 0.17 (0.07), 0.16 | 0.16 (0.07), 0.15 | 0.17 (0.07), 0.16 |
| 4 | Years = 20 | 0.17 (0.05), 0.17 | 0.17 (0.05), 0.17 | 0.17 (0.05), 0.17 |
| 5 | *n* = 10 | 0.16 (0.10), 0.14 | 0.14 (0.09), 0.12 | 0.16 (0.10), 0.14 |
| 6 | sites = 20 | 0.15 (0.09), 0.14 | 0.13 (0.08), 0.12 | 0.15 (0.09), 0.14 |
| | | Unbalanced data sets | | |
| 7 | Unbalanced-1[b] | 0.15 (0.12), 0.12 | 0.13 (0.10), 0.11 | 0.15 (0.11), 0.13 |
| 8 | Unbalanced-2[c] | 0.15 (0.12), 0.13 | 0.13 (0.10), 0.11 | 0.15 (0.12), 0.13 |

[a]See footnote 'a,' Table 1, for analytical details.

[b]All cells filled (in 3 of 5 years, 70% of sites contain only 1 observation).

[c]Some cells empty (in 3 of 5 years, 70% of sites contain no observations).

Table 3. Among-year variance component, $\sigma^2_{\text{year}}$, estimates *on the logit scale* from random effects logistic models by estimation method. Unless otherwise indicated, number of sites = 20, number of years = 5, number of Bernoulli observations per site-year ($n$) = 5, mean probability on the logit scale ($\beta_{00}$) = -2 (i.e., median $p_{jk}$ = 0.12), $\sigma^2_{\text{site}} = \sigma^2_{\text{year}}$ = 0.3 and $\sigma^2_{\text{site-year}}$ = 0.15.[a]

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on logit scale by method (Monte Carlo standard deviation) | | | | |
|---|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace | MCMC |
| | | Sample size modifications | | | | |
| 1 | Years = 5 (reference) | 0.17 (0.17) | 0.19 (0.19) | 0.25 (0.24) | 0.21 (0.22) | 0.33 (0.41) |
| 2 | Years = 10 | 0.22 (0.16) | 0.24 (0.15) | 0.27 (0.17) | 0.26 (0.17) | 0.31 (0.24) |
| 3 | Years = 20 | 0.25 (0.12) | 0.26 (0.11) | 0.27 (0.12) | 0.28 (0.12) | 0.30 (0.14) |
| 4 | $n$ = 10 | 0.18 (0.16) | 0.21 (0.20) | 0.26 (0.25) | 0.22 (0.22) | 0.31 (0.30) |
| 5 | Sites = 40 | 0.19 (0.15) | 0.21 (0.17) | 0.27 (0.22) | 0.21 (0.19) | 0.31 (0.30) |
| | | Mean and variance component modifications | | | | |

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on logit scale by method (Monte Carlo standard deviation) | | | | |
|---|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace | MCMC |
| 6 | $\beta_{00} = 0$ | 0.17 (0.13) | 0.20 (0.17) | 0.26 (0.21) | 0.23 (0.19) | 0.36 (0.35) |
| 7 | $\sigma^2_{\text{site}} = 1$ | 0.16 (0.14) | 0.22 (0.21) | 0.26 (0.26) | 0.23 (0.24) | 0.31 (0.34) |
| 8 | $\sigma^2_{\text{year}} = 1$ | 0.53 (0.44) | 0.65 (0.55) | 0.84 (0.69) | 0.76 (0.68) | 1.19 (1.09) |

[a]Estimates represent means of simulation-specific estimates; methods include marginal quasi-likelihood (MQL1), penalized quasi-likelihood (PQL1), restricted PQL (RPQL1), Laplacian estimation (Laplace) and Bayesian analysis using Markov chain Monte Carlo (MCMC). MQL, PQL, RPQL and Laplacian estimation was performed using SAS' GLIMMIX procedure (SAS 2009) while MCMC estimation was implemented in WinBugs via the R package R2WinBugs (Lunn et al. 2000, R Development Core Team 2006). For the Bayesian analyses, priors for all variance components ($\sigma^2_{\text{site}}, \sigma^2_{\text{year}}, \sigma^2_{\text{site-year}}$) were gamma(0.001, 0.001), with means 1 and variances 1000. Each Bayesian estimate was obtained using 4000 posterior samples taken from a total of 22,000 iterations (the first 2000 iterations were used as burn in; samples were obtained every fifth iteration from the remaining 20,000 samples). Quasi-likelihood methods employ first-order Taylor series linearizations; percent model convergence per scenario ≥ 0.92 for QL methods, and 1.00 for Laplacian estimation; replicates per scenario = 200. R, SAS and WinBugs code are available at the volume's web site. All models presumed *p* varied as a logit-normal random variable.

Table 4. Among-year variance, $\sigma^2_{\text{year}}$, estimates *on the probability scale* from random effects logistic models by estimation method. Unless otherwise indicated, number of sites = 20, number of years = 5, number of Bernoulli observations per site-year ($n$) = 5, mean probability on the logit scale ($\beta_{00}$) = -2 (i.e., median $p_{jk}$ = 0.12), variances on the logit-scale of $\sigma^2_{\text{site}} = \sigma^2_{\text{year}}$ = 0.3 and $\sigma^2_{\text{site-year}}$ = 0.15.[a]

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on probability scale by method (Monte Carlo standard deviation) | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| | | Sample size modifications | | | |
| 1 | Years = 5 (reference) | 0.0029 (0.0033) | 0.0029 (0.0033) | 0.0036 (0.0036) | 0.0027 (0.0035) |
| 2 | Years = 10 | 0.0039 (0.0032) | 0.0036 (0.0028) | 0.0039 (0.0030) | 0.0036 (0.0029) |
| 3 | Years = 20 | 0.0042 (0.0023) | 0.0036 (0.0017) | 0.0037 (0.0020) | 0.0038 (0.0023) |
| 4 | $n$ = 10 | 0.0032 (0.0033) | 0.0031 (0.0031) | 0.0036 (0.0036) | 0.0030 (0.0034) |
| 5 | Sites = 40 | 0.0033 (0.0031) | 0.0031 (0.0029) | 0.0038 (0.0033) | 0.0029 (0.0029) |

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on probability scale by method (Monte Carlo standard deviation) | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| | | Mean and variance component modifications | | | |
| 6 | $\beta_{00} = 0$ | 0.0083 (0.0061) | 0.0089 (0.0067) | 0.0110 (0.0083) | 0.0099 (0.0077) |
| 7 | $\sigma^2_{\text{site}} = 1$ | 0.0030 (0.0032) | 0.0032 (0.0032) | 0.0039 (0.0042) | 0.0030 (0.0031) |
| 8 | $\sigma^2_{\text{year}} = 1$ | 0.0113 (0.0013) | 0.0099 (0.0087) | 0.0130 (0.0115) | 0.0110 (0.0112) |

[a]See footnote ‘a,’ Table 3, for estimation and design details. Estimates derive from the alternate or difference method described in Appendix 2.

Table 5. Among-year variance partition coefficient, $VPC_{year}$, estimates on the probability scale from random effects logistic models by estimation method and VPC estimation approach.[a]

| Scenario | Attribute varied | $VPC_{year}$ estimates on the probability scale by method (Monte Carlo standard deviation) | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| | | Simulation approach | | | |
| 1 | Years = 5 (reference) | 0.021 (0.021) | 0.021 (0.022) | 0.027 (0.025) | 0.021 (0.024) |
| 2 | Years = 10 | 0.027 (0.020) | 0.027 (0.019) | 0.029 (0.020) | 0.028 (0.020) |
| 3 | Years = 20 | 0.030 (0.016) | 0.028 (0.012) | 0.029 (0.014) | 0.030 (0.016) |
| 4 | $n$ = 10 | 0.022 (0.020) | 0.023 (0.021) | 0.027 (0.024) | 0.023 (0.023) |
| 5 | Sites = 40 | 0.023 (0.020) | 0.023 (0.020) | 0.029 (0.024) | 0.022 (0.021) |
| 6 | $\beta_{00} = 0$ | 0.034 (0.025) | 0.036 (0.027) | 0.045 (0.034) | 0.040 (0.031) |
| 7 | $\sigma^2_{site} = 1$ | 0.018 (0.018) | 0.022 (0.020) | 0.026 (0.025) | 0.021 (0.020) |

| Scenario | Attribute varied | $VPC_{year}$ estimates on the probability scale by method (Monte Carlo standard deviation) | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| 8 | $\sigma^2_{year}$ = 1 | 0.067 (0.058) | 0.068 (0.053) | 0.086 (0.065) | 0.076 (0.066) |
| | | Latent variable approach | | | |
| 1 | Years = 5 (reference) | 0.043 (0.040) | 0.047 (0.044) | 0.059 (0.050) | 0.048 (0.048) |
| 2 | Years = 10 | 0.056 (0.036) | 0.060 (0.036) | 0.065 (0.039) | 0.063 (0.039) |
| 3 | Years = 20 | 0.064 (0.030) | 0.064 (0.026) | 0.066 (0.028) | 0.069 (0.031) |
| 4 | $n$ = 10 | 0.046 (0.038) | 0.052 (0.042) | 0.061 (0.048) | 0.053 (0.046) |
| 5 | Sites = 40 | 0.048 (0.038) | 0.052 (0.041) | 0.064 (0.048) | 0.051 (0.044) |
| 6 | $\beta_{00}$ = 0 | 0.045 (0.033) | 0.050 (0.037) | 0.062 (0.047) | 0.056 (0.044) |
| 7 | $\sigma^2_{site}$ = 1 | 0.037 (0.032) | 0.047 (0.040) | 0.055 (0.047) | 0.046 (0.041) |
| 8 | $\sigma^2_{year}$ = 1 | 0.121 (0.087) | 0.136 (0.093) | 0.166 (0.105) | 0.151 (0.108) |

[a]See footnote "a," Table 3, for estimation and design details. Estimates under "Simulation approach" derive from the alternate or difference method described in Appendix 2.

Table 6. Among-year variance, $\sigma^2_{\text{year}}$, estimates on the log scale from random effects count models by estimation method. Unless otherwise indicated, number of sites = 20, number of years = 5, number of counts per site-year ($n$) = 5, median count mean ($\lambda_{jk}$) = 5, $\sigma^2_{\text{site}} = \sigma^2_{\text{year}}$ = 0.3 and $\sigma^2_{\text{site-year}}$ = 0.15.[a]

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on the log scale by method (Monte Carlo standard deviation), % convergence | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| | | Negative binomial outcomes (dispersion parameter = 0.5) | | | |
| NB1 | Years = 5 (reference) | 0.21 (0.16), 96 | 0.21 (0.13), 49 | 0.21 (0.16), 46 | 0.23 (0.16), 100 |
| NB2 | Years = 10 | 0.27 (0.16), 96 | 0.27 (0.10), 14 | 0.26 (0.11), 14 | 0.27 (0.12), 100 |
| NB3 | Years = 20 | 0.33 (0.20), 93 | 0.27 (0.09), 29 | 0.29 (0.09), 27 | 0.29 (0.09), 100 |
| NB4 | $n$ = 10 | 0.21 (0.16), 93 | 0.25 (0.18), 22 | 0.27 (0.22), 29 | 0.23 (0.16), 100 |
| NB5 | Sites = 40 | 0.23 (0.18), 92 | 0.17 (0.12), 14 | 0.23 (0.14), 15 | 0.23 (0.15), 100 |
| NB6 | $\lambda_{jk}$ = 2 | 0.20 (0.15), 90 | 0.19 (0.13), 46 | 0.24 (0.19), 42 | 0.23 (0.16), 100 |
| NB7 | $\lambda_{jk}$ = 20 | 0.21 (0.16), 98 | 0.23 (0.16), 47 | 0.25 (0.17), 49 | 0.23 (0.16), 100 |

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on the log scale by method (Monte Carlo standard deviation), % convergence | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| NB8 | $\sigma^2_{\text{site}} = 1$ | 0.21 (0.16), 88 | 0.25 (0.17), 37 | 0.27 (0.19), 33 | 0.24 (0.16), 92 |
| NB9 | $\sigma^2_{\text{year}} = 1$ | 0.60 (0.46), 87 | 0.66 (0.47), 34 | 0.77 (0.53), 31 | 0.62 (0.35), 90 |
| NB10 | Poisson assumption | 0.20 (0.15), 87 | 0.22 (0.14), 77 | 0.27 (0.20), 73 | 0.23 (0.16), 100 |
| | | Poisson outcomes | | | |
| P1 | Years = 5 (reference) | 0.21 (0.17), 99 | 0.22 (0.15), 71 | 0.24 (0.17), 67 | 0.22 (0.15), 100 |
| P2 | Years = 10 | 0.29 (0.24), 78 | 0.26 (0.12), 83 | 0.28 (0.13), 85 | 0.27 (0.12), 100 |
| P3 | Years = 20 | 0.32 (0.23), 72 | 0.28 (0.09), 48 | 0.29 (0.10), 46 | 0.29 (0.09), 100 |
| P4 | $n = 10$ | 0.22 (0.16), 90 | 0.23 (0.15), 72 | 0.27 (0.18), 72 | 0.23 (0.15), 100 |
| P5 | Sites = 40 | 0.21 (0.15), 84 | 0.23 (0.15), 78 | 0.27 (0.18), 82 | 0.23 (0.15), 100 |
| P6 | $\lambda_{jk} = 2$ | 0.22 (0.17), 94 | 0.22 (0.15), 66 | 0.27 (0.19), 64 | 0.23 (0.16), 100 |
| P7 | $\lambda_{jk} = 20$ | 0.22 (0.16), 98 | 0.23 (0.15), 92 | 0.27 (0.18), 94 | 0.23 (0.15), 100 |

| Scenario | Attribute varied | $\sigma^2_{\text{year}}$ estimates on the log scale by method (Monte Carlo standard deviation), % convergence | | | |
|---|---|---|---|---|---|
| | | MQL1 | PQL1 | RPQL1 | Laplace |
| P8 | $\sigma^2_{\text{site}} = 1$ | 0.22 (0.17), 94 | 0.25 (0.17), 60 | 0.28 (0.19), 61 | 0.24 (0.16), 100 |
| P9 | $\sigma^2_{\text{year}} = 1$ | 0.63 (0.48), 96 | 0.70 (0.51), 65 | 0.81 (0.56), 66 | 0.74 (0.50), 100 |

[a]Estimates represent means of simulation-specific estimates; replicates per scenario = 200. Methods include marginal quasi-likelihood (MQL1), penalized quasi-likelihood (PQL1), restricted PQL1 (RPQL1) and Laplacian estimation (Laplace). Quasi-likelihood methods employ first-order Taylor series linearizations.