BIOMETRICS 55, 137–148
March 1999

Estimating Intraclass Correlation for Binary Data

Martin S. Ridout,[1,*] Clarice G. B. Demétrio,[2] and David Firth[3]

[1]Horticulture Research International, East Malling, West Malling, Kent ME19 6BJ, U.K.
[2]Departamento de Matemática e Estatística, ESALQ/USP,
CP 13418-900, Piracicaba, São Paulo, Brazil
[3]Nuffield College, Oxford University, New Road, Oxford OX1 1NF, U.K.
*email: martin.ridout@hri.ac.uk

SUMMARY. This paper reviews many different estimators of intraclass correlation that have been proposed for binary data and compares them in an extensive simulation study. Some of the estimators are very specific, while others result from general methods such as pseudo-likelihood and extended quasi-likelihood estimation. The simulation study identifies several useful estimators, one of which does not seem to have been considered previously for binary data. Estimators based on extended quasi-likelihood are found to have a substantial bias in some circumstances.

KEY WORDS: Beta-binomial; Extended quasi-likelihood; Kappa coefficient; Pseudo-likelihood.

1. Introduction

The intraclass correlation coefficient ρ provides a quantitative measure of similarity between individuals within groups. Donner (1986) reviews the extensive literature on estimators of ρ for continuous variates. Often, however, the variable of interest is binary. For example, Donovan, Ridout, and James (1994) used intraclass correlation to quantify the extent of variation in rooting ability among somaclones of the apple cultivar Greensleeves. The binary response here measured whether or not rooting occurred, and the number of observations varied between somaclones. As another example, Gibson and Austin (1996) used an estimator of intraclass correlation to characterize the spatial pattern of disease incidence in an orchard and compare the observed disease pattern with patterns simulated from a model. Groups of trees were sampled in different parts of the orchard and classified as diseased or not diseased. The intraclass correlation provides an index of disease aggregation.

The purpose of this paper is to review numerous estimators of ρ that have been proposed for binary data and to report the results of an extensive simulation study of these estimators. Some of the estimators are those that have been proposed for continuous data. Others arise from the literature on observer agreement studies due to the link between intraclass correlation and chance-corrected measures of agreement (e.g., Fleiss and Cuzick, 1979). The standard model for intraclass correlation, which we discuss in Section 2, is widely used as a model for overdispersed binary data, and this provides a third source of estimators.

The remainder of the paper is organized as follows. Section 2 describes the assumed statistical model, the so-called common-correlation model. Section 3 defines various estimators of ρ, and Section 4 briefly explores some of their properties. Section 5 reviews previous comparative studies of the estimators. Section 6 describes an extensive simulation study, the results of which are reported in Section 7.

2. The Underlying Model

Suppose that there are k groups of individuals. The ith group contains n_i individuals with each having a binary response X_{ij} ($i = 1, \ldots, k$; $j = 1, \ldots, n_i$). We refer to the two possible values of X_{ij} as success and failure, coded as one and zero, respectively, and let $Y_i = \Sigma_j X_{ij}$ denote the total number of successes in the ith group.

The probability of success is assumed to be the same for all individuals, irrespective of the individual's group; specifically, $\Pr(X_{ij} = 1) = \pi$ for all i, j. Furthermore, the responses of individuals from different groups are assumed to be independent, while within each group, the correlation between any pair of responses (X_{ij}, X_{il}) $(j \neq l)$ is ρ. This model is therefore sometimes called the common-correlation model. In particular, the correlation is assumed not to vary with group size. It follows from these assumptions that

$$\mathrm{var}(Y_i) = n_i\pi(1-\pi)\{1 + (n_i - 1)\rho\}, \tag{1}$$

so that the Y_i are overdispersed relative to the binomial distribution if $\rho > 0$ and underdispersed if $\rho < 0$.

For continuous variables, the intraclass correlation ρ must satisfy the inequality $\rho \geq -1/(n_{\max} - 1)$, where $n_{\max}$ is the size of the largest group. For binary variables, there is generally a more stringent constraint (Prentice, 1986) that can be shown to be

$$\rho \geq \frac{-1}{(n_{\max} - 1)} + \frac{\omega(1-\omega)}{n_{\max}(n_{\max} - 1)\pi(1-\pi)}, \tag{2}$$

where $\omega = n_{\max}\pi - \text{int}(n_{\max}\pi)$ and $\text{int}(\cdot)$ denotes the integer part.

If ρ is positive, there is an alternative specification of the model in which the probability of success varies from group to group (but not between individuals in the same group) according to a distribution with mean π and variance $\rho\pi(1-\pi)$ and, conditional on this probability, individuals in the same group (as well as individuals from different groups) respond independently of one another (Kupper and Haseman, 1978).

One well-known example of a common-correlation model occurs when the group totals Y_i follow a beta-binomial distribution with

$$\Pr(Y_i = y_i) = \binom{n_i}{y_i} \frac{\prod_{j=0}^{y_i-1}(\pi + j\theta)\prod_{j=0}^{n_i-y_j-1}(1-\pi+j\theta)}{\prod_{j=0}^{n_i-1}(1+j\theta)}. \tag{3}$$

In this parametrization, π is the (marginal) probability of success for any individual and the intraclass correlation is $\rho = \theta/(1+\theta)$. The beta-binomial distribution is usually derived as a mixture distribution in which the probability of success varies from group to group according to a beta distribution with parameters α and β and, conditional on this probability, Y_i is binomially distributed. In terms of the parameters α and β, $\pi = \alpha/(\alpha+\beta)$ and $\theta = 1/(\alpha+\beta)$. In this formulation, ρ is necessarily nonnegative. However, Prentice (1986) showed that equation (3) defines a valid probability distribution for certain negative values of ρ, specifically

$$\rho > -\min\left(\frac{\pi}{n_{\max}-\pi-1}, \frac{1-\pi}{n_{\max}+\pi-2}\right). \tag{4}$$

Other specific common-correlation models that have appeared in the literature include the correlated binomial distribution of Kupper and Haseman (1978) and Altham (1978) and the correlated probit distribution of Ochi and Prentice (1984).

3. Estimators of Intraclass Correlation

3.1 *The Analysis of Variance Estimator*

This estimator is given by

$$\hat{\rho}_{\text{AOV}} = \frac{MS_b - MS_w}{MS_b + (n_0 - 1)MS_w},$$

where MS_b and MS_w are, respectively, the between-group and within-group mean squares from a one-way analysis of variance of the binary data X_{ij} and where

$$n_0 = \frac{1}{(k-1)}\left[N - \sum_{i=1}^{k}\frac{n_i^2}{N}\right] \quad \text{with } N = \sum_{i=1}^{k} n_i.$$

This estimator was proposed originally for continuous variables, but various authors (e.g., Elston, 1977) have subsequently suggested its use for binary variables. For binary data, explicit formulas for MS_b and MS_w are

$$MS_b = \frac{1}{k-1}\left[\sum_{i=1}^{k}\frac{Y_i^2}{n_i} - \frac{1}{N}\left(\sum_{i=1}^{k} Y_i\right)^2\right]$$

and

$$MS_w = \frac{1}{N-k}\left[\sum_{i=1}^{k} Y_i - \sum_{i=1}^{k}\frac{Y_i^2}{n_i}\right].$$

Fleiss (1981, pp. 226–228) used a modification of this estimator, here denoted by $\hat{\rho}^*_{\text{AOV}}$, in which the divisor of MS_b is k rather than $(k-1)$.

3.2 *Moment Estimators*

Let $\tilde{\pi}_i = Y_i/n_i$ denote the observed proportion of successes in the ith group and define

$$\tilde{\pi}_w = \sum_{i=1}^{k} w_i \tilde{\pi}_i \tag{5}$$

and

$$S_w = \sum_{i=1}^{k} w_i(\tilde{\pi}_i - \tilde{\pi}_w)^2, \tag{6}$$

where the w_i are weights summing to one. By equating $\tilde{\pi}_w$ and S_w to their expected values under the common-correlation model, Kleinman (1973) derived a class of estimators of the form

$$\hat{\rho} = \frac{S_w - \tilde{\pi}_w(1-\tilde{\pi}_w)\sum_{i=1}^{k}\frac{w_i(1-w_i)}{n_i}}{\tilde{\pi}_w(1-\tilde{\pi}_w)\left[\sum_{i=1}^{k} w_i(1-w_i) - \sum_{i=1}^{k}\frac{w_i(1-w_i)}{n_i}\right]}. \tag{7}$$

He considered two specific estimators, one with equal weights ($w_i = 1/k$) and the other with weights proportional to group size ($w_i = n_i/N$). Here we label these estimators $\hat{\rho}_{\text{KEQ}}$ and $\hat{\rho}_{\text{KPR}}$, respectively.

Kleinman (1973) also proposed slight variants of these estimators, here labelled $\hat{\rho}^*_{\text{KEQ}}$ and $\hat{\rho}^*_{\text{KPR}}$, that are obtained by replacing S_w in equation (7) by $S^*_w = (k-1)S_w/k$. The motivation for this is that, when the Y_i are from a beta-binomial distribution, and provided ρ is not too large (say $\rho < 0.8$), S^*_w has smaller mean square error as an estimator of $\text{E}(S_w)$ than does S_w itself.

In a more general context, Williams (1982) suggested an iterative scheme for estimating ρ, based on equation (7), in which the weights w_i are taken to be proportional to $n_i/[1+\hat{\rho}(n_i-1)]$, where $\hat{\rho}$ is the current estimate of ρ. This leads to an estimator that is here labelled $\hat{\rho}^*_{\text{W}}$. Again, this can be modified by replacing S_w by S^*_w in equation (7) to give a new estimator $\hat{\rho}_{\text{W}}$.

Tamura and Young (1987) were concerned with estimation of the parameter $\eta = \rho/(1-\rho)$ in connection with an empirical Bayes method of incorporating historical control data when testing for trend in binary dose–response data. The empirical Bayes procedure can be sensitive to small changes in the historical control data, and Tamura and Young (1987) proposed a stabilized estimator $\hat{\eta}_{\text{STAB}} = \hat{\rho}_{\text{STAB}}/(1-\hat{\rho}_{\text{STAB}})$ for η, where $\hat{\rho}_{\text{STAB}}$ is a modification of the estimator $\hat{\rho}_{\text{KPR}}$. Explicitly,

$$\hat{\rho}_{\text{STAB}} = \frac{1}{Nn_0 - 1}\left\{\frac{NS_w}{(k-1)\hat{\pi}(1-\hat{\pi})} + \kappa - 1\right\},$$

where S_w is given by equation (6) with weights $w_i = n_i/N$ and κ is the stabilization parameter. When $\kappa = 0$, $\hat{\rho}_{\text{STAB}} = \hat{\rho}_{\text{KPR}}$. For their application, Tamura and Young found values of κ in the range 0.3–0.6 to be effective, and in our simulation study, we follow Yamamoto and Yanagimoto (1992) in taking $\kappa = 0.45$.

Recently, Yamamoto and Yanagimoto (1992) proposed a new moment estimator that results from an unbiased estimating equation. After some rearrangement, this estimator can be written as

$$\hat{\rho}_{\mathrm{UB}} = 1 - \frac{N n_0 (m-1) MS_w}{\sum_{i=1}^k Y_i \left(n_0(m-1) - \sum_{i=1}^k Y_i \right) - \sum_{i=1}^k Y_i^2}.$$

(Note that the equation defining $\hat{\rho}_{\mathrm{UB}}$ in Yamamoto and Yanagimoto (1992, p. 280) contains two typographical errors.)

3.3 *Estimators with a Direct Probabilistic Interpretation*

Let the probability that two individuals have the same response (i.e., both zero or both one) be α when the two individuals are from the same group and β when they are from different groups. It is easily shown from the assumptions of Section 2 that $\alpha = 1 - 2\pi(1-\pi)(1-\rho)$, $\beta = 1 - 2\pi(1-\pi)$, and hence that

$$\rho = \frac{\alpha - \beta}{1 - \beta}. \tag{8}$$

Fleiss and Cuzick (1979) and Mak (1988) have obtained estimators of ρ by substituting estimators of α and β into this equation.

An unbiased estimator of α using data from the ith group is

$$1 - \frac{2Y_i(n_i - Y_i)}{n_i(n_i-1)}.$$

Fleiss and Cuzick estimated α as a weighted average of these within-group estimators, with weights proportional to $(n_i - 1)$, while Mak used the unweighted average.

Fleiss and Cuzick then estimated β by $1 - 2\hat{\pi}(1-\hat{\pi})$, where $\hat{\pi} = (\Sigma Y_i)/(\Sigma n_i)$ is the overall proportion of successes in the data, leading to the estimator

$$\hat{\rho}_{\mathrm{FC}} = 1 - \frac{1}{(N-k)\hat{\pi}(1-\hat{\pi})} \sum_{i=1}^k \frac{Y_i(n_i - Y_i)}{n_i}.$$

Alternatively, an unbiased estimator of β using data from the ith and jth groups is

$$\frac{Y_i Y_j + (n_i - Y_i)(n_j - Y_j)}{n_i n_j} = 1 - \frac{Y_i(n_j - Y_j) + (n_i - Y_i)Y_j}{n_i n_j},$$

and Mak estimated β as the (unweighted) average of this quantity taken over all $k(k-1)/2$ pairs of groups. This in turn leads, via equation (8), to the estimator

$$\hat{\rho}_{\mathrm{MAK}} = 1 - \frac{(k-1)\sum_{i=1}^k \frac{Y_i(n_i - Y_i)}{n_i(n_i-1)}}{\sum_{i=1}^k \frac{Y_i^2}{n_i^2} + \left(\sum_{i=1}^k \frac{Y_i}{n_i}\right)\left(k - 1 - \sum_{i=1}^k \frac{Y_i}{n_i}\right)}.$$

Clearly, two further estimators of ρ may be obtained by using the Fleiss–Cuzick estimator of α in conjunction with the Mak estimator of β and *vice versa*. However, calculation of the asymptotic variance of these estimators, similarly to Mak (1988), shows that, at least for beta-binomial data, these new estimators are inferior to $\hat{\rho}_{\mathrm{FC}}$ and $\hat{\rho}_{\mathrm{MAK}}$. Simulation studies endorse this conclusion for small samples, and we do not consider these estimators further.

3.4 *Estimators Based on Direct Calculation of Correlation Within Each Group*

Donner (1986) noted that the oldest method of estimating intraclass correlation is simply to calculate the Pearson correlation coefficient over all possible pairs of observations that can be constructed within groups. However, this tends to give too much weight to large groups, and Karlin, Cameron, and Williams (1981) have proposed the more general weighted estimator

$$\hat{\rho}_{\mathrm{PW}} = \left[\sum_{i=1}^k w_i \sum\sum_{j \neq l}^{n_i} (X_{ij} - \hat{\mu})(X_{il} - \hat{\mu})\right] \Bigg/ \left[\sum_{i=1}^k w_i(n_i - 1) \sum_{j=1}^{n_i} (X_{ij} - \hat{\mu})^2\right],$$

where

$$\hat{\mu} = \sum_{i=1}^k w_i(n_i - 1) \sum_{j=1}^{n_i} X_{ij}$$

and the weights w_i satisfy

$$\sum_{i=1}^k n_i(n_i - 1) w_i = 1.$$

For binary data,

$$\hat{\rho}_{\mathrm{PW}} = \frac{\sum_{i=1}^k w_i Y_i(Y_i - 1) - \hat{\mu}^2}{\hat{\mu}(1-\hat{\mu})},$$

and we consider here three choices of the weights w_i. First, taking w_i constant, i.e., the classical estimator giving equal weight to every pair of observations, gives

$$\hat{\mu}_{\mathrm{PEQ}} = \frac{\sum_{i=1}^k (n_i - 1) Y_i}{\sum_{i=1}^k (n_i - 1) n_i}$$

and

$$\hat{\rho}_{\mathrm{PEQ}} = \frac{1}{\hat{\mu}_{\mathrm{PEQ}}(1 - \hat{\mu}_{\mathrm{PEQ}})} \left[\frac{\sum_{i=1}^k Y_i(Y_i - 1)}{\sum_{i=1}^k n_i(n_i - 1)} - \hat{\mu}_{\mathrm{PEQ}}^2\right].$$

Second, taking $w_i = 1/[k n_i(n_i - 1)]$, i.e., giving equal weight to each group irrespective of its size, gives

$$\hat{\mu}_{\mathrm{PGP}} = \frac{1}{k} \sum_{i=1}^k \frac{Y_i}{n_i} = \tilde{\pi}$$

and

$$\hat{\rho}_{\mathrm{PGP}} = \frac{1}{\tilde{\pi}(1-\tilde{\pi})} \left[\frac{1}{k} \sum_{i=1}^k \frac{Y_i(Y_i - 1)}{n_i(n_i - 1)} - \tilde{\pi}^2\right].$$

Last, taking $w_i = 1/[N(n_i - 1)]$, i.e., weighting each pair according to the total number of pairs in which the individuals appear, gives

$$\hat{\mu}_{\mathrm{PPR}} = \frac{1}{N} \sum_{i=1}^k Y_i = \hat{\pi}$$

and

$$\hat{\rho}_{\mathrm{PPR}} = \frac{1}{\hat{\pi}(1-\hat{\pi})}\left[\frac{1}{N}\sum_{i=1}^{k}\frac{Y_i(Y_i-1)}{n_i-1} - \hat{\pi}^2\right].$$

The estimator $\hat{\rho}_{\mathrm{PGP}}$ can be shown to be identical to an estimator proposed by Lipsitz, Laird, and Brennan (1994, their equation (17)) and earlier by Schouten (1986).

3.5 *Extended Quasi-Likelihood and Pseudo-Likelihood Estimators*

For the common-correlation model, the extended quasi-(log)likelihood function of Nelder and Pregibon (1987) is, ignoring constant terms,

$$l_Q = -\frac{1}{2}\sum_{i=1}^{k}\left\{\log[1+(n_i-1)\rho] + \frac{D_i(Y_i,\pi)}{[1+(n_i-1)\rho]}\right\},$$

where D_i is the binomial deviance function

$$D_i(Y_i,\pi) = 2\left[Y_i\log\left(\frac{Y_i}{n_i\pi}\right) + (n_i-Y_i)\log\left(\frac{n_i-Y_i}{n_i-n_i\pi}\right)\right].$$

Differentiation of l_Q with respect to π and ρ leads to the pair of estimating equations

$$\sum_{i=1}^{k}\frac{Y_i-n_i\pi}{\pi(1-\pi)\phi_i} = 0 \tag{9}$$

and

$$\sum_{i=1}^{k}(n_i-1)\left(\frac{D_i-\phi_i}{\phi_i^2}\right) = 0, \tag{10}$$

where $\phi_i = 1+(n_i-1)\rho$. For a given value of ρ, the solution to equation (9) is simply $\tilde{\pi}_w$, given by equation (5), with weights $w_i = (1/\phi_i)/(\Sigma\,1/\phi_i)$. We write the estimator of ρ that results from simultaneous solution of the equations as $\hat{\rho}^*_{\mathrm{EQL}}$.

Equation (9) is an unbiased estimating equation, but equation (10) is not because, in general, $\mathrm{E}(D_i) \neq \phi_i$. The estimator $\hat{\rho}^*_{\mathrm{EQL}}$ is therefore inconsistent (Davidian and Carroll, 1988). Nelder and Lee (1992) have argued that this inconsistency, although technically undesirable, is often offset by high efficiency in finite samples.

The pseudo-likelihood method (cf., Carroll and Ruppert, 1988, Section 3.2) replaces D_i in equation (10) with

$$X_i^2 = \frac{(Y_i-n_i\pi)^2}{n_i\pi(1-\pi)}$$

and leads to the pseudo-likelihood estimator $\hat{\rho}^*_{\mathrm{PL}}$. Unlike $\hat{\rho}^*_{\mathrm{EQL}}$, $\hat{\rho}^*_{\mathrm{PL}}$ results from an unbiased estimating equation because $\mathrm{E}(X_i^2) = \phi_i$.

A variant of each of these methods that has been proposed to reduce bias in small samples is the degrees-of-freedom correction (see McCullagh and Nelder, 1989, Section 10.5.2); this multiplies D_i (or X_i^2) in equation (10) by $k/(k-1)$ to allow for estimation of π. We write the resulting estimators as $\hat{\rho}_{\mathrm{EQL}}$ and $\hat{\rho}_{\mathrm{PL}}$.

3.6 *Maximum Likelihood Estimator for Beta-Binomial Data*

Crowder (1979) discussed likelihood-based inference for ρ when the data are from a beta-binomial distribution. Smith (1983) gave an algorithm for calculating the maximum likelihood estimator $\hat{\rho}_{\mathrm{ML}}$ over the parameter space $0 \leq \rho \leq 1$. In the simulation study, we used a modification of this algorithm that maximizes the likelihood over the extended parameter space described in Section (2) (Smith and Ridout, 1995).

4. Some Properties of the Estimators

4.1 *Equivalence of Various Estimators When Group Sizes Are Equal*

When the group sizes n_i are equal, estimators within the following sets are equivalent: $\{\hat{\rho}_{\mathrm{AOV}}, \hat{\rho}_{\mathrm{MAK}}, \hat{\rho}_{\mathrm{UB}}\}$, $\{\hat{\rho}_{\mathrm{KEQ}}, \hat{\rho}_{\mathrm{KPR}}, \hat{\rho}_{\mathrm{W}}, \hat{\rho}_{\mathrm{PL}}\}$ and $\{\hat{\rho}^*_{\mathrm{AOV}}, \hat{\rho}_{\mathrm{FC}}, \hat{\rho}^*_{\mathrm{KEQ}}, \hat{\rho}^*_{\mathrm{KPR}}, \hat{\rho}^*_{\mathrm{W}}, \hat{\rho}^*_{\mathrm{PL}}, \hat{\rho}_{\mathrm{PEQ}}, \hat{\rho}_{\mathrm{PGP}}, \hat{\rho}_{\mathrm{PPR}}\}$. Yamamoto and Yanagimoto (1992) present some analytical comparisons between estimators for equal n_i; they show, e.g., that the value of the estimators in the first set cannot be less than the value of the estimators in the second set.

4.2 *Asymptotic Behaviour as $k \to \infty$*

Estimators marked with asterisks (e.g., $\hat{\rho}^*_{\mathrm{AOV}}$) differ from the corresponding unmarked estimators (*e.g.* $\hat{\rho}_{\mathrm{AOV}}$) only in the use of a divisor k in place of $(k-1)$ in their definition. Thus, all such pairs are asymptotically equivalent as $k \to \infty$.

The following estimators are also equivalent asymptotically as $k \to \infty$: $\{\hat{\rho}_{\mathrm{AOV}}, \hat{\rho}_{\mathrm{FC}}, \hat{\rho}_{\mathrm{UB}}, \hat{\rho}_{\mathrm{KPR}}\}$. Asymptotic equivalence of $\hat{\rho}_{\mathrm{AOV}}$ and $\hat{\rho}_{\mathrm{FC}}$ is demonstrated by Fleiss and Cuzick (1979), and it can be shown that $\hat{\rho}_{\mathrm{UB}}$ is also asymptotically equivalent to these two estimators. The asymptotic equivalence of $\hat{\rho}_{\mathrm{KPR}}$ and $\hat{\rho}_{\mathrm{FC}}$ follows from the easily established relationship $\hat{\rho}^*_{\mathrm{KPR}} = (\bar{n}-1)\hat{\rho}_{\mathrm{FC}}/(n_0-1)$, where $\bar{n} = N/k$.

4.3 *Parameter Estimates Lying Outside the Parameter Space*

Yamamoto and Yanagimoto (1992) noted that the estimators $\hat{\rho}_{\mathrm{KEQ}}$ and $\hat{\rho}_{\mathrm{STAB}}$ could exceed one for some data sets. This can also occur for several other estimators, including $\hat{\rho}_{\mathrm{KPR}}$, $\hat{\rho}^*_{\mathrm{KPR}}$, and $\hat{\rho}_{\mathrm{PGP}}$.

Table 1 gives a data set for which the estimators $\hat{\rho}_{\mathrm{AOV}}$, $\hat{\rho}^*_{\mathrm{AOV}}$, $\hat{\rho}_{\mathrm{FC}}$, $\hat{\rho}_{\mathrm{MAK}}$, $\hat{\rho}_{\mathrm{KEQ}}$, $\hat{\rho}^*_{\mathrm{KEQ}}$, $\hat{\rho}_{\mathrm{KPR}}$, $\hat{\rho}^*_{\mathrm{KPR}}$, $\hat{\rho}_{\mathrm{PEQ}}$, $\hat{\rho}_{\mathrm{PGP}}$, $\hat{\rho}_{\mathrm{PPR}}$, and $\hat{\rho}_{\mathrm{UB}}$ are all less than the lower bound for ρ given by equation (4). This set of estimators comprises all the estimators, except for $\hat{\rho}_{\mathrm{STAB}}$, that we have described that do not require iterative calculation. Our algorithms for obtaining the estimators that required iterative calculation were constrained to stay within the parameter space.

5. Previous Simulation Comparisons of Estimators

Yamamoto and Yanagimoto (1992) compared the estimators $\hat{\rho}_{\mathrm{ML}}$, $\hat{\rho}_{\mathrm{UB}}$, $\hat{\rho}_{\mathrm{W}}$, $\hat{\rho}^*_{\mathrm{W}}$, and $\hat{\rho}_{\mathrm{STAB}}$. They used $\pi = 0.05, 0.2, 0.4$ and, for each value of π, five values of ρ in the range [0.01, 0.4]. For each combination of π and ρ, they generated 10,000 samples from the beta-binomial distribution, each sample consisting of 20 groups of size 10. Samples that generated estimates of ρ greater than one or which led to undefined estimates were excluded, and negative estimates were replaced by zero.

Estimators were compared on the basis of mean and median bias and the ratio

$$E = (\text{mean square error})/(\text{sample mean})^2.$$

Table 1
Data set for which the estimators $\hat{\rho}_{\mathrm{AOV}}$, $\hat{\rho}^*_{\mathrm{AOV}}$, $\hat{\rho}_{\mathrm{FC}}$, $\hat{\rho}_{\mathrm{MAK}}$, $\hat{\rho}_{\mathrm{KEQ}}$, $\hat{\rho}^*_{\mathrm{KEQ}}$, $\hat{\rho}_{KPR}$, $\hat{\rho}^*_{\mathrm{KPR}}$, $\hat{\rho}_{\mathrm{PEQ}}$, $\hat{\rho}_{\mathrm{PGP}}$, $\hat{\rho}_{\mathrm{PPR}}$, and $\hat{\rho}_{\mathrm{UB}}$ are less than the lower bound for ρ given by equation (4)

i	y_i	n_i	i	y_i	n_i
1	1	10	6	2	13
2	1	9	7	1	17
3	1	15	8	2	14
4	1	13	9	2	13
5	1	10	10	2	13

For small values of ρ in conjunction with $\pi = 0.05$ or $\pi = 0.2$, many negative estimates of ρ were replaced by zero and consequently all estimators had positive bias. Otherwise, estimators generally had negative or negligible bias, except for $\hat{\rho}_{\mathrm{STAB}}$, which was always positively biased. The magnitude of the bias was generally less for $\hat{\rho}_{\mathrm{UB}}$ and $\hat{\rho}_{\mathrm{W}}$ than for $\hat{\rho}^*_{\mathrm{W}}$ and $\hat{\rho}_{\mathrm{ML}}$, and values of the efficiency measure E were correspondingly greater. Values of E for $\hat{\rho}_{\mathrm{STAB}}$ were similar to those for other estimators when $\pi = 0.4$, but for other values of π, they were often much less. On the basis of this study, Yamamoto and Yanagimoto (1992, p. 280) conclude that the estimator $\hat{\rho}_{\mathrm{UB}}$ is "best for practical use," though differences between this estimator and $\hat{\rho}_{\mathrm{W}}$ appear very slight.

Feng and Grizzle (1992), motivated by the problem of calculating sample sizes for group randomized studies, compared the estimators $\hat{\rho}_{\mathrm{AOV}}$, $\hat{\rho}^*_{\mathrm{AOV}}$, and $\hat{\rho}_{\mathrm{W}}$. They used $\pi = 0.05$, 0.2, and 0.35 and $\rho = 0$, 0.03, and 0.06, though for $\rho = 0$ and $\rho = 0.03$, results for different values of π were similar and were averaged. For each combination of π and ρ, they generated 500 samples from a beta-binomial distribution, each sample consisting of $k = 10$, 30, or 90 groups of equal size $n = 30$, 90, or 270. These parameter values were intended to cover the range commonly encountered in group randomized trials. Estimators were compared on the basis of bias and standard deviation. For small k, $\hat{\rho}_{\mathrm{AOV}}$ and $\hat{\rho}_{\mathrm{W}}$ had similar bias and were less biased than $\hat{\rho}^*_{\mathrm{AOV}}$. All three estimators had similar standard deviations.

Lipsitz et al. (1994) simulated 1825 sets of beta-binomial data with $\pi = 0.36$ and $\rho = 0.41$ for 26 groups of varying group size, ranging from 3 to 6, based on an example from Fleiss (1971). They compared the estimators $\hat{\rho}_{\mathrm{FC}}$, $\hat{\rho}_{\mathrm{PGP}}$ (the weighted estimator in their Table 4), and $\hat{\rho}_{\mathrm{ML}}$, together with another estimator (the unweighted estimator in their Table 4). This latter estimator performed poorly in the simulation study and also in the asymptotic relative efficiency calculations done by Lipsitz et al. (1994). The estimator $\hat{\rho}_{\mathrm{FC}}$ was slightly better than $\hat{\rho}_{\mathrm{PGP}}$. Although Lipsitz et al. (1994) used a Newton–Raphson algorithm to determine $\hat{\rho}_{\mathrm{ML}}$, the algorithm failed to converge in 120 simulation runs (6.6%). This led them to recommend that the $\hat{\rho}_{\mathrm{MLE}}$ should probably not be used in small samples. We repeated their simulations with our own computer program, obtaining similar results but with no convergence problems for $\hat{\rho}_{\mathrm{ML}}$. Indeed, although the algorithm we used (Smith, 1983; Smith and Ridout, 1995) detects various types of nonconvergence, none of these faults was reported in any of our simulation runs. In another application, we have encountered convergence problems (Xu and Ridout, 1996), but on further investigation, these were found to be related to difficulties in calculating asymptotic standard errors of estimates rather than in calculating the estimates themselves.

6. Simulation Study

Simulations were run for three values of π (0.05, 0.2, 0.5), five values of ρ (0, 0.05, 0.2, 0.5, 0.8), two values of k (10, 50), three different probability distributions (beta-binomial and two finite mixture distributions), and two distributions of group sizes. A fully factorial combination of these five factors was used, giving a total of 180 simulation runs.

In view of the widespread use of the common-correlation model in toxicology (e.g., Morgan, 1992, Chapter 6), the first distribution of group sizes was the empirical distribution of 523 litter sizes, quoted by Kupper et al. (1986) as being representative of that encountered in teratology screening studies. The litter sizes range from 1 to 19 with a mean of 12.0 and standard deviation of 2.98.

The second distribution of group sizes is a negative binomial distribution, truncated below 1 and above 15. This distribution, which is based on human sibship data for the U.S. (Brass, 1958), has been used by Donner and Koval (1980) in a study of estimators of intraclass correlation for continuous data. The mean and standard deviation of the distribution are, respectively, 3.1 and 2.11.

Data were simulated from three underlying distributions. The first was the beta-binomial distribution that was discussed in Section 2. The other two distributions were mixtures of two binomial distributions of the form

$$\gamma\,\mathrm{bin}(n, \pi_1) + (1 - \gamma)\mathrm{bin}(n, \pi_2).$$

The first distribution has $\pi_1 = 0$, $\pi_2 = \rho + \pi(1 - \rho)$, and $\gamma = 1 - \pi/\pi_2$ and the second has $\pi_1 = \pi(1 - \rho)$, $\pi_2 = 1$, and $\gamma = (1 - \pi)/(1 - \pi_1)$. These distributions are at opposite extremes of the class of mixtures of two binomial distributions that have the variance–mean relationship that results from the common-correlation model (equation (1)). The first of these distributions is equivalent to the Markov-like susceptibility model of Zucker and Wittes (1992).

For each simulation run, we used a Fortran program to generate data from the appropriate distribution. Uniform random numbers were generated by the NAG library routine G05CAF (Numerical Algorithms Group, 1989), which uses a multiplicative congruential method. To avoid problems with undefined estimators, the program rejected data sets that satisfied any of the following criteria:

(i) $y_i = 0$ for all $i = 1, \ldots, k$,
(ii) $y_i = n_i$ for all $i = 1, \ldots, k$,
(iii) $n_i = 1$ for all $i = 1, \ldots, k$.

Rejecting these simulated data sets is reasonable since one would not attempt to estimate intraclass correlation from such data sets in practice.

Simulations continued until 1000 acceptable data sets had been generated. Estimates of ρ that exceeded one or fell below the minimum possible value (equation (4)) were replaced by the appropriate extreme value. For each estimator, and for each combination of simulation factors, we calculated the mean bias, the standard deviation, and the mean square error (MSE) over the 1000 simulation runs. The MSE values were used to calculate the efficiency of each estimator relative to the estimator $\hat{\rho}_{\mathrm{ML}}$, which is asymptotically fully efficient for the simulation runs that generated data from the beta-binomial distribution, using the formula

$$\mathrm{RE(estimator)} = \frac{\mathrm{MSE}(\hat{\rho}_{\mathrm{ML}})}{\mathrm{MSE(estimator)}}.$$

Analysis of variance was used to identify the most important simulation factors affecting the bias, standard deviation, and MSE of the estimators.

7. Results

Figure 1 summarizes the bias, standard deviation, and mean square error of the 20 estimators across all 180 simulation runs as box-plots, modified along the lines suggested by Tufte (1983, p. 125). The plotted point shows the median of the distribution. The upper and lower end-points of the vertical line below the median indicate, respectively, the 25th and 5th percentile of the distribution, and the end-points of the vertical line above the median indicate similarly the 75th and 95th percentiles of the distribution.

The median bias over all simulation runs is close to zero for all estimators except for $\hat{\rho}_{\mathrm{STAB}}$, which is generally positively biased, and $\hat{\rho}_{\mathrm{PPR}}$, which is generally negatively biased. For most estimators, the distribution of bias is skewed to the left, with the 95th percentile often close to the median, indicating that the estimators were negatively biased under some simulation conditions but were seldom positively biased.

The distribution of bias for the extended quasi-likelihood estimators, $\hat{\rho}_{\mathrm{EQL}}$ and $\hat{\rho}^*_{\mathrm{EQL}}$, differs greatly from that of other estimators. Table 2 shows the effect of underlying probability (π), group size (k), and distribution of group sizes on the bias of $\hat{\rho}^*_{\mathrm{EQL}}$ for different true values of ρ. The results in Table 2 are for simulations from the beta-binomial distribution, but similar patterns were observed in the simulations from the two-component binomial mixture distributions. The bias is strongly dependent on π: when $\pi = 0.05$, the bias is negative; when $\pi = 0.2$, it is usually close to zero; and when $\pi = 0.5$, it is positive. The magnitude of the bias generally increases as the true value of ρ increases and, except for $\rho = 0.8$, is greater for the Brass distribution of group sizes than for the Kupper et al. distribution. We have already commented that the bias of $\hat{\rho}^*_{\mathrm{EQL}}$ does not disappear asymptotically as $k \to \infty$. Table 2 shows that, in most instances, the magnitude of the bias increases as the number of groups increases from $k = 10$ to $k = 50$.

The median standard deviation from the 180 simulation runs was similar for all estimators except for $\hat{\rho}_{\mathrm{PGP}}$, which had a slightly higher median than other estimators (Figure 1b). The distribution of standard deviation is skewed to the right for all estimators, but particularly for the pseudo-likelihood estimators $\hat{\rho}^*_{\mathrm{PL}}$ and $\hat{\rho}_{\mathrm{PL}}$, indicating that these estimators had large standard deviations under some simulation conditions.

Figure 1c shows the distribution of mean square error. The estimators with highest MSE were those that were most biased, i.e., $\hat{\rho}_{\mathrm{EQL}}$, $\hat{\rho}^*_{\mathrm{EQL}}$, and $\hat{\rho}_{\mathrm{STAB}}$, and also the correlation estimator that gives equal weight to each group, $\hat{\rho}_{\mathrm{PGP}}$. The estimators $\hat{\rho}_{\mathrm{AOV}}$, $\hat{\rho}^*_{\mathrm{AOV}}$, $\hat{\rho}_{\mathrm{FC}}$, $\hat{\rho}^*_{\mathrm{KPR}}$, $\hat{\rho}^*_{\mathrm{W}}$, and $\hat{\rho}_{\mathrm{UB}}$, which are asymptotically equivalent, behaved similarly to one another under all simulation conditions and had the lowest median MSE, together with the estimators $\hat{\rho}_{\mathrm{PEQ}}$, $\hat{\rho}^*_{\mathrm{KEQ}}$, and $\hat{\rho}_{\mathrm{ML}}$. The distributions of MSE of $\hat{\rho}_{\mathrm{PEQ}}$ and $\hat{\rho}^*_{\mathrm{KEQ}}$ were skewed more strongly to the right than those of the other estimators, indicating that they had somewhat higher MSE in some simulation runs. Figures 2, 3, and 4 show, for $\rho = 0.05$, 0.2, and 0.5, respectively, how the relative efficiency of four estimators is affected by the underlying distribution of group sizes, the distribution from which the data arise, and the true value of the intraclass correlation ρ. We chose $\hat{\rho}_{\mathrm{PEQ}}$ because it has good overall performance, based on Figure 1c, $\hat{\rho}_{\mathrm{FC}}$ as a representative of a group of similarly behaved estimators, $\hat{\rho}_{\mathrm{MAK}}$ because its behaviour can be very different than that of $\hat{\rho}_{\mathrm{FC}}$, at least asymptotically (Mak, 1988), and $\hat{\rho}^*_{\mathrm{EQL}}$, which appears to be a poor estimator under certain conditions. The relative efficiencies shown in Figures 2–4 are based on samples of $k = 10$ groups.

Overall, Figures 2–4 show that differences between estimators in their relative efficiency depend in quite complex ways on the underlying simulation factors, emphasizing the difficulty of drawing general conclusions. The asymptotic efficiency results of Mak (1988) suggest that $\hat{\rho}_{\mathrm{MAK}}$ may be more efficient than $\hat{\rho}_{\mathrm{FC}}$ for large ρ but less efficient for small ρ. However, it is only for $\pi = 0.5$ and $\rho = 0.8$ (Figure 4) that the efficiency of $\hat{\rho}_{\mathrm{MAK}}$ noticeably exceeds that of $\hat{\rho}_{\mathrm{FC}}$. Elsewhere, the estimators either have similar efficiency or $\hat{\rho}_{\mathrm{FC}}$ is more efficient, sometimes for all values of ρ (e.g., Figure 3d). The estimator $\hat{\rho}_{\mathrm{PEQ}}$ is particularly efficient when ρ is small, particularly when π is also small (Figure 2). Conversely, it performs poorly when ρ and π are large. In general, when other simulation factors are held fixed, the efficiency of $\hat{\rho}_{\mathrm{PEQ}}$ is more dependent on ρ than is the efficiency of $\hat{\rho}_{\mathrm{FC}}$, and this may be one reason for prefering $\hat{\rho}_{\mathrm{FC}}$.

The efficiency of $\hat{\rho}^*_{\mathrm{EQL}}$ is generally lower for the Kupper et al. distribution than for the Brass distribution. Although $\hat{\rho}^*_{\mathrm{EQL}}$ is a poor estimator in general, it is actually more efficient than $\hat{\rho}_{\mathrm{MAK}}$ and $\hat{\rho}_{\mathrm{FC}}$ when π and ρ are small (Figure 2).

Differences between the underlying distributions (beta-binomial, two-component mixtures) were more pronounced for the Kupper et al. distribution of group sizes than for the Brass distribution. This is to be expected because the former distribution gives larger group sizes with more opportunity for distributional differences. For groups of size 2, the distributions are identical because they are determined entirely by the mean and variance and, for the Brass distribution, the mean group size is only 3.1.

Because of limitations of space, we do not present the detailed results for $k = 50$. However, the efficiency of the estimators relative to $\hat{\rho}_{\mathrm{ML}}$ was generally higher than for $k = 10$, except for $\hat{\rho}^*_{\mathrm{EQL}}$. For this latter estimator, the efficiency decreased from $k = 10$ to $k = 50$ because of the increase in bias (Table 2).

8. Discussion

Where comparisons are possible, our results are in broad agreement with previous studies. The asymptotically equivalent estimators $\hat{\rho}_{\mathrm{AOV}}$, $\hat{\rho}^*_{\mathrm{AOV}}$, $\hat{\rho}_{\mathrm{FC}}$, $\hat{\rho}^*_{\mathrm{KPR}}$, $\hat{\rho}^*_{\mathrm{W}}$, and $\hat{\rho}_{\mathrm{UB}}$ all performed well in our simulations, and none of them appeared to have any consistent small-sample advantage. The estimator $\hat{\rho}_{\mathrm{PEQ}}$, which does not seem to have been used previously for binary data, also performed well, though its performance was more dependent on the true value of ρ. It appears to be particularly useful when

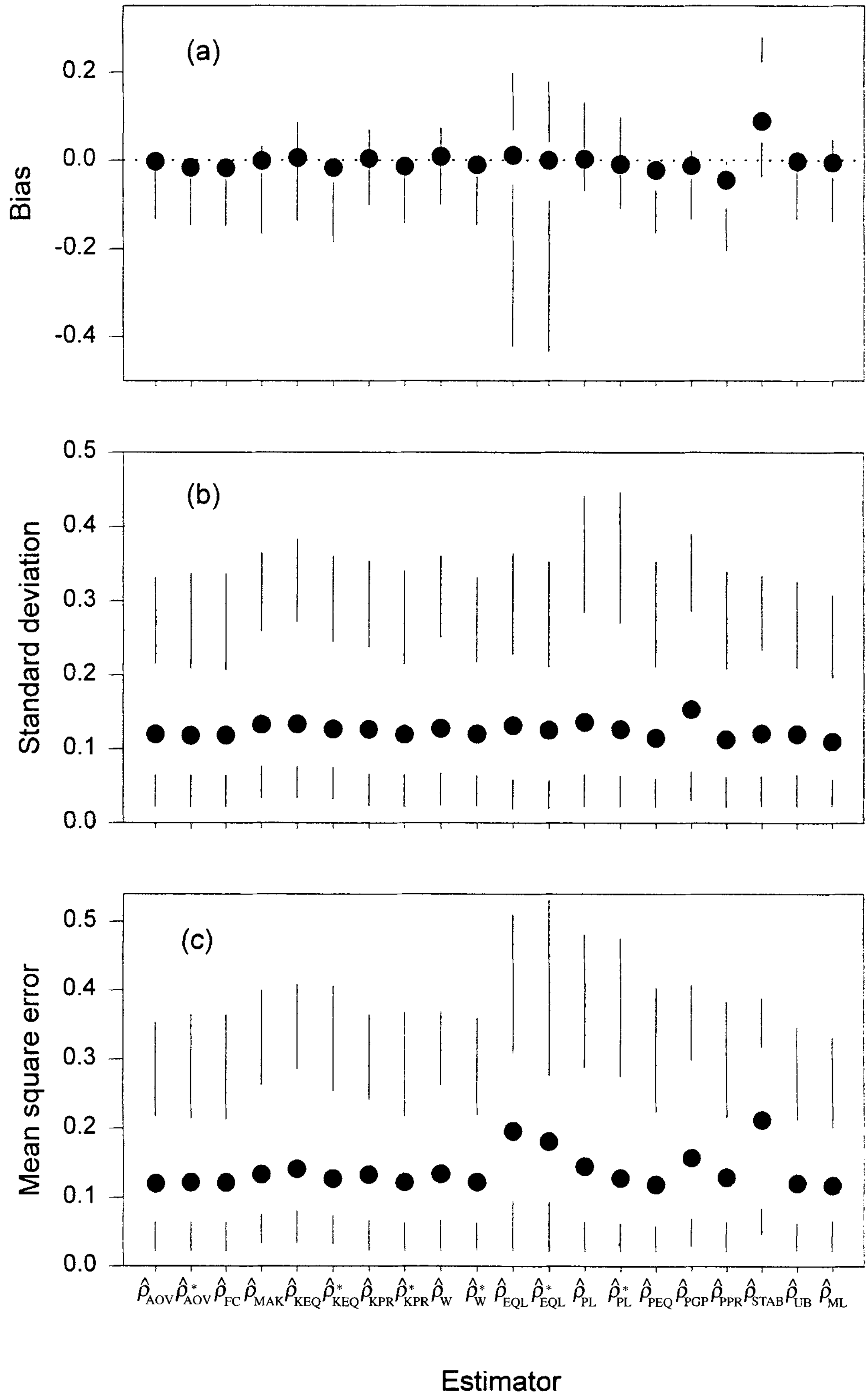

Figure 1. Bias, standard deviation, and mean square error of 20 estimators of ρ. The figure summarizes the distribution of these performance measures over 180 simulation runs. The plotted point is the median. Lower and upper end-points of the vertical lines indicate the 5th, 25th, 75th, and 95th percentiles of the distribution.

Table 2

The effect of probability (π), number of groups (k), distribution of group sizes, and true value of the intraclass correlation (ρ) on the bias of the estimator $\hat{\rho}^*_{\mathrm{EQL}}$. The table gives the mean of the difference $\hat{\rho}^*_{\mathrm{EQL}} - \rho$ in 1000 simulations from a beta-binomial distribution.

			ρ				
π	k	Groups	0	0.05	0.20	0.50	0.80
0.05	10	Kupper	−0.001	−0.028	−0.119	−0.303	−0.389
		Brass	−0.064	−0.100	−0.174	−0.315	−0.349
	50	Kupper	0.005	−0.017	−0.107	−0.323	−0.529
		Brass	−0.046	−0.082	−0.202	−0.429	−0.649
0.20	10	Kupper	0.001	−0.003	−0.018	−0.044	−0.105
		Brass	0.002	0.004	−0.024	−0.062	−0.106
	50	Kupper	0.008	0.012	0.014	0.000	−0.009
		Brass	0.036	0.034	0.040	0.011	−0.037
0.50	10	Kupper	−0.005	−0.004	0.023	0.104	0.133
		Brass	0.041	0.035	0.091	0.157	0.117
	50	Kupper	0.002	0.009	0.045	0.158	0.195
		Brass	0.050	0.083	0.158	0.295	0.199

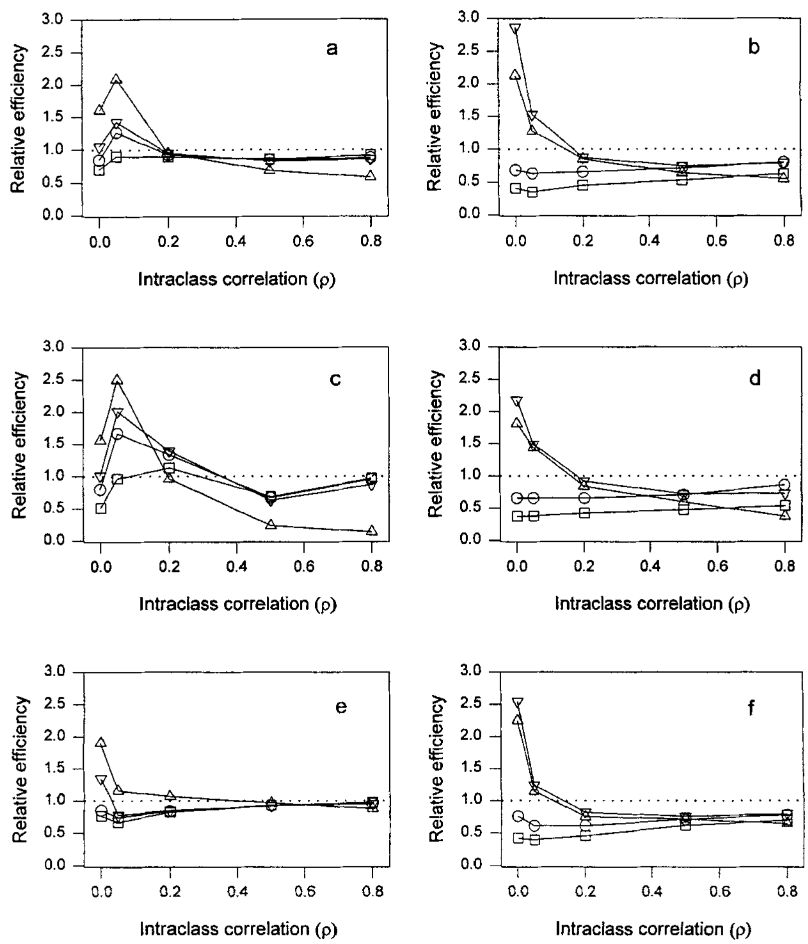

Figure 2. Relative efficiency of four estimators of ρ [$\hat{\rho}_{\mathrm{FC}}$ (O), $\hat{\rho}_{\mathrm{MAK}}$ (□), $\hat{\rho}_{\mathrm{PEQ}}$ (∇), and $\hat{\rho}^*_{\mathrm{EQL}}$ (Δ)] under varying simulation conditions. All simulations represented here had $k = 10$ and $\pi = 0.05$. The underlying distribution was beta-binomial (**a, b**), the first two-component binomial mixture (**c, d**), or the second two-component binomial mixture (**e, f**). The distribution of group sizes was from Kupper et al. (1986) (**a, c, e**) or Brass (1958) (**b, d, f**).

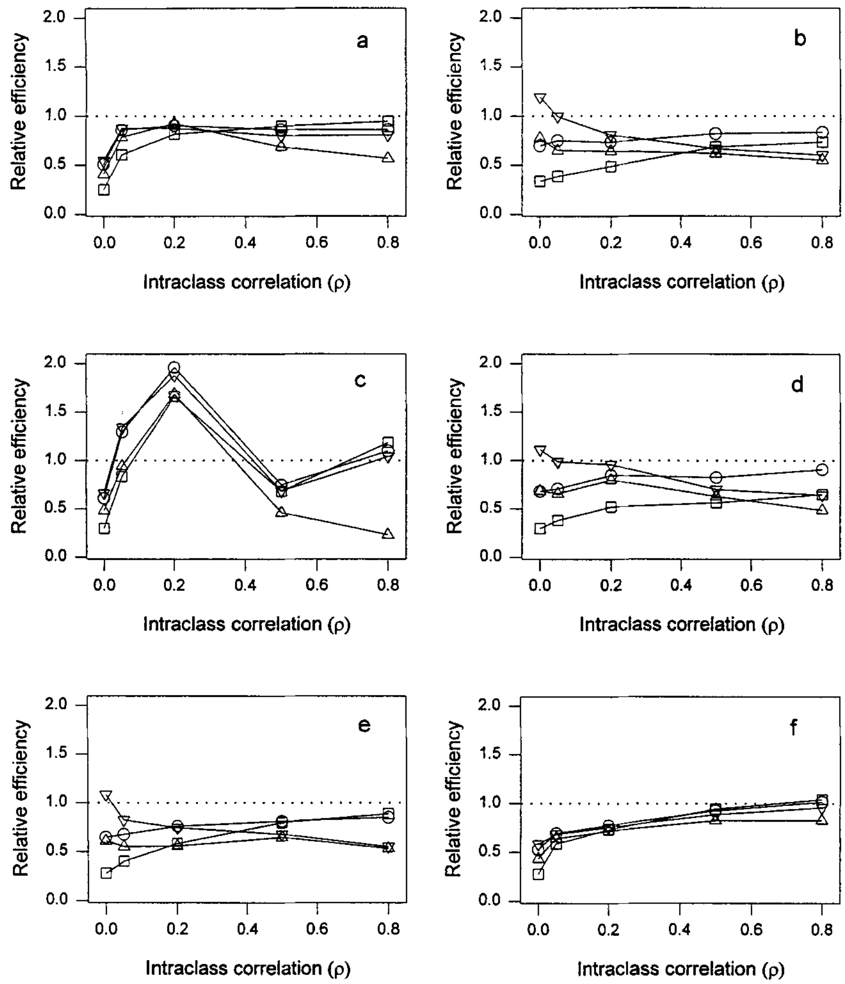

Figure 3. As for Figure 2 but with $\pi = 0.2$.

ρ is expected to be small and may therefore be useful as the basis of a test for $\rho = 0$. It is sometimes argued that this estimator gives too much weight to large groups (e.g., Donner, 1986), but in our simulations, it performed much better than the estimator $\hat{\rho}_{\mathrm{PGP}}$, which assigns equal weight to each group.

The performance of the extended quasi-likelihood estimators $\hat{\rho}_{\mathrm{EQL}}$ and $\hat{\rho}^*_{\mathrm{EQL}}$ indicates that the lack of consistency of these estimators can be a practical as well as a theoretical disadvantage. In their original paper on extended quasi-likelihood, Nelder and Pregibon (1987) used the method to reanalyze some seed germination data of Crowder (1978) using a model with constant ρ (θ in their notation) but with π dependent on treatment factors. The extended quasi-likelihood method gave results similar to maximum likelihood estimation assuming the data to be beta-binomially distributed. However, the estimate of ρ was small (0.012), and our simulations showed $\hat{\rho}_{\mathrm{EQL}}$ and $\hat{\rho}^*_{\mathrm{EQL}}$ to have reasonable, and sometimes high, efficiency for small ρ. In contrast to our results, Clark and Perry (1989) and Piegorsch (1990) recommended the use of extended quasi-likelihood estimation for the somewhat analogous problem of estimating the dispersion parameter of a negative binomial distribution.

We have assumed a model in which the underlying probability π is the same in all groups and many of the estimators that we have discussed were developed for this specific problem. However, the estimators ρ_{W}, ρ^*_{W}, ρ_{EQL}, ρ^*_{EQL}, ρ_{PL}, and ρ^*_{PL} are all directly applicable in the more general situation in which π depends on covariates. Among these estimators, ρ^*_{W} performed best in our simulations. Lipsitz et al. (1994) and Molenberghs, Fitzmaurice, and Lipsitz (1996) discuss alternative estimators derived from generalized estimating equations. In most instances, ρ will be positive, implying that the data are overdispersed relative to the binomial distribution. The common-correlation model is only one of several ways in which such data might be modelled, and we mention briefly two alternative approaches.

If all pairs of responses Y_{ij} and Y_{ik} in the ith group have the same correlation ρ, then the odds ratio

$$\psi = \frac{\Pr(Y_{ij} = 1, Y_{ik} = 1)\Pr(Y_{ij} = 0, Y_{ik} = 0)}{\Pr(Y_{ij} = 1, Y_{ik} = 0)\Pr(Y_{ij} = 0, Y_{ik} = 1)}$$

is also constant and is related to ρ by the equation

$$\psi = 1 + \frac{\rho}{(1-\rho)^2}\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right).$$

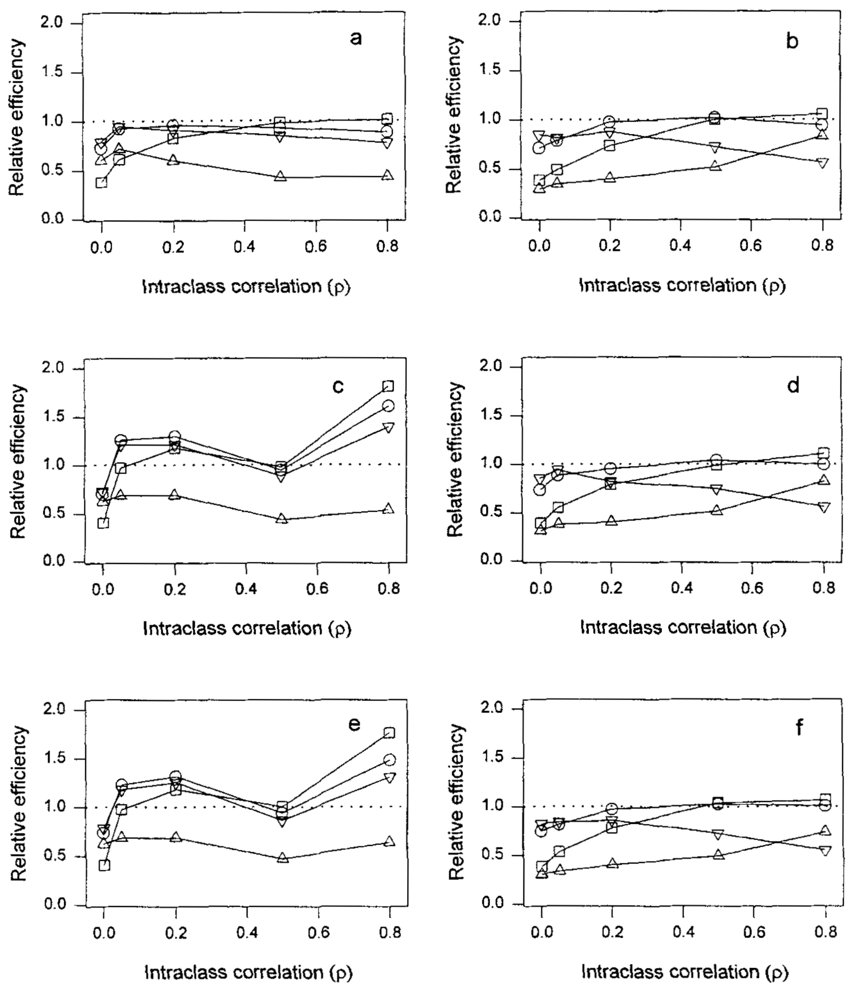

Figure 4. As for Figure 2 but with $\pi = 0.5$.

Thus, if π varies, models with constant ρ are distinct from models with constant ψ. Models based on odds ratios have been considered by, e.g., Dale (1986), Lang and Agresti (1994), Glonek and McCullagh (1995), and Lipsitz, Laird, and Harrington (1991).

An alternative approach is to use a generalized linear mixed model in which the effects of the covariates are assumed to be linear on a transformed (e.g., logit) scale, and a random group effect is added on the transformed scale. Williams (1982) contrasts one such model (his model III) with the common-correlation model. Estimation of parameters in generalized linear mixed models is a topic of great current interest (cf., Schall, 1991; Breslow and Clayton, 1993; McGilchrist, 1994; Clayton, 1996).

Acknowledgements

Most of this work was done while the second author was visiting HRI, East Malling. We are grateful to the Royal Society and to FAPESP for funding. We thank J. S. Fenlon, Professor J. A. Nelder, Dr A. R. Thompson, the associate editor, and two referees for their comments.

Résumé

Cet article fait la revue de différents estimateurs de corrélation intraclasse qui ont été proposés pour des données binaires, et les compare par une étude extensive de simulation. Certains estimateurs sont vraiment particuliers, tandis que d'autres résultent de l'utilisation de méthodes générales comme les estimations par pseudo-vraisemblance et quasi-vraisemblance étendue. L'étude de simulation identifie plusieurs estimateurs utiles, l'un d'entre eux ne semble pas avoir été considéré auparavant pour des données binaires. Elle montre aussi que des estimateurs fondés sur une quasi-vraisemblance étendue peuvent avoir un biais substantiel dans certaines circonstances.

References

Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Applied Statistics* **27,** 162–167.

Brass, W. (1958). Models of birth distribution in human populations. *Bulletin of the International Statistical Institute* **36**(2), 165–179.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* London: Chapman and Hall.

Clark, S. J. and Perry, J. N. (1989). Estimation of the negative binomial parameter κ by maximum quasi-likelihood. *Biometrics* **45,** 309–316.

Clayton, D. G. (1996). Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice,* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), 275–301. London: Chapman and Hall.

Crowder, M. J. (1978). Beta-binomial ANOVA for proportions. *Applied Statistics* **27,** 34–37.

Crowder, M. J. (1979). Inference about the intraclass correlation coefficient in the beta-binomial ANOVA for proportions. *Journal of the Royal Statistical Society, Series B* **41,** 230–234.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42,** 909–917.

Davidian, M. and Carroll, R. J. (1988). A note on extended quasi-likelihood. *Journal of the Royal Statistical Society, Series B* **50,** 74–82.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review* **54,** 67–82.

Donner, A. and Koval, J. J. (1980). The estimation of intraclass correlation in the analysis of family data. *Biometrics* **36,** 19–25.

Donovan, A. M., Ridout, M. S., and James, D. J. (1994). Assessment of somaclonal variation in apple. II. Rooting ability and shoot proliferation. *Journal of Horticultural Science* **69,** 115–122.

Elston, R. C. (1977). Response to query: Estimating "heritability" of a dichotomous trait. *Biometrics* **33,** 232–233.

Feng, Z. and Grizzle, J. E. (1992). Correlated binomial variates: Properties of estimator of intraclass correlation and its effect on sample size calculation. *Statistics in Medicine* **11,** 1607–1614.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* **76,** 378–382.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions,* 2nd edition. New York: Wiley.

Fleiss, J. L. and Cuzick, J. (1979). The reliability of dichotomous judgements: Unequal numbers of judges per subject. *Applied Psychological Measurement* **3,** 537–542.

Gibson, G. J. and Austin, E. J. (1996). Fitting and testing spatio-temporal stochastic models with applications in plant pathology. *Plant Pathology* **45,** 172–184.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* **57,** 533–546.

Karlin, S., Cameron, P. E., and Williams, P. (1981). Sibling and parent–offspring correlation with variable family age. *Proceedings of the National Academy of Science, U.S.A.* **78,** 2664–2668.

Kleinman, J. C. (1973). Proportions with extraneous variance: Single and independent samples. *Journal of the American Statistical Association* **68,** 46–54.

Kupper, L. L. and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* **34,** 69–76.

Kupper, L. L., Portier, C., Hogan, M. D., and Yamamoto, E. (1986). The impact of litter effects on dose–response modelling in teratology. *Biometrics* **42,** 85–98.

Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* **89,** 625–632.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* **78,** 153–160.

Lipsitz, S. R., Laird, N. M., and Brennan, T. A. (1994). Simple moment estimates of the κ-coefficient and its variance. *Applied Statistics* **43,** 309–323.

Mak, T. K. (1988). Analysing intraclass correlation for dichotomous variables. *Applied Statistics* **37,** 344–352.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

McGilchrist, C. A. (1994). Estimation in generalized mixed models. *Journal of the Royal Statistical Society, Series B* **56,** 61–69.

Molenberghs, G., Fitzmaurice, G. M., and Lipsitz, S. R. (1996). Efficient estimation of the intraclass correlation for a binary trait. *Journal of Agricultural, Biological, and Environmental Statistics* **1,** 78–96.

Morgan, B. J. T. (1992). *Analysis of Quantal Response Data.* London: Chapman and Hall.

Nelder, J. A. and Lee, Y. (1992). Likelihood, quasi-likelihood and pseudo-likelihood: Some comparisons. *Journal of the Royal Statistical Society, Series B* **54,** 273–284.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74,** 221–232.

Numerical Algorithms Group. (1989). *NAG FORTRAN Library Manual, Mark 13.* Oxford: Numerical Algorithms Group.

Ochi, Y. and Prentice, R. L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika* **71,** 531–543.

Piegorsch, W. G. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics* **46,** 863–867.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81,** 321–327.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78,** 719–727.

Schouten, H. J. A. (1986). Nominal scale agreement among observers. *Psychometrika* **51,** 453–466.

Smith, D. M. (1983). Algorithm AS189: Maximum likelihood estimation of the parameters of the beta binomial distribution. *Applied Statistics* **32,** 196–204.

Smith, D. M. and Ridout, M. S. (1995). Algorithm AS R93: A remark on algorithm AS189: Maximum likelihood estimation of the parameters of the beta binomial distribution. *Applied Statistics* **44,** 545–547.

Tamura, R. N. and Young, S. S. (1987). A stabilized moment estimator for the beta-binomial distribution. *Biometrics* **43,** 813–824. Correction: (1994), **50,** 321.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information.* Cheshire, U.K.: Graphics Press.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* **31,** 144–148.

Xu, X.-M. and Ridout, M. S. (1996). Analysis of disease incidence data using a stochastic simulation model. *Aspects of Applied Biology* **46,** 155–158.

Yamamoto, E. and Yanagimoto, T. (1992). Moment estimators for the binomial distribution. *Journal of Applied Statistics* **19,** 273–283.

Zucker, D. and Wittes, J. (1992). Testing the effect of treatment in experiments with correlated binary outcomes. *Biometrics* **48,** 695–710.

Received November 1996. *Revised September* 1997.
Accepted January 1998.