$$\begin{aligned} \text{pr}(S = s) &= \frac{n!}{s!(n-s)!}\alpha\theta^s(1-\theta)^{n-s}, \\ \text{pr}(S = 0) &= 1 - \alpha\{1 - (1-\theta)^n\}. \end{aligned} \tag{2}$$

Distribution (2) reduces to the binomial distribution for $\alpha = 1$, but $E(S) = n\pi$ for $0 < \alpha \leqslant 1$. The parameters $\alpha$ and $\theta$ are poorly identified from a set of indiscernible seeds. However, the seeds on the same plate are not marginally independent since $\text{pr}(Y_{ij} = Y_{ik} = 1) = \alpha\theta_i^2$, and the marginal dependence ratio

$$\frac{\text{pr}(Y_{ij} = Y_{ik} = 1)}{\text{pr}(Y_{ij} = 1)\,\text{pr}(Y_{ik} = 1)} = \frac{1}{\alpha} = \tau.$$

Distribution (2) is overdispersed compared with the binomial distribution exactly as the beta–binomial; more particularly,

$$\text{var}(S) = n\pi(1-\pi)\left\{1 + (n-1)\frac{\pi}{1-\pi}(\tau - 1)\right\}.$$

For the data analysed in Section 6.3 the estimate of the overdispersion factor based on the Pearson statistic is 1.862 leading to $\tilde{\tau} = 1.057$ and $\tilde{\alpha} = 0.946$. The standard errors of the GLM estimates adjusted for the overdispersion are still smaller than the corresponding standard errors for the GLMM and the HGLM models.

For data with explanatory variables varying between seeds or with many plates with the same number of seeds it is possible to do likelihood inference for both $\tau$ and the regression parameters. The marginal dependence ratio of order $k$,

$$\frac{\text{pr}(Y_{ij_1} = \ldots = Y_{ij_k} = 1)}{\text{pr}(Y_{ij_1} = 1)\ldots\text{pr}(Y_{ij_k} = 1)} = \tau^{k-1},$$

where $2 \leqslant k \leqslant n$. Ekholm *et al.* (1995) referred to this as a homogeneous association structure and fitted it, with and without regression, to familial data for a disease among siblings.

**Andrew Gelman** (University of California, Berkeley): This paper nicely illustrates the convergence of classical and Bayesian methods for increasingly complex data analyses.

Setting aside computational issues, and in the absence of prior information, the main difference between the methods in this paper and fully Bayesian inference (e.g. as described by Zeger and Karim (1991)) is that the latter accounts for uncertainty in the hierarchical variance parameters. This can make a practical difference if the variances are near 0 and have highly skewed posterior distributions. In some cases, the usual point estimates (including those described in the paper) of the hierarchical variance are 0 (see, for example, the educational testing study described in Rubin (1981) and also discussed by Draper (1995) and Gelman *et al.* (1995), section 5.5). Setting a variance parameter to 0 is generally *not* desirable because it would lead to falsely precise estimates of random effects. Setting the variance to some non-zero value would require additional work which, in practice, would not be done since it would offer no advantages over the Bayesian approach of averaging over the uncertainty in the variance parameter.

It might be argued that such examples—in which the restricted maximum likelihood (REML) estimate of the hierarchical variance is at or near 0—are pathological and unlikely to occur in practice. I would argue, in contrast, that all hierarchical generalized linear models aspire to this status, as informative covariates are added to a model and the hierarchical variance decreases. For example, in forecasting US Presidential elections, the 'individual level' variables are at the state level and the 'hierarchical' variables are at the regional and national level (see Boscardin and Gelman (1996) and Gelman *et al.* (1995), section 13.2). Including better national level variables, such as economic indicators and opinion poll results, leads to a reduction in the hierarchical variance parameter corresponding to nationwide vote swings. At no point, however, would we want to set the variance to exactly 0 (even if that were the REML estimate), because this would critically understate the forecast uncertainty.

To the extent that the methods in this paper give different answers from the full Bayesian treatment, I would trust the latter. In the future, complexities such as Appendixes A–G of the paper will become unnecessary. For example, in mapping for small area estimation, consider the progression from 'empirical Bayes'—i.e. using point estimates for variance components—in Clayton and Kaldor (1987) to full Bayes in Clayton and Bernardinelli (1992).