Biometrika (2009), **96**, 2, *pp.* 457–468
doi: 10.1093/biomet/asp003

Advance Access publication 7 April 2009
Printed in Great Britain

Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models

BY SHARON L. LOHR

Department of Mathematics and Statistics, Arizona State University, Tempe, Arizona 85287-1804, U.S.A.
sharon.lohr@asu.edu

AND J. N. K. RAO

School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6, Canada
jrao@math.carleton.ca

SUMMARY

Empirical Bayes predictors of small area parameters of interest are often obtained under a linear mixed model for continuous response data or a generalized linear mixed model for binary responses or count data. However, estimation of the unconditional mean squared error of prediction is complicated, particularly for a nonlinear mixed model. Jiang et al. (2002) proposed a jackknife method for estimating the unconditional mean squared error and showed that the resulting estimator is nearly unbiased. The leading term of this estimator does not depend on the area-specific responses in the nonlinear case, whereas the posterior variance of the small area parameter given the model parameters is area-specific. Rao (2003) proposed an alternative method that leads to a computationally simpler jackknife estimator with an area-specific leading term. We show that a modification of Rao's method leads to a nearly unbiased area-specific jackknife estimator, which is also nearly unbiased for the conditional mean squared error given the area-specific responses. We examine the relative performances of the jackknife estimators, conditionally as well as unconditionally, in a simulation study, and apply the proposed method to estimate small area mean squared errors in disease mapping problems.

Some key words: Area-specific; Beta-binomial model; Binary response; Disease mapping; Empirical Bayes; Generalized linear mixed model.

1. INTRODUCTION

Small area estimation methods are commonly used to make inferences for subpopulations, or small areas, in which the sample sizes are too small for direct estimators to be sufficiently reliable. Typically, models are employed to borrow strength from other small areas, using covariates such as census and administrative information; see Rao (2003) for an overview.

Suppose there are m small areas. For area i, the parameter of interest is denoted by θ_i. For example, θ_i might be the poverty rate for county i, or the proportion of persons with diabetes in demographic subgroup i. Although in a purely design-based framework θ_i is often viewed as a fixed but unknown quantity, the small sample size in area i necessitates use of a model to incorporate auxiliary information which can improve the accuracy of the small area estimator. In the related problem of disease mapping, Wakefield (2007) argued that multilevel models are effective for estimating disease rates in regions with small numbers of cases or exposure.

In many commonly used small area estimation methods, a two-stage hierarchical model is adopted. In the first stage, the data vector from area i, y_i, follows $y_i \mid \theta_i \sim f_1(y_i \mid \theta_i, \phi_1)$. The second-stage model relates the characteristic of interest θ_i to other areas and covariates. A general form for such a model is $\theta_i \mid x_i \sim f_2(\phi_2)$, where x_i is a vector of covariates for area i. The pairs (y_i, θ_i) are assumed to be independent. Under this model, θ_i is a random quantity and $\phi = (\phi_1, \phi_2)$ is a vector of unknown parameters. We want to find the best predictor of θ_i or of a function $h(\theta_i)$, and an estimator of the mean squared error of prediction under the model.

Consider the beta-binomial model, which is often used for binary response data. In this model, the parameter of interest θ_i is a proportion, often denoted by p_i. Conditionally on p_i, the n_i responses y_{ij} in area i are assumed to be independently generated from Bernoulli random variables with parameter p_i, so that, for $y_{i+} = \sum_{j=1}^{n_i} y_{ij}$, we have $y_{i+} \mid p_i \sim \text{Bi}(n_i, p_i)$ $(i = 1, \ldots, m)$. The second-stage model linking the proportions p_i across areas is $p_i \sim \text{Be}(\alpha, \beta)$, with $\alpha > 0$ and $\beta > 0$. As in the linear model, the pairs (y_{i+}, p_i) are assumed to be independent across areas. If the parameters α and β are known, then $p_i \mid y_{i+}, \alpha, \beta \sim \text{Be}(y_{i+} + \alpha, n_i - y_{i+} + \beta)$, and the best predictor of p_i is

$$\hat{p}_i^{\text{B}} = E(p_i \mid y_{i+}, \alpha, \beta) = \frac{y_{i+} + \alpha}{n_i + \alpha + \beta}.$$

The posterior variance of p_i, if the values of α and β are known, is

$$\begin{aligned} \text{var}\left(p_i \mid y_{i+}, \alpha, \beta\right) &= E\left\{\left(\hat{p}_i^{\text{B}} - p_i\right)^2 \mid y_{i+}, \alpha, \beta\right\} \\ &= \frac{(y_{i+} + \alpha)(n_i - y_{i+} + \beta)}{(n_i + \alpha + \beta + 1)(n_i + \alpha + \beta)^2}, \end{aligned}$$

which depends on area-specific data y_{i+}. The unconditional mean squared error of $\hat{p}_i^{\text{B}}$ is

$$\begin{aligned} \text{MSE}\left(\hat{p}_i^{\text{B}}\right) &= E\{\text{var}\left(p_i \mid y_{i+}, \alpha, \beta\right)\} \\ &= k_i(\alpha, \beta), \end{aligned}$$

where the expectation is with respect to the marginal distribution of y_{i+}. In the beta-binomial case, $k_i(\phi)$ with $\phi = (\alpha, \beta)$ has the closed-form expression

$$\begin{aligned} k_i(\alpha, \beta) &= E\left\{\left(\hat{p}_i^{\text{B}} - p_i\right)^2 \mid \alpha, \beta\right\} \\ &= \frac{\alpha}{(n_i + \alpha + \beta + 1)(n_i + \alpha + \beta)^2} \left\{n_i + \beta + \frac{n_i(n_i - 1)\beta}{(\alpha + \beta)(\alpha + \beta + 1)} + \frac{n_i(\beta - \alpha)}{\alpha + \beta}\right\}. \end{aligned}$$

However, for most other nonlinear cases, including most generalized linear mixed models, calculating $k_i(\phi)$ requires numerical integration.

If α and β are unknown, then $\hat{p}_i^{\text{B}}$ cannot be evaluated. Instead, the empirical Bayes, or the empirical best, predictor $\hat{p}_i^{\text{EB}} = (y_{i+} + \hat{\alpha})/(n_i + \hat{\alpha} + \hat{\beta})$ may be used, where the estimators $\hat{\alpha}$ and $\hat{\beta}$ are calculated from the data. The unconditional mean squared error of $\hat{p}_i^{\text{EB}}$ is

$$\text{MSE}\left(\hat{p}_i^{\text{EB}}\right) = k_i(\alpha, \beta) + E\left(\hat{p}_i^{\text{EB}} - \hat{p}_i^{\text{B}}\right)^2. \tag{1}$$

It follows from (1) that the use of the naive estimator $k_i(\hat{\alpha}, \hat{\beta})$ could lead to serious underestimation if the number of areas, m, is small.

For linear or generalized linear mixed models, Jiang et al. (2002) proposed a novel jackknife method for estimating $\text{MSE}\,(\hat{\theta}_i^{\text{EB}})$ in the case of independent pairs (y_i, θ_i) across areas. They used the orthogonal decomposition

$$\begin{aligned}\text{MSE}\,(\hat{\theta}_i^{\text{EB}}) &= E\,(\hat{\theta}_i^{\text{B}} - \theta_i)^2 + E\,(\hat{\theta}_i^{\text{EB}} - \hat{\theta}_i^{\text{B}})^2 \\ &= k_i(\phi) + M_{2i} \\ &= M_{1i} + M_{2i},\end{aligned} \tag{2}$$

where $\hat{\theta}_i^{\text{B}} = \hat{\theta}_i^{\text{B}}(\phi)$ and $\hat{\theta}_i^{\text{EB}} = \hat{\theta}_i^{\text{B}}(\hat{\phi})$, and then employed the jackknife method to estimate M_{1i} and M_{2i} separately. In the jth jackknife iteration, area j is deleted. The data from the remaining $(m-1)$ areas are used to calculate the estimator $\hat{\phi}_{(-j)}$ of ϕ. The delete-j estimator of θ_i is then given by $\hat{\theta}_{i(-j)}^{\text{EB}} = \hat{\theta}_i^{\text{B}}\{\hat{\phi}_{(-j)}\}$. The term M_{1i} in (2) is estimated by

$$\hat{M}_{1i} = k_i(\hat{\phi}) - \frac{m-1}{m}\sum_{j=1}^{m}\{k_i(\hat{\phi}_{(-j)}) - k_i(\hat{\phi})\}. \tag{3}$$

The quantity $\hat{M}_{1i}$ in (3) equals the naive estimator $k_i(\hat{\phi})$ plus a Quenouille (1956) bias correction. The term M_{2i} in (2) is estimated by

$$\hat{M}_{2i} = \frac{m-1}{m}\sum_{j=1}^{m}(\hat{\theta}_{i(-j)}^{\text{EB}} - \hat{\theta}_i^{\text{EB}})^2. \tag{4}$$

Jiang et al. (2002) showed that the jackknife estimator, $\text{mse}_{\text{J}}(\hat{\theta}_i^{\text{EB}}) = \hat{M}_{1i} + \hat{M}_{2i}$, is nearly unbiased in the sense of having unconditional bias of order $o(1/m)$ under regularity conditions.

The Jiang et al. (2002) jackknife was constructed to estimate the unconditional mean squared error given in (2). However, the leading term of order $O(1)$ in (3) is not area specific: it does not depend on y_i, even when the posterior variance of θ_i when ϕ is known depends on y_i. The Jiang et al. (2002) jackknife is thus, in general, biased for estimating the conditional mean squared error

$$\text{CMSE}\,(\hat{\theta}_i^{\text{EB}}) = E\{(\hat{\theta}_i^{\text{EB}} - \theta_i)^2 \mid y_i\}, \tag{5}$$

which depends on y_i in nonlinear models.

Booth & Hobert (1998) argued that inferences about θ_i should depend on area-specific data y_i and maintained that one should estimate the conditional mean squared error, since it corresponds to the posterior variance $\text{var}\,(\hat{\theta}_i^{\text{B}} \mid y_i, \phi)$ when ϕ is known; the conditional mean squared error can be considered as a frequentist analogue of the posterior variance. They wrote that 'it is our belief that much of the literature on this topic is confusing and perhaps misleading for practitioners, in part because of the emphasis on unconditional inference in the literature.' Fuller (1990) earlier proposed an estimator of conditional mean squared error in the linear case.

Rao (2003, § 9.4) proposed a modification of the Jiang et al. (2002) jackknife method that leads to a computationally simpler jackknife estimator with the leading term depending on y_i. We shall show that a modification of Rao's proposal leads to a nearly unbiased area-specific jackknife estimator of the conditional as well as the unconditional mean squared error.

2. Area-specific jackknife

Rao (2003, § 9.4) noted that $k_i(\phi) = E\{g_i(\phi, y_i)\}$, where $g_i(\phi, y_i) = \text{var}\,(\theta_i \mid y_i, \phi)$, and proposed applying the Quenouille (1956) bias-reduction method to $g_i(\hat{\phi}, y_i)$ instead of to $k_i(\hat{\phi})$. An estimator with leading term $g_i(\hat{\phi}, y_i)$ instead of $k_i(\hat{\phi})$ tracks the conditional mean squared error as recommended by Booth & Hobert (1998); in addition, since $E\{\text{CMSE}\,(\hat{\theta}_i^{\text{EB}})\} = \text{MSE}\,(\hat{\theta}_i^{\text{EB}})$, an unbiased estimator of the conditional mean squared error is also unbiased for estimating the unconditional mean squared error. Calculation of $k_i(\hat{\phi})$ often requires numerical integration of

$g_i(\hat{\phi}, y_i)$ using the marginal distribution of y_i, so it was expected that considerable computational simplification could be attained by eliminating the additional integrations. Rao (2003) estimated the term M_{1i} in (2) by

$$\hat{M}^*_{\mathrm{A}1i}(y_i) = g_i(\hat{\phi}, y_i) - \frac{m-1}{m}\sum_{j=1}^{m}\{g_i(\hat{\phi}_{(-j)}, y_i) - g_i(\hat{\phi}, y_i)\}. \tag{6}$$

However, since the function g_i depends on y_i, $\hat{M}^*_{\mathrm{A}1i}$ has bias of order $O(m^{-1})$, as shown below in Theorem 1. Consequently, Rao's (2003) area-specific jackknife estimator, given by $\mathrm{mse}^*_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}}) = \hat{M}^*_{\mathrm{A}1i}(y_i) + \hat{M}_{2i}$, where $\hat{M}_{2i}$ is given by (4), also has bias of order $O(m^{-1})$. Hence, it is not nearly unbiased.

To reduce the bias, we propose instead estimating M_{1i} by

$$\hat{M}_{\mathrm{A}1i}(y_i) = g_i(\hat{\phi}, y_i) - \sum_{j \neq i}^{m}\{g_i(\hat{\phi}_{(-j)}, y_i) - g_i(\hat{\phi}, y_i)\}. \tag{7}$$

This leads to the area-specific jackknife estimator

$$\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}}) = \hat{M}_{\mathrm{A}1i}(y_i) + \hat{M}_{2i}. \tag{8}$$

It is clear from (8) that the leading term of $\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}})$ is area specific. Theorem 1 gives properties of the area-specific jackknife when the maximum likelihood method is used to estimate ϕ. Conditions N1–N9 of Hoadley (1971) are standard conditions used to prove asymptotic normality of maximum likelihood estimators when observations are independent but not necessarily identically distributed. We assume in the theorem that the maximum likelihood estimators $\hat{\phi}$ and $\hat{\phi}_{(-j)}$ exist; if not, they may be set equal to ϕ^* for some fixed point in the interior of the parameter space, as done in Jiang et al. (2002).

THEOREM 1. *Assume that ϕ is an interior point of the parameter space Φ. Let $\sum_{j=1}^{m} \ell_j(\phi; y_j)$ be the loglikelihood function of ϕ, and let $\hat{\phi}$ be the maximum likelihood estimator of ϕ. Assume that conditions* N1–N9 *in* Hoadley (1971) *are met, and that $\gamma_j = E\{\ell_j''(\phi; y_j)\}$ are uniformly bounded away from* 0 *for $j = 1, \ldots, m$. In addition, assume that there exist $\varepsilon > 0$ and $\delta > 0$ such that $E\{|\ell'(\phi; y_j)|^{4+\delta}\}$, $E\{|\ell''(\phi; y_j)|^{4+\delta}\}$ and $E\{\sup_{c\in(-\varepsilon,\varepsilon)} |\ell'''(\phi + c; y_j)|^{4+\delta}\}$ are uniformly bounded for all j. Then, as $m \to \infty$,*

$$E\{\hat{M}^*_{\mathrm{A}1i}(y_i) - g_i(\phi, y_i)|y_i\} = -g_i'(\phi, y_i)\frac{\ell_i'(\phi; y_i)}{\sum_{j=1}^{m}\gamma_j} + r_1(y_i)\, o(m^{-1}),$$

$$E\{\hat{M}_{\mathrm{A}1i}(y_i) - g_i(\phi, y_i)|y_i\} = r_2(y_i)\, o(m^{-1}), \tag{9}$$

$$E\{\hat{M}_{2i} - (\hat{\theta}_i^{\mathrm{EB}} - \hat{\theta}_i^{\mathrm{B}})^2|y_i\} = r_3(y_i)\, o(m^{-1}), \tag{10}$$

where $|r_k(y_i)| \leqslant \max\{1, |\ell'(\phi; y_i)|^3, |\ell''(\phi; y_i)|^3, |\ell'''(\phi; y_i)|^3\} \equiv R(y_i)$ for $k = 1, 2, 3$.

The proof of Theorem 1 is given in the Appendix. It follows from Theorem 1 that

$$E\{\hat{M}_{\mathrm{A}1i}(y_i) - g_i(\phi, y_i)\} = o(m^{-1}), \quad E(\hat{M}_{2i}) = M_{2i} + o(m^{-1}).$$

Hence, while the unconditional bias of Rao's (2003) jackknife estimator $\mathrm{mse}^*_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}})$ is of order $O(m^{-1})$, the unconditional bias of $\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}})$ is of order $o(m^{-1})$, which is the same as that of the nearly unbiased jackknife estimator $\mathrm{mse}_{\mathrm{J}}(\hat{\theta}_i^{\mathrm{EB}})$ of Jiang et al. (2002). It also follows from (9) and (10) that

$$E\{\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}}) - \mathrm{CMSE}\,(\hat{\theta}_i^{\mathrm{EB}}) \mid y_i\} = o_p(m^{-1}).$$

Consequently, $\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}})$ can be used to estimate the conditional mean squared error $\mathrm{CMSE}\,(\hat{\theta}_i^{\mathrm{EB}})$, given in (5), with low conditional bias. On the other hand,

$$E\{k_i(\hat{\phi}) - g_i(\phi, y_i) \mid y_i\} = k_i(\phi) - g_i(\phi, y_i) + O_p(m^{-1}).$$

Hence, the conditional bias of $\mathrm{mse}_{\mathrm{J}}(\hat{\theta}_i^{\mathrm{EB}})$ is of order $O_p(1)$ and does not diminish as $m \to \infty$.

For the linear mixed model used in the Fay–Herriot (1979) estimator, the conditional and unconditional mean squared errors are approximately the same. In that situation, let $\bar{y}_i$ denote the estimator of the population mean in area i. With first-stage model $\bar{y}_i \mid \theta_i \sim N(\theta_i, \psi_i)$ and second-stage model $\theta_i \mid \beta, \sigma^2 \sim N(x_i^{\mathrm{T}}\beta, \sigma^2)$, the posterior variance of θ_i if $\phi = (\beta, \sigma^2)$ is known is $g_i(\phi, y_i) = k_i(\phi) = \psi_i\sigma^2/(\sigma^2 + \psi_i)$, which does not depend on the data in area i. Consequently, for the linear Fay–Herriot model, the area-specific jackknife in (8) is approximately the same as the Jiang et al. (2002) jackknife. In nonlinear models, however, the area-specific jackknife has low bias for the conditional as well as unconditional mean squared error, while the Jiang et al. (2002) jackknife has low bias only for the unconditional mean squared error.

The area-specific jackknife is easier to compute than the Jiang et al. (2002) jackknife, since it does not require finding the expected value of $g_i(\phi, y_i)$. When $k_i(\phi) = E\{g_i(\phi, y_i)\}$ must be calculated numerically, the Jiang et al. (2002) jackknife uses an additional $m + 1$ numerical integrations. The reduced computational complexity for the area-specific jackknife leads to less computation time; in many cases, it also leads to fewer numerical errors. The summands in (3), (4) and (7) are typically small and the sum is very sensitive to numerical errors. Computing $g_i(\phi, y_i)$, without taking the additional step of computing its expectation, can reduce the numerical errors of the jackknife.

Maximum likelihood estimation of ϕ is typically implemented using an iterative algorithm such as Newton–Raphson, which is subject to numerical inaccuracies (Householder, 1970, Ch. 4). If the estimate $\hat{\phi}$ from the full dataset has numerical errors, and those errors are propagated in the jackknifed quantities $\hat{\phi}_{(-j)}$, the jackknife corrections can result in decreased, rather than increased, accuracy of the estimated mean squared error. For use with the jackknife, $\hat{\phi}$ should be calculated to as close to machine precision as possible.

With a numerically accurate estimate $\hat{\phi}$, the Newton–Raphson algorithm may be used to accelerate the computation of $\hat{\phi}_{(-j)}$. In the jackknife iterations, the maximum likelihood estimate $\hat{\phi}$ from the full data may be used as an initial value, and Newton–Raphson with numerical derivatives used to calculate $\hat{\phi}_{(-j)}$ for each jackknife iteration. The results of Theorem 1 hold if the jackknife iterations use at least two steps of a root-finding algorithm with quadratic convergence or one step of an algorithm with cubic convergence; if desired, two-step Newton–Raphson can be used to estimate $\hat{\phi}_{(-j)}$ with initial value $\hat{\phi}$ instead of iterating to full convergence. One step of Newton–Raphson is not sufficient, however, since it does not provide the accuracy needed for the jackknife to remove the terms of order $1/m$ in the bias.

In small samples, $\hat{M}_{1i}$ in (3) or $\hat{M}_{\mathrm{A}1i}(y_i)$ in (7) may be negative. In that case, we recommend substituting $k_i(\hat{\phi})$ for $\hat{M}_{1i}$, and $g_i(\hat{\phi}, y_i)$ for $\hat{M}_{\mathrm{A}1i}(y_i)$.

3. Simulation study

We performed a simulation study to compare the area-specific jackknife method in (8) with the naive estimators of mean squared error and the jackknife estimator proposed in Jiang et al. (2002). The beta-binomial model described in § 1 was used to generate the data. A full factorial design was employed with four factors: $m \in \{20, 40, 60\}$, $\alpha \in \{0{\cdot}5, 1, 3\}$, $\beta \in \{0{\cdot}5, 1, 3\}$ and equal or unequal within-area sample sizes. The sample size for all areas was set to $n_i = 5$ when all within-area sample sizes were equal; for the unequal setting, half of the m small areas had $n_i = 3$ and the other half had $n_i = 7$.

All computations were performed in R, version 2·1·1 (R Development Core Team, 2008). Ten thousand simulation runs were performed for each factor combination. For computational simplicity, following Kleinman (1973) and Smith (1983), we calculated method-of-moments estimators of $\mu = \alpha/(\alpha + \beta)$ and $\eta = 1/(\alpha + \beta)$, restricting to the ranges $0 \leqslant \mu \leqslant 1$ and $\eta \geqslant 0$. When $\eta = 0$ there is no overdispersion. Let $\hat{\mu} = \sum_{i=1}^m y_{i+}/\sum_{i=1}^m n_i$ and $\hat{\sigma}^2 = \{\sum_{i=1}^m y_{i+}(y_{i+} - 1)\}/\{\sum_{i=1}^m n_i(n_i - 1)\} - \hat{\mu}^2$. Then $\hat{\eta} = \hat{r}/(1 - \hat{r})$, where $\hat{r} = \min[\hat{\sigma}^2\{\hat{\mu}(1 - \hat{\mu})\}^{-1}, 1]$ if $\hat{\mu}(1 - \hat{\mu}) \neq 0$ and $\hat{\sigma}^2 > 0$, and $\hat{r} = 0$ otherwise. When $n_i = 1$ for all areas, $\hat{\sigma}^2$ is not defined; we must have $n_i > 1$ in a sufficient number of areas so that the estimators are consistent as required by the Hoadley (1971) conditions in Theorem 1.

For each setting of the experimental factors with equal sample sizes, we calculated the average empirical unconditional mean squared error as

$$\text{EUMSE} = \frac{1}{10\,000m} \sum_{k=1}^{10\,000} \sum_{i=1}^{m} \left(\hat{p}_{ik}^{\text{EB}} - p_{ik}\right)^2,$$

where p_{ik} is the true value of p_i generated for area i in the kth simulation run, and $\hat{p}_{ik}^{\text{EB}}$ is the estimate of p_i obtained in run k. The empirical conditional mean squared error for $j = 0, \ldots, 5$, denoted by ECMSE(j), was calculated by averaging $(\hat{p}_{ik}^{\text{EB}} - p_{ik})^2$ over the areas and runs in which $y_{ik+} = j$. Other choices for the empirical estimators of the conditional mean squared error were also explored and gave similar results. The average unconditional relative bias of each estimator was calculated as $100\{\sum_{i=1}^m \sum_{k=1}^{10\,000} \text{mse}_{ik}/(10\,000m) - \text{EUMSE}\}/\text{EUMSE}$, where mse_{ik} is the value of the estimator mse for area i on run k. The conditional relative bias was calculated similarly, using the average of the estimate over runs and areas in which $y_{ik+} = j$. We calculated the unconditional coefficient of variation as

$$\text{CV} = \frac{\left\{(10\,000m)^{-1} \sum_{k=1}^{10\,000} \sum_{i=1}^{m} (\text{mse}_{ik} - \text{EUMSE})^2\right\}^{1/2}}{\text{EUMSE}};$$

for the conditional coefficient of variation we used the empirical mean squared error of the values of mse corresponding to area i in which $y_{ik+} = j$. When sample sizes were unequal, similar calculations were performed separately for each value of n_i.

Results from the simulation study demonstrate empirically the theoretical result that the naive estimators of the unconditional mean squared error underestimate the true value. The patterns were similar for all of the settings studied; partial results are presented in Table 1. The relative biases of $g_i(\hat{\phi}, y_i)$ and $k_i(\hat{\phi})$ were negative for every situation studied. The unconditional relative bias is much smaller for the Jiang et al. (2002) and area-specific jackknife estimators, and the bias decreases as m increases.

However, the unconditional coefficient of variation can be significantly smaller for mse_{J} than for mse_{AS}, and the difference persists as m increases. The area-specific jackknife has greater coefficient of variation because the true values of $g_i(\phi, y_i)$ differ for each value of y_i.

A different pattern emerges for estimating the conditional mean squared error. The area-specific jackknife has small conditional relative bias, and the bias decreases with increasing m. On the other hand, the Jiang et al. (2002) jackknife has a strong pattern for the conditional bias; for the results in Table 1 with $\alpha = \beta = 1$, the conditional bias is large and positive when y_{i+} is small or large, and the conditional bias is large and negative when y_{i+} is close to $n_i/2$. Moreover, the

Table 1. *Relative bias and coefficient of variation of mean squared error estimators from simulation study when $\alpha = \beta = 1$*

	Relative bias				*Coefficient of variation*			
	G	K	J	AS	G	K	J	AS
				$n_i = 5$ for all areas; $m = 20$				
Unconditional	−16	−16	3	3	32	19	17	29
Conditional, $y_{i+} = 0$	−26	18	52	6	30	21	60	40
Conditional, $y_{i+} = 1$	−13	−20	−4	2	18	22	13	12
Conditional, $y_{i+} = 2$	−13	−32	−20	2	21	33	22	14
Conditional, $y_{i+} = 3$	−13	−32	−20	2	21	33	22	13
Conditional, $y_{i+} = 4$	−12	−20	−3	3	17	22	13	12
Conditional, $y_{i+} = 5$	−26	19	54	7	30	22	62	40
				$n_i = 5$ for all areas; $m = 40$				
Unconditional	−7	−7	1	1	28	9	6	26
Conditional, $y_{i+} = 0$	−15	34	51	1	19	35	51	20
Conditional, $y_{i+} = 1$	−6	−13	−6	1	9	13	7	5
Conditional, $y_{i+} = 2$	−6	−26	−22	−0	12	27	22	8
Conditional, $y_{i+} = 3$	−6	−26	−22	0	11	27	22	8
Conditional, $y_{i+} = 4$	−5	−12	−5	1	9	13	7	5
Conditional, $y_{i+} = 5$	−13	37	53	2	18	37	54	20
				$n_i \in \{3, 7\}$; $m = 40$				
$n_i = 3$								
Unconditional	−15	−17	2	2	29	20	13	22
Conditional, $y_{i+} = 0$	−20	1	26	3	23	11	28	18
Conditional, $y_{i+} = 1$	−6	−25	−14	2	19	26	17	16
Conditional, $y_{i+} = 2$	−11	−28	−16	1	21	29	19	15
Conditional, $y_{i+} = 3$	−27	−9	22	3	29	16	27	24
$n_i = 7$								
Unconditional	−12	−13	4	4	33	16	23	32
Conditional, $y_{i+} = 0$	−26	36	76	9	31	38	89	59
Conditional, $y_{i+} = 1$	−13	−9	12	5	15	13	25	21
Conditional, $y_{i+} = 2$	−9	−24	−12	4	16	25	18	11
Conditional, $y_{i+} = 3$	−6	−29	−20	4	17	30	23	13
Conditional, $y_{i+} = 4$	−8	−31	−22	1	17	32	25	12
Conditional, $y_{i+} = 5$	−10	−27	−15	0	15	28	20	9
Conditional, $y_{i+} = 6$	−12	−9	11	4	13	12	25	18
Conditional, $y_{i+} = 7$	−28	40	83	13	34	43	99	63

Methods of estimating mean squared error: G, $g_i(\hat{\phi}, y_i)$; K, $k_i(\hat{\phi})$; J, Jiang et al. (2002) jackknife; AS, area-specific jackknife.

conditional bias of the Jiang et al. (2002) method does not necessarily decrease as m increases. Table 1 also shows that the conditional coefficient of variation of the area-specific jackknife is much smaller than the conditional coefficient of variation of the Jiang et al. (2002) jackknife.

Similar results occur for other values of α and β with both equal and unequal within-area sample sizes: the area-specific jackknife has lower conditional bias but significantly higher unconditional coefficient of variation, while the Jiang et al. (2002) jackknife has more stability unconditionally but at the cost of high conditional bias and high conditional coefficient of variation.

We also compared the area-specific jackknife in (8) with Rao's (2003) original proposal of $\text{mse}^*_{\text{AS}}(\hat{\theta}_i^{\text{EB}})$. The second term, $\hat{M}_{2i}$, is the same for the two methods, so we compared the relative biases of $\hat{M}^*_{\text{A}1i}(y_i)$ in (6) and $\hat{M}_{\text{A}1i}(y_i)$ in (7) for estimating $g_i(\phi, y_i)$. In almost all situations, the estimator in (7) had smaller relative bias for estimating $g_i(\phi, y_i)$, both conditionally and unconditionally. For example, for the simulation setting in Table 1 with $n_i = 5$ and $m = 40$,

$\hat{M}_{A1i}(y_i)$ had unconditional relative bias 0·3 while $\hat{M}^*_{A1i}(y_i)$ had unconditional relative bias −1·5.

In Table 1, the conditional relative bias for the area-specific jackknife, mse_{AS}, is higher for the simulations in which $n_i = 5$ for all areas when $y_{i+} = 0$ and $y_{i+} = 5$ than for the other values of y_{i+}. This pattern diminishes as m increases. When m is small, the large conditional bias for $y_{i+} = 5$ arises because the extreme value of 5 is less common in the data. There are thus simulation runs in which the full-data estimates $\hat{\mu}$ and $\hat{\eta}$ are in the interior of the parameter space, but deleting area i with $y_{i+} = 5$ results in $\hat{\eta}_{(-i)} = 0$, which is on the boundary of the parameter space. In such a situation, then, $\hat{\theta}^{\text{EB}}_{i(-i)}$ will not depend on y_{i+} but will instead be the average of the responses in the other areas, and $(\hat{\theta}^{\text{EB}}_{i(-i)} - \hat{\theta}^{\text{EB}}_i)^2$ and $\hat{M}_{2i}$ will be very large. We explored this using an asymptotically equivalent modification of $\hat{M}_{2i}$, $\hat{M}^*_{2i} = \sum_{j \neq i}(\hat{\theta}^{\text{EB}}_{i(-j)} - \hat{\theta}^{\text{EB}}_i)^2$, but $\hat{M}^*_{2i}$ tends to overcompensate in small samples and results in a negative conditional bias for areas in which $y_{i+} = 0$ or $y_{i+} = 5$. If one considers only the simulation runs in which all parameter estimates are in the interior of the parameter space, the area-specific jackknife has low conditional bias even at the extremes.

When a jackknifed parameter estimate is on the boundary of the parameter space, neither the Jiang et al. (2002) jackknife nor the area-specific jackknife performs well; Theorem 1 assumes that the probability of such an occurrence is negligible as $m \to \infty$, and the theoretical results depend on $\hat{\phi}$ being an interior point. We suggest viewing either jackknife as a conservative estimator of mean squared error when any of the parameter estimates are on the boundary.

4. APPLICATION TO DISEASE MAPPING

Many governmental health agencies produce cancer atlases which show the incidence of cancer by geographical regions such as counties. Often relative risks are presented so that areas with unusually high cancer incidence can be identified. Let y_i be the number of observed cases of cancer in area i, and e_i be the expected number of cases in area i; then the standardized mortality ratio is y_i/e_i. However, the standardized mortality ratio has high variability in areas in which e_i is small. Clayton & Kaldor (1987) and Wakefield (2007), among others, discussed using small area estimation techniques to obtain more reliable estimators of relative risk.

We apply the area-specific jackknife to the lip cancer data from Clayton & Kaldor (1987). The data consist of numbers of cases of lip cancer in 56 Scottish counties from 1975 to 1980. The expected number of cases is calculated using the reported standardized mortality ratio for counties with lip cancer cases; for counties with no lip cancer cases e_i is taken from Table 6 of Breslow & Clayton (1993). We consider two models for the response: a Poisson-gamma model and a Poisson-lognormal model. For both models, assume that, conditionally on the relative risk θ_i for county i, the count y_i of lip cancer cases in county i follows a Poisson distribution with mean $e_i\theta_i$. Wakefield (2007) evaluated the fits of various models for the lip cancer data; although he preferred a spatial model for inference about regression parameters, he argued that the predictive nature of disease mapping allows great flexibility for which the model is adopted. The Poisson-lognormal model is easily modified to include covariates or spatial information, as done in Breslow & Clayton (1993) for the lip cancer data; however, since our primary interest is in estimating the mean squared error, we fit the simple lognormal model without covariates for easier comparison with the Poisson-gamma model.

In the Poisson-gamma model, assume that θ_i is a gamma random variable with scale parameter α and shape parameter ν. Then $\hat{\theta}^{\text{B}}_i = (y_i + \nu)/(e_i + \alpha)$ and the area-specific variance is $g_i(\alpha, \nu, y_i) = (y_i + \nu)/(e_i + \alpha)^2$. Since g_i has a closed form, both the area-specific jackknife and

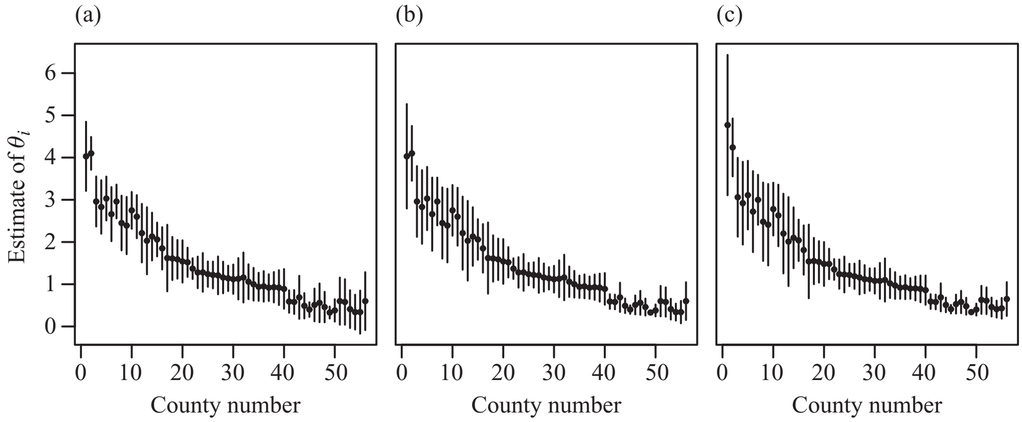

Fig. 1. Lip cancer illustration. The vertical line for each county displays $\hat{\theta}_i^{\mathrm{EB}} \pm \surd\ \mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}})$, with $\hat{\theta}_i^{\mathrm{EB}}$ marked by the circle, for (a) Poisson-gamma model with Jiang et al. (2002) jackknife, (b) Poisson-gamma model with area-specific jackknife and (c) Poisson-lognormal model with area-specific jackknife. The county numbers are given in Breslow & Clayton (1993).

the Jiang et al. (2002) jackknife are easy to calculate. For this model, $k_i(\alpha, \nu) = \nu/\{\alpha(e_i + \alpha)\}$. The maximum likelihood estimate of (α, ν) is $(1{\cdot}316,\ 1{\cdot}874)$. Figures 1(a) and (b) display $\hat{\theta}_i^{\mathrm{EB}} \pm \surd\,\mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}})$ using the two jackknife methods to estimate the mean squared error. The Jiang et al. (2002) jackknife, estimating $k_i(\alpha, \nu) + M_{2i}$, smooths out the conditional variances over the areas. Consider County 1, with $e_1 = 1{\cdot}4$ and $y_1 = 9$, and County 56, with $e_{56} = 1{\cdot}8$ and $y_{56} = 0$. Since the e_i values are similar, the Jiang et al. (2002) jackknife gives approximately the same estimated mean squared error for the two counties. However, the estimated conditional mean squared error using the area-specific jackknife is much larger for County 1 than for County 56. County 1 has an unusually high number of observed cancer cases relative to e_1, and the larger area-specific mean squared error reflects the higher uncertainty associated with this unusual data point.

For the Poisson-lognormal model, assume that for the second stage $\log(\theta_i) \sim N(\mu, \sigma^2)$. Let $h_i(k, y_i, \mu, \sigma) = E_Z[\exp\{-e_i \exp(\sigma Z + \mu) + (y_i + k)(\sigma Z + \mu)\}]$, where $Z \sim N(0, 1)$. The best predictor of θ_i is

$$\hat{\theta}_i^{\mathrm{B}} = E(\theta_i \mid y_i) = \frac{h_i(1, y_i, \mu, \sigma)}{h_i(0, y_i, \mu, \sigma)},$$

with conditional variance

$$g_i(\mu, \sigma, y_i) = \frac{h_i(2, y_i, \mu, \sigma)}{h_i(0, y_i, \mu, \sigma)} - (\hat{\theta}_i^{\mathrm{B}})^2.$$

Each of these quantities must be calculated numerically. We used 21-point Gauss–Hermite quadrature to calculate the integrals. The maximum likelihood estimates are $\hat{\mu} = 0{\cdot}02$ and $\hat{\sigma} = 0{\cdot}86$. Figure 1(c) displays $\hat{\theta}_i^{\mathrm{EB}} \pm \surd\,\mathrm{mse}_{\mathrm{AS}}(\hat{\theta}_i^{\mathrm{EB}})$ for each county. The estimates are similar to those from the Poisson-gamma model.

To use the Jiang et al. (2002) jackknife with the Poisson-lognormal model,

$$k_i(\mu, \sigma) = E_{y_i}\{g_i(\mu, \sigma, y_i)\} = \sum_{j=0}^{\infty} g_i(\mu, \sigma, j)\,\mathrm{pr}(y_i = j)$$

needs to be calculated numerically using the discrete marginal distribution of y_i. Each term in the sum requires three numerical integrations, and enough terms in the sum must be calculated until an integral approximation of the remainder is sufficiently accurate. With the Poisson-lognormal model, the area-specific jackknife is much simpler to compute and is subject to fewer numerical errors. We have not calculated the Jiang et al. (2002) jackknife for the Poisson-lognormal case because of the above-noted computational complexity and proneness to numerical errors. Lohr (2007) showed that maximum likelihood calculations in the Poisson-lognormal model are very sensitive to the number of quadrature points used; thus, reducing the number of computational steps by using the area-specific jackknife allows more control over the numerical errors.

5. DISCUSSION

All of the jackknife methods presented in this paper estimate the model-based mean squared error given in (2). Longford (2007) argued that, for mean squared error estimation, the quantities θ_i should be treated as fixed effects and proposed estimating the design-based mean squared error by taking a weighted average of the estimated design-based mean squared error and the estimated model-based mean squared error. This approach is interesting, but is not easily extended to complex models such as those considered in this paper because of difficulties in determining the optimal weights. Following standard practice in small area estimation, we treat the quantities θ_i as realizations of a random process from a model with random small area effects and, consequently, estimate the mean squared error under this model. Since all inferences are performed under the model, it is important in practice to check the validity of the model.

Resampling methods for estimating the mean squared error of small area predictors allows one to estimate the extra variability due to estimating the parameters ϕ without having to calculate the extra terms analytically. Even in the Fay–Herriot (1979) model, the extra terms differ when a different estimator of σ^2 is used (Datta et al., 2005).

The area-specific jackknife is a promising estimator for both conditional and unconditional mean squared error in small area estimation problems and other problems involving prediction from a generalized linear mixed model. Generalized linear mixed models and other two-stage models often present many computational challenges for implementation; the area-specific jackknife provides a method for estimating the mean squared error of predicted values with reduced computational complexity.

ACKNOWLEDGEMENT

This research was partially supported by grants from the U.S. National Science Foundation and the Natural Sciences and Engineering Research Council of Canada. We thank the referees, associate editor and Professor D. M. Titterington for their helpful comments and suggestions.

APPENDIX

Proof of Theorem 1

We prove the theorem for a univariate parameter ϕ. The multivariate case is similar. In the following, let ℓ'_j, ℓ''_j and ℓ'''_j denote $\ell'_j(\phi, y_j)$, $\ell''_j(\phi, y_j)$ and $\ell'''_j(\phi, y_j)$, respectively; let $\Gamma = \sum_{j=1}^{m} \gamma_j$, and let $r(y_i)$ denote a generic remainder term that is bounded in absolute value by $R(y_i)$.

We examine the terms in (6) and (7) separately. First note that

$$g_i(\hat{\phi}, y_i) - g_i(\phi, y_i) = g'_i(\phi, y_i)(\hat{\phi} - \phi) + \frac{g''_i(\phi, y_i)}{2}(\hat{\phi} - \phi)^2 + \frac{g'''_i(\phi^*, y_i)}{6}(\hat{\phi} - \phi)^3, \tag{A1}$$

for some ϕ^* between ϕ and $\hat\phi$. Also,

$$g_i\left(\hat\phi_{(-j)}, y_i\right) - g_i(\hat\phi, y_i) = g_i'(\hat\phi, y_i)\left(\hat\phi_{(-j)} - \hat\phi\right) + \frac{g''(\hat\phi, y_i)}{2}\left(\hat\phi_{(-j)} - \hat\phi\right)^2 + \frac{g'''\left(\hat\phi^*_{(-j)}, y_i\right)}{6}\left(\hat\phi_{(-j)} - \hat\phi\right)^3, \tag{A2}$$

for some $\hat\phi^*_{(-j)}$ between $\hat\phi_{(-j)}$ and $\hat\phi$.

We use Householder's (1970, Theorem 4.4.1) method to approximate the solution $\hat\phi$ of the equation $f(z) = \sum_{j=1}^m \ell_j'(z, y_j) = 0$: in iteration $(\nu + 1)$,

$$z_{\nu+1} = z_\nu - \frac{f(z_\nu)}{f'(z_\nu)}\left[1 + \frac{f(z_\nu) f''(z_\nu)}{2\{f'(z_\nu)\}^2}\right].$$

Taking initial value $z_\nu = \phi$, we have

$$\begin{aligned}
\hat\phi - \phi &= -\frac{\sum_{j=1}^m \ell_j'}{\sum_{j=1}^m \ell_j''}\left\{1 + \frac{\sum_{k=1}^m \ell_k' \sum_{r=1}^m \ell_r'''}{2\left(\sum_{k=1}^m \ell_k''\right)^2}\right\} + O_p(|\hat\phi - \phi|^3) \\
&= -\frac{\sum_{j=1}^m \ell_j'}{\Gamma}\left\{1 + \frac{\sum_{k=1}^m \ell_k' \sum_{r=1}^m \ell_r'''}{2\left(\sum_{k=1}^m \ell_k''\right)^2}\right\}\left(1 - \frac{\sum_{k=1}^m \ell_k'' - \Gamma}{\sum_{k=1}^m \ell_k''}\right) + O_p(|\hat\phi - \phi|^3),
\end{aligned} \tag{A3}$$

because of the cubic convergence of the algorithm and the conditions on the derivatives of ℓ. Thus, using Theorem 2.1 of Jiang et al. (2002), we have

$$E(\hat\phi - \phi \mid y_i) = \frac{-\ell_i'(\phi; y_i) + p}{\Gamma} + r(y_i)\, o(m^{-1}), \tag{A4}$$

where

$$p = \frac{\sum_{j=1}^m E(\ell_j' \ell_j'')}{\Gamma} - \frac{\sum_{j=1}^m \sum_{k=1}^m E\{(\ell_j')^2\} E(\ell_k''')}{2\Gamma^2}.$$

Similarly,

$$\begin{aligned}
\hat\phi_{(-j)} - \hat\phi &= \frac{\ell_j'(\hat\phi; y_j)}{\sum_{k\neq j}^m \ell_k''(\hat\phi; y_k)}\left[1 - \frac{\ell_j'(\hat\phi; y_j)\sum_{k\neq j} \ell_k'''(\hat\phi; y_k)}{2\left\{\sum_{k\neq j}^m \ell_k''(\hat\phi; y_k)\right\}^2}\right] + O_p\left(|\hat\phi_{(-j)} - \hat\phi|^3\right) \\
&= \frac{\ell_j' + \ell_j''(\hat\phi - \phi) + \ell_j''(\hat\phi - \phi)^2/2}{\Gamma}\left[1 - \frac{\ell_j'(\hat\phi; y_j)\sum_{k\neq j} \ell_k'''(\hat\phi; y_k)}{2\left\{\sum_{k\neq j}^m \ell_k''(\hat\phi; y_k)\right\}^2}\right] \\
&\quad \times \left\{1 - \frac{\sum_{k\neq j}^m \ell_k''(\hat\phi; y_k) - \Gamma}{\sum_{k\neq j}^m \ell_k''(\hat\phi; y_k)}\right\} + O_p\left(|\hat\phi_{(-j)} - \hat\phi|^3\right).
\end{aligned} \tag{A5}$$

Evaluating term by term gives

$$\sum_{j\neq i}^m E\left(\hat\phi_{(-j)} - \hat\phi \mid y_i\right) = \frac{-\ell_i'(\phi; y_i) + p}{\Gamma} + r(y_i)\, o(m^{-1}), \tag{A6}$$

$$E\left(\hat\phi_{(-i)} - \hat\phi \mid y_i\right) = \frac{\ell_i'(\phi; y_i)}{\Gamma} + r(y_i) o(m^{-1}). \tag{A7}$$

The remaining terms in $E\{\hat M_{\mathrm{A}1i} - g_i(\phi, y_i) \mid y_i\}$ and $E\{\hat M^*_{\mathrm{A}1i} - g_i(\phi, y_i) \mid y_i\}$, involving g'', are similarly shown to be of order $O_p(\Gamma^{-2})$. Thus, combining (A1), (A2), (A4), (A6) and (A7), we have

$$\begin{aligned}
E\{\hat M^*_{\mathrm{A}1i} - g_i(\phi, y_i) \mid y_i\} &= -g_i'(\phi, y_i)\ell_i'(\phi; y_i)/\Gamma + r(y_i)\, o(m^{-1}), \\
E\{\hat M_{\mathrm{A}1i} - g_i(\phi, y_i) \mid y_i\} &= r(y_i)\, o(m^{-1}).
\end{aligned}$$

To show that $\hat{M}_{2i}$ is approximately unbiased conditionally, let $h(\phi, y_i) = E(\theta_i \mid y_i, \phi)$, and note that

$$\hat{\theta}^{\mathrm{EB}}_{i(-j)} - \hat{\theta}^{\mathrm{EB}}_i = h\big(\hat{\phi}_{(-j)}, y_i\big) - h(\hat{\phi}, y_i)$$
$$= h'(\hat{\phi}, y_i)\big(\hat{\phi}_{(-j)} - \hat{\phi}\big) + \frac{h''\big(\hat{\phi}^*_{(-j)}, y_i\big)}{2}\big(\hat{\phi}_{(-j)} - \hat{\phi}\big)^2,$$

for some $\hat{\phi}^*_{(-j)}$ between $\hat{\phi}_{(-j)}$ and $\hat{\phi}$. Using (A3) and (A5) along with an additional Taylor expansion of $h'(\hat{\phi}, y_i)$ about ϕ, we obtain

$$\sum_{j=1}^{m} E\big\{\big(\hat{\theta}^{\mathrm{EB}}_{i(-j)} - \hat{\theta}^{\mathrm{EB}}_i\big)^2 \mid y_i\big\} = \{h'(\phi, y_i)\}^2 \frac{\sum_{j=1}^m E\{(\ell'_j)^2\}}{\Gamma^2} + r(y_i)\, o(m^{-1}).$$

Similarly,

$$E\big\{\big(\hat{\theta}^{\mathrm{EB}}_i - \hat{\theta}^{B}_i\big)^2 \mid y_i\big\} = \{h'(\phi, y_i)\}^2 \frac{\sum_{j=1}^m E\{(\ell'_j)^2\}}{\Gamma^2} + r(y_i)\, o(m^{-1}).$$

REFERENCES

BOOTH, J. G. & HOBERT, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *J. Am. Statist. Assoc.* **93**, 262–72.

BRESLOW, N. E. & CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.

CLAYTON, D. & KALDOR, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–81.

DATTA, G. S., RAO, J. N. K. & SMITH, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92**, 183–96.

FAY, R. E. & HERRIOT, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Am. Statist. Assoc.* **74**, 269–77.

FULLER, W. A. (1990). Prediction of true values for the measurement error model. In *Statistical Analysis of Measurement Error Models and Applications*, Contemporary Mathematics **112**, Ed. P. J. Brown and W. A. Fuller, pp. 41–57. Providence, RI: American Mathematical Society.

HOADLEY, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.* **42**, 1977–91.

HOUSEHOLDER, A. S. (1970). *The Numerical Treatment of a Single Nonlinear Equation*. New York: McGraw-Hill.

JIANG, J., LAHIRI, P. & WAN, S.-M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Ann. Statist.* **30**, 1782–810.

KLEINMAN, J. C. (1973). Proportions with extraneous variance: single and independent samples. *J. Am. Statist. Assoc.* **68**, 46–54.

LOHR, S. (2007). Discussion of papers on resampling methods in small area estimation. In *Proc. Survey Res. Meth. Sect. Am. Statist. Assoc.,* pp. 3265–7. Alexandria, VA: American Statistical Association.

LONGFORD, N. (2007). On standard errors of model-based small area estimators. *Survey Methodol.* **33**, 69–79.

QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353–60.

RAO, J. N. K. (2003). *Small Area Estimation*. New York: Wiley.

R DEVELOPMENT CORE TEAM (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

SMITH, D. M. (1983). Algorithm AS 189: Maximum likelihood estimation of the parameters of the beta binomial distribution. *Appl. Statist.* **32**, 196–204.

WAKEFIELD, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics* **8**, 158–83.

[*Received October* 2007. *Revised August* 2008]