Journal of Educational and Behavioral Statistics
December 2007, Vol. 32, No. 4, pp. 419–430
DOI: 10.3102/1076998607302629

Redundant Overdispersion Parameters in Multilevel Models for Categorical Responses

Anders Skrondal
London School of Economics
Norwegian Institute of Public Health

Sophia Rabe-Hesketh
University of California, Berkeley
University of London

In some distributions, such as the binomial distribution, the variance is determined by the mean. However, in practice, overdispersion is often observed where the variance is larger than that predicated by the mean, and underdispersion is sometimes observed where the variance is smaller. It is well known that overdispersion or underdispersion cannot be modeled for dichotomous responses having a Bernoulli distribution. Redundant overdispersion parameters are nevertheless often included when multilevel or hierarchical models for categorical responses are estimated using quasi-likelihood methods and in generalized estimating equations. This may be due to the popularity of an algorithmic model formulation.

Keywords: *overdispersion; underdispersion; extrabinomial variation; quasi-likelihood; scale parameter; multilevel model; hierarchical model; generalized linear mixed model; generalized estimating equations*

1. Introduction

Cox (1983) pointed out that analysis of data via single-parameter distributions such as the Poisson, binomial, and exponential distributions implies that the variance is determined by the mean. A very common practical complication is the presence of overdispersion, or more rarely underdispersion, violating the variance–mean relationship. Overdispersion occurs when the variance is larger than that predicated by the mean and underdispersion when the variance is smaller.

In this article, we focus on counts of events or "successes" that can occur only at discrete times or "trials." The observed count y is then less than or equal to the number of trials n, and the response can be expressed as a sample proportion $p = y/n$. We will focus mainly on overdispersion because it is more prevalent than underdispersion.

The purpose of the article is to emphasize that overdispersion (or underdispersion) is impossible when $n = 1$. This is true in very simple models as well as in complex models, although it is certainly less transparent in complex models. Perhaps for this reason, the inclusion of redundant overdispersion parameters, or more generally "scale parameters," is common in multilevel or hierarchical modeling.

In Section 2, we introduce single-level models and the notion of overdispersion. In Section 3, we investigate the handling of overdispersion in multilevel modeling. Finally, we end with a short conclusion.

2. Overdispersion in Single-Level Models

2.1. Proportions and the Binomial Distribution

If trials are independent and the probability of success π is constant over trials, the number of successes Y in n trials has a binomial distribution; $Y \sim \text{Binomial}(n, \pi)$.

In generalized linear models for proportions, this distribution is typically specified for the count of successes Y_i in n_i trials for unit i, but now conditional on covariates $\mathbf{x}_i$,

$$Y_i|\mathbf{x}_i \sim \text{Binomial}(\mathrm{n_i}, \pi_\mathrm{i}).$$

The covariates determine the probability π_i via

$$\pi_i \equiv E(P_i|\mathbf{x}_i) = h(\eta_i), \quad \eta_i = \mathbf{x}_i'\boldsymbol{\beta}, \tag{1}$$

where $P_i = Y_i/n_i$, $h(\cdot)$ is the inverse of a link function such as a logit or probit, and η_i is the linear predictor. This specification simultaneously determines the variance as a function of the expectation

$$\text{Var}(P_i|\mathbf{x}_i) = \frac{\pi_i(1-\pi_i)}{n_i} = \frac{E(P_i|\mathbf{x}_i)[1-E(P_i|\mathbf{x}_i)]}{n_i}. \tag{2}$$

However, in practice, either the independence or constant probability assumption is often violated, leading to overdispersion (variance larger than the binomial variance) or underdispersion (variance smaller than the binomial variance).

In the remainder of this section, we do not explicitly condition on $\mathbf{x}_i$ or use the i subscript, but the results also hold if π_i depends on covariates on which we condition. Consider the case in which the probability of success is not a constant π over trials but varies systematically with random unobserved covariates. Because these covariates are not included in the model, the parameter π behaves like a random variable with a probability distribution. Let Y given π be binomial, $Y|\pi \sim \text{Binomial}(\mathrm{n}, \pi)$, and let π have some distribution with mean $\bar{\pi}$ and

positive variance σ^2. The expectation of P becomes

$$E(P) = E_\pi[E(P|\pi)] = E_\pi[\pi] = \bar{\pi},$$

and the variance can be expressed as

$$\begin{aligned}\mathrm{Var}(P) &= \mathrm{Var}_\pi[E(P|\pi)] + E_\pi[\mathrm{Var}(P|\pi)] \\ &= \mathrm{Var}_\pi[\pi] + E_\pi\left[\frac{\pi(1-\pi)}{n}\right] \\ &= \sigma^2 + \frac{1}{n}E_\pi[\pi - \pi^2] = \sigma^2 + \frac{1}{n}[\bar{\pi} - (\sigma^2 + \bar{\pi}^2)] \\ &= \frac{\bar{\pi}(1-\bar{\pi})}{n} + \left(1 - \frac{1}{n}\right)\sigma^2, \end{aligned} \tag{3}$$

where we have used that $\sigma^2 = E_\pi[\pi^2] - \bar{\pi}^2$. Importantly, if $n > 1$, it follows that

$$\mathrm{Var}(P) > \frac{\bar{\pi}(1-\bar{\pi})}{n} = \frac{E(P)[1-E(P)]}{n},$$

the variance is larger than the binomial variance in equation 2, and we have overdispersion.

2.2. Dichotomous Responses and the Bernoulli Distribution

When $n = 1$ and Y is dichotomous or binary, $Y \in \{0, 1\}$, the distribution is Bernoulli(π) (or equivalently Binomial[1, π]), and the relationship between the variance and expectation π becomes

$$\mathrm{Var}(P) = \pi(1-\pi) = E(P)[1-E(P)]. \tag{4}$$

When there are omitted covariates leading to randomly varying π, we see from Equation 3 that $\mathrm{Var}(P) = E(P)[1-E(P)]$, as in Equation 4, when $n = 1$ so there is no overdispersion.

It is easily demonstrated that the relationship between the mean and variance is always fixed in samples, so there is no scope for overdispersion or underdispersion. To facilitate generalization to polytomous responses (with more than two categories) in Section 2.3, we let y_i^s be 1 if s is the observed response category for a sample unit i ($i = 1, \ldots, N$) and 0 otherwise ($s = 0, 1$), and we let $\sum_{i=1}^{N} y_i^s = N^s$. The sample mean of y_i^s or sample proportion in category s is

$$\bar{y}^s = \frac{1}{N}\sum_{i=1}^{N} y_i^s = \frac{N^s}{N},$$

and, noting that $y_i^s = (y_i^s)^2$, the sample variance is

$$\mathrm{Var}(y^s) = \frac{1}{N}\sum_{i=1}^{N} y_i^s - (\bar{y}^s)^2 = \frac{N^s}{N} - (\bar{y}^s)^2 = \bar{y}^s - (\bar{y}^s)^2 = \bar{y}^s(1-\bar{y}^s). \tag{5}$$

These relationships between the sample variance and sample mean hold whether we condition on covariates (by considering subgroups of units with the same covariate values) or not. In a regression model for dichotomous responses, the sample variance for given covariate values can therefore deviate from that implied by the predicted mean $\hat{\pi}$ only if the predicted mean differs from the sample mean, for instance because of the omission of an interaction term.

That overdispersion cannot be modeled for dichotomous or binary responses is implicit in much of the literature on categorical data analysis. Unfortunately, it has rather seldom been pointed out explicitly, exceptions being Fahrmeir and Tutz (2001, p. 35) and Collett (2002, pp. 170, 195). Collett also pointed out that the deviance cannot be used for model assessment in the dichotomous case because it only depends on fitted probabilities (Collett, 2002, pp. 68–69, 196; see also McCullagh & Nelder, 1989, p. 121).

2.3. Polytomous Responses

Consider now the case of polytomous responses in which the response of each unit i is one of $S > 2$ mutually exclusive outcome categories. Let y_i^s be 1 if s is the observed response category and 0 otherwise ($s = 0, \ldots, S-1$). Using the results shown for dichotomous responses, the sample mean of y_i^s is $\bar{y}^s$, and the sample variance is $\bar{y}^s(1-\bar{y}^s)$, as shown in Equation 5. The sample covariance of responses for categories s and s' becomes

$$\mathrm{Cov}(y^s, y^{s'}) = \frac{1}{N}\sum_{i=1}^{N} y_i^s y_i^{s'} - \left(\frac{1}{N}\sum_{i=1}^{N} y_i^s\right)\left(\frac{1}{N}\sum_{i=1}^{N} y_i^{s'}\right) = -\bar{y}^s\bar{y}^{s'}, \tag{6}$$

consistent with the relationship for the population under the binomial distribution. No assumptions were made in deriving these relationships, so they hold in all samples (regardless of whether we condition on covariates), and there is thus no scope for modeling overdispersion.

3. Overdispersion in Multilevel Models

3.1. Multilevel and Generalized Linear Mixed Models

In multilevel designs, elementary units are nested in clusters that may be nested in superclusters, and so on. For notational simplicity, we consider two-level designs in which units i ($i = 1, \ldots, n_j$) are nested in clusters j. Unobserved heterogeneity between clusters induces dependence within clusters, even after conditioning on covariates representing observed heterogeneity. For instance,

two students within a school (the cluster) are expected to be more similar than two students from different schools.

Two-level generalized linear mixed models are specified as in Equation 1 but with random effects $\boldsymbol{\zeta}_j$ in the linear predictor η_{ij} for unit i in cluster j,

$$\pi_{ij} = h(\eta_{ij}) = h(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_j),$$

where $\mathbf{z}_{ij}$ are covariates, typically a subset of $\mathbf{x}_{ij}$, and $\boldsymbol{\zeta}_j \sim N(0, \Psi)$.

3.2. Overdispersion in Multilevel Models for Proportions ($n_{ij} > 1$)

If the response for unit i in cluster j is a count of "successes" out of n_{ij} "trials," overdispersion can be modeled by either introducing a Level 1 random effect r_{ij} in the linear predictor (e.g., Breslow & Clayton, 1993),

$$\pi_{ij} = h(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_j + r_{ij}),$$

or by using a quasi-likelihood approach without r_{ij} in the linear predictor but with a modified variance function (e.g., Goldstein, 1991; Schall, 1991),

$$\mathrm{Var}(P_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_j) = \phi \frac{\pi_{ij}(1-\pi_{ij})}{n_{ij}}.$$

These two approaches are straightforward adaptations of common methods for ordinary generalized linear models without random effects.

Note that the random-effects approach does not allow for underdispersion, because variances are nonnegative and it implies additive overdispersion if $n > 1$ (see Equation 3), whereas the quasi-likelihood approach implies multiplicative overdispersion (or underdispersion).

In multilevel models, the most natural approach seems to be to add a random effect r_{ij} at Level 1 because variability at the higher levels is also modeled by random effects. This approach has the advantage over the quasi-likelihood method that it corresponds to a statistical model or data-generating mechanism and can thus be estimated by maximum likelihood (e.g., Rabe-Hesketh, Skrondal, & Pickles, 2005) or using Markov chain Monte Carlo (e.g., Browne, Subramanian, Jones, & Goldstein, 2005). Furthermore, statistical models can be used to make predictions and perform simulations.

As shown in Equation 3, a random π_{ij} produces overdispersion only if $n_{ij} > 1$ and serves no purpose if $n_{ij} = 1$. In fact, if $n_{ij} = 1$, the variance of r_{ij} is not identified from the first- and second-order moments of P, so the redundancy of r_{ij} in this case will be noticed by users through slow convergence or nonconvergence of the estimation algorithms. In contrast, this problem is not flagged in the quasi-likelihood approach, because the overdispersion parameter ϕ is identified in the linearized model.

3.3. Overdispersion in Multilevel Models for Dichotomous Responses ($n_{ij} = 1$)

3.3.1. Data-Generating Model Formulations

An appropriate model for dichotomous responses Y_{ij} is

$$Y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_j \sim \text{Bernoulli}(\pi_{ij}). \tag{7}$$

As for single-level models with dichotomous responses, it does not make sense to include an overdispersion parameter here. This is because the variance will be determined by the expectation via Equation 4 for units in a given cluster (i.e., with given value of $\boldsymbol{\zeta}_j$) sharing the same covariate values. In fact, the same variance-expectation relationship also holds marginally with respect to the random effects (i.e., considering all units in all clusters sharing the same covariate values; e.g., Skrondal & Rabe-Hesketh, 2004, p. 127). Furthermore, Equation 5 also holds for the sample variance, both conditionally and marginally. By specifying the model as in Equation 7 it is evident that there is no scope for overdispersion.

Another data-generating model formulation is via a continuous latent response Y^*_{ij} related to the observed response via

$$Y_{ij} = \begin{cases} 1 & \text{if } Y^*_{ij} > 0 \\ 0 & \text{else} \end{cases}.$$

A linear regression model is specified for the latent response

$$Y^*_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\boldsymbol{\zeta}_j + \varepsilon_{ij},$$

where the residual errors ε_{ij} have mean zero and are independently distributed across units i and clusters j and are independent from the $\boldsymbol{\zeta}_j$. A logistic regression model is obtained if ε_{ij} is logistic (with variance $\pi^2/3$) and a probit model if ε_{ij} is standard normal. If an attempt is made to estimate the variance of ε_{ij}, convergence problems will again occur because the variance is not identified at all in the probit case and only very weakly in the logistic case (Rabe-Hesketh & Skrondal, 2001).

3.3.2. "Algorithmic" Model Formulation

Goldstein (1995, 2003), and many others, expressed the model as

$$P_{ij} = \pi_{ij} + \varepsilon_{ij} z^{(1)}_{ij}, \qquad z^{(1)}_{ij} = \sqrt{\pi_{ij}(1-\pi_{ij})/n_{ij}}. \tag{8}$$

Constraining the Level 1 variance of ε_{ij} to 1, $\mathrm{Var}(\varepsilon_{ij}) = 1$, we obtain the required binomial variance $\pi_{ij}(1-\pi_{ij})/n_{ij}$. Overdispersion or extrabinomial variation can be introduced into the model by estimating $\mathrm{Var}(\varepsilon_{ij}) = \phi$ instead of fixing it to 1.

The reason for writing a random-effects binomial regression model in this way is that some estimation methods, such as marginal quasi-likelihood and penalized quasi-likelihood (PQL), rely on a clever linearization of the model (e.g., Goldstein, 1991) by expanding $h(\eta_{ij})$ as a Taylor series. By including an additive error term ε_{ij}, the linearized model becomes a linear mixed model, which is easy to estimate.

Although the above "model" correctly produces the expectation and variance of a random-effects binomial regression model, it is clear that it is not a proper model, because only contrived distributions for ε_{ij} can generate valid proportions in the sense that $P_{ij}n_{ij}$ are integers. This problem is perhaps most transparent if an algorithmic formulation is adopted for dichotomous responses Y_{ij},

$$Y_{ij} = \pi_{ij} + \varepsilon_{ij} z_{ij}^{(1)}, \qquad z_{ij}^{(1)} = \sqrt{\pi_{ij}(1-\pi_{ij})},$$

as in Renard and Molenberghs (2002). Because Y_{ij} is either 1 or 0, it follows that the random variable ε_{ij} can only take on the values $(1-\pi_{ij})/\sqrt{\pi_{ij}(1-\pi_{ij})}$ and $-\pi_{ij}/\sqrt{\pi_{ij}(1-\pi_{ij})}$, with probabilities π_{ij} and $1-\pi_{ij}$, respectively.

The basic problem with writing the model as in Equation 8 is that it does not correspond to the data-generating mechanism. The resulting confusion is apparent in many articles describing this model. For instance, $e_{ij} = \varepsilon_{ij} z_{ij}^{(1)}$ is sometimes said to be a binomial random variable (e.g., Ribaudo & Thompson, 2002; Wright, 1997, 1998). More important, by using the algorithmic formulation, the data-generating mechanism is easily forgotten, and one may be tempted to estimate the variance of the "error term" as in ordinary linear regression even if the response is dichotomous. Perhaps for these reasons, there is a widespread use of overdispersion parameters for dichotomous responses both in the applied literature (e.g., Dryler, 1999; McCulloch, 2001; Pinilla, Gonzalez, Barber, & Santana, 2002; Wright, 1998) and in the methodological literature (e.g., Goldstein, 1995, p. 102, 2003, p. 99; Guo & Zhao, 2000; Hox, 2002, pp. 107–108; Jacob, 2001; Luke, 2004, pp. 57–58; McLeod, 2001; Renard & Molenberghs, 2002; Venables & Ripley, 2002; Wolfinger & O'Connell, 1993; Wright, 1997; Yang, Goldstein, & Heath, 2000).

The same problem applies to generalized estimating equations, which is a quasi-likelihood method and thus does not require the specification of a data-generating mechanism. Redundant overdispersion parameters are therefore also common in generalized estimating equations. See Hardin and Hilbe (2003, p. 83) and Molenberghs and Verbeke (2005, sect. 8.9) for examples.

3.3.3. Potential Utility of the Overdispersion Parameter in PQL

It has been argued that ϕ does have a role to play even for dichotomous responses when PQL is used. Yang et al. (2000) interpreted an estimated scale parameter very different from 1 as indicating model misspecification, thus using ϕ as a diagnostic. Venables and Ripley (2002, p. 298) pointed out for a particular application that estimating ϕ instead of setting it to 1 yielded estimates of the other parameters that were closer to the maximum likelihood estimates.

Little research appears to have been done on the use of ϕ as a diagnostic or for improving estimates. Exceptions are Wright (1997) and Jacob (2001), who simulated data from multilevel logistic regression models for dichotomous responses and used PQL-2 to estimate the models, including an overdispersion parameter. They found that the overdispersion parameter is underestimated if there are only small numbers of Level 1 units (fewer than 10 or 5) per Level 2 unit, yielding mean estimates of ϕ below 0.9 for some models even under correct model specification. Similarly, Yang (1997) obtained mean estimates of about 0.9 for ordered and unordered polytomous responses with clusters of size 30 and correctly specified models. These findings cast some doubt on the usefulness of ϕ as a diagnostic, although Jacob suggested that overdispersion parameters larger than 1 could indicate that higher level random effects have been omitted. A similar but more convincing use of ϕ as a diagnostic was proposed by Schall (1991) for $n > 1$, who interpreted $\hat{\phi} > 1$ as suggesting that further random effects should be included in the model.

Unfortunately, the parameter ϕ is typically interpreted as an overdispersion parameter, as if such a phenomenon were possible for dichotomous responses. For instance, Yang et al. (2000) remarked that "the extrabinomial parameter is well below 1, suggesting that the assumption of binomial error for the model is not appropriate" (p. 57). Furthermore, we are not aware of any multilevel textbook or software manual explicitly stating that it is impossible to model overdispersion for dichotomous responses, although the HLM 6 software (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004, p. 138) does not allow for overdispersion in this case.

It should be noted that the potential usefulness of overdispersion parameters discussed here is relevant only for estimation using linearization, such as PQL. Unfortunately, these methods often perform poorly in practice (e.g., Breslow, 2005; Rodriguez & Goldman, 2001).

3.4. Overdispersion in Multilevel Models for Polytomous Responses ($n_{ij} = 1$)

Consider a trial with S possible outcome categories. The proportion of trials resulting in outcome s for unit i in cluster j is denoted P_{ij}^{s}. Goldstein (1995, 2003) specified a random-effects multinomial logit model as

$$\log\left(\frac{\pi_{ij}^{s}}{\pi_{ij}^{S}}\right) = \mathbf{x}_{ij}'\boldsymbol{\beta}^{s} + \mathbf{z}_{ij}'\boldsymbol{\zeta}_{j}^{s}, \qquad s = 1, \ldots, S-1,$$

where $\pi_{ij}^{s} = E[P_{ij}^{s}|\mathbf{x}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\zeta}_{j}^{s}]$.

Goldstein (1995) invoked the standard assumption of a multinomial distribution for the response proportions,

$$\mathrm{Cov}(P_{ij}^{s'}, P_{ij}^{s}|\mathbf{x}_{ij}, \mathbf{z}_{ij}) = \begin{cases} \pi_{ij}^{s}(1-\pi_{ij}^{s})/n_{ij} & (s = s') \\ -\pi_{ij}^{s'}\pi_{ij}^{s}/n_{ij} & (s \neq s') \end{cases}, \tag{9}$$

and used a similar but more complex specification than Equation 8 to produce the required variances and covariances.

Goldstein (1995, p. 105) extended the idea of overdispersion or extrabinomial variation from the binomial case to extramultinomial variation in the multinomial case by allowing all variances and covariances to be multiplied by a common scale factor ϕ. However, if $n_{ij} = 1$, the extramultinomial parameter is redundant because the relationships in Equations 5 and 6 are retained in the sample. Multilevel models for individual unordered categorical responses Y_{ij} with extramultinomial variation have nevertheless been used (e.g., Yang, 2001).

For ordered categorical responses, the appropriate distribution is again multinomial. The ordinal nature of the response only affects the relationship between π_{ij}^{s} and the covariates. It is nevertheless often attempted to model overdispersion for individual ordinal responses Y_{ij}^{s} when $n_{ij} = 1$. Examples of such use of scale parameters in multilevel models include Fielding, Yang, and Goldstein (2003) and Fielding and Yang (2005).

As in the dichotomous case, a possible justification is to "allow for model misspecification" or "improved estimation of the model parameters" (Fielding & Yang, 2005, pp. 172, 173), although the same authors also interpreted the parameter as allowing the "multinomial distribution assumption" to be relaxed (p. 174).

4. Conclusion

In situations in which overdispersion can occur, such as multinomial and binomial distributions with denominators greater than 1 or Poisson distributions, we recommend modeling overdispersion by including a Level 1 random intercept in the linear predictor. When the multinomial or binomial denominators are equal to 1, overdispersion (or underdispersion) is impossible and should not be modeled. Unfortunately, redundant overdispersion parameters are often estimated in this case using quasi-likelihood methods. Even when an overdispersion parameter is included for other reasons, it is important to emphasize that the parameter does not represent any such phenomenon as overdispersion or extrabinomial variation.

Confusion may simply be a price one has to pay for not specifying a statistical model for the data-generating mechanism. Having started this article by

referring to D. R. Cox, it seems fitting to close it with a quotation from Cox (1990): "The essential idea is that if the investigator cannot use the model directly to simulate artificial data, how can 'Nature' have used anything like that method to generate real data?" (p. 172).

References

Breslow, N. E. (2005). Whither PQL? In Lin, D. Y., & Heagerty, P. J. (Eds.), *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data* (pp. 1–22). New York: Springer.

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

Browne, W. J., Subramanian, S. V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society, Series A*, *168*, 599–613.

Collett, D. (2002). *Modelling binary data* (2nd ed.). London: Chapman & Hall/CRC.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, *70*, 269–274.

Cox, D. R. (1990). Role of models in statistical analysis. *Statistical Science*, *5*, 169–174.

Dryler, H. (1999). The impact of school and classroom characteristics on educational choices by boys and girls: A multilevel analysis. *Acta Sociologica*, *42*, 299–318.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer.

Fielding, A., & Yang, M. (2005). Generalized linear mixed models for ordered responses in complex multilevel structures: Effects beneath the school or college in education. *Journal of the Royal Statistical Society, Series A*, *168*, 159–183.

Fielding, A., Yang, M., & Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Statistical Modelling*, *3*, 127–153.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, *78*, 45–51.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Arnold.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Guo, G., & Zhao, H. (2000). Multilevel modeling of binary data. *Annual Review of Sociology*, *26*, 441–462.

Hardin, J., & Hilbe, J. (2003). *Generalized estimating equations*. Boca Raton, FL: Chapman & Hall/CRC.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Jacob, M. (2001). Extra-binomial variation in logistic multilevel models—A simulation. *Multilevel Modelling Newsletter*, *12*, 8–14.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McCulloch, A. (2001). Teenage childbearing in Great Britain and the special concentration of poverty households. *Journal of Epidemiology and Community Health*, *55*, 16–23.

McLeod, A. (2001). Multivariate multilevel regression. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 59–73). Chichester, UK: Wiley.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer.

Pinilla, J., Gonzalez, B., Barber, P., & Santana, Y. (2002). Smoking in young adolescents: An approach with multilevel discrete choice models. *Journal of Epidemiology and Community Health*, *56*, 227–232.

Rabe-Hesketh, S., & Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics*, *57*, 1256–1264.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*, 301–323.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Renard, D., & Molenberghs, G. (2002). Multilevel modeling of complex survey data. In M. Aerts, H. Geys, G. Molenberghs, & L. M. Ryan (Eds.), *Topics in modelling clustered data* (pp. 263–272). Boca Raton, FL: Chapman & Hall/CRC.

Ribaudo, H. J., & Thompson, S. G. (2002). The analysis of repeated multivariate binary quality of life data: A hierarchical model approach. *Statistical Methods in Medical Research*, *11*, 69–83.

Rodriguez, G., & Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case study. *Journal of the Royal Statistical Society, Series A*, *164*, 339–355.

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

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Wolfinger, R. D., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computing and Simulation*, *48*, 233–243.

Wright, D. B. (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical and Statistical Psychology*, *50*, 21–29.

Wright, D. B. (1998). Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology*, *12*, 339–357.

Yang, M. (1997). Multilevel models for multiple category responses—A simulation. *Multilevel Modelling Newsletter*, *9*, 10–16.

Yang, M. (2001). Multinomial regression. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 107–123). Chichester, UK: Wiley.

Yang, M., Goldstein, H., & Heath, A. (2000). Multilevel models for repeated binary outcomes: Attitudes and voting over the electoral cycle. *Journal of the Royal Statistical Society, Series A*, *163*, 49–62.

Authors

ANDERS SKRONDAL is a professor of statistics and chair in social statistics in the Department of Statistics and director of The Methodology Institute, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom;

a.skrondal@lse.ac.uk. He is also a senior scientist, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway. His research interests in applied statistics, psychometrics, and biostatistics include latent variable, multilevel, and longitudinal modeling.

SOPHIA RABE-HESKETH is a professor in the Graduate School of Education and the Graduate Group in Biostatistics at the University of California, Berkeley, 4329 Tolman Hall, Berkeley, CA 94720; sophiarh@berkeley.edu. She is also chair of social statistics at the Institute of Education, University of London, London, United Kingdom. Her research interests include multilevel and latent variable modeling and software development.

Manuscript received August 17, 2005
Accepted June 19, 2006