A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil

Author(s): Narayan Sastry

Source: *Journal of the American Statistical Association*, Vol. 92, No. 438 (Jun., 1997), pp. 426-435
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2965690
Accessed: 19-01-2018 15:48 UTC



# A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil

Narayan SASTRY

This article presents a multivariate hazard model for survival data that are clustered at two hierarchical levels. The model provides corrected parameter estimates and standard errors, as well as estimates of the intragroup correlation at both levels. The model is estimated using the expectation–maximization (EM) algorithm. We apply the model to an analysis of the covariates of child survival using survey data from northeast Brazil collected via a hierarchically clustered sampling scheme. We find that family and community frailty effects are fairly small in magnitude but are of importance because they alter the results in a systematic pattern.

KEY WORDS: EM algorithm; Hierarchically clustered data; Multilevel hazard model; Nested random effects; Unobserved heterogeneity.

## 1. INTRODUCTION

This article presents a multivariate hazard model for survival data that are clustered at two hierarchical levels. The model provides corrected parameter estimates and standard errors—as well as estimates of intragroup correlation of survival times at both levels—with survey data collected via a hierarchically clustered sampling scheme. It is also applicable to problems in which data are naturally clustered at two levels, independently from the study design. The model accounts for the hierarchical clustering of the data by including two nested random frailty effects. The two random effects are assumed to be mutually independent and to each follow the gamma distribution. The parameters of the hazard model and the mixing distributions are estimated using the expectation–maximization (EM) algorithm. The incomplete-data log-likelihood function is used to calculate standard errors.

Frailty models have been applied to the analysis of event-history data in a number of research areas, including the study of unemployment durations (McCall 1994), consumer purchase behavior (Gönül and Srinivasan 1993), spells on welfare (Blank 1989), migration (Lindstrom 1996), fertility (Larsen and Vaupel 1993), and marriage and divorce (Lillard, Brien, and Waite 1995). Although (nested) multilevel random-effects models are often appropriate in these types of studies (Goldstein 1995), multilevel random-effects models for survival analysis have been proposed only recently (Rodríguez 1994; Sargent 1996).

In this article we apply the nested frailty model to an analysis of the covariates of child survival in northeast Brazil. Most demographic surveys in the developing world—including the World Fertility Survey and the Demographic and Health Surveys—collect child survival data that are clustered at the community level and at the family level due to sampling design. Independence among observations is a standard but important assumption of regression analysis. However, clustered child survival data are correlated at the community level and at the family level. Because ignoring the full dependence among observations may lead to standard errors that are understated and parameter estimates that are both biased and inconsistent when estimating nonlinear models, results obtained from applying our model to these datasets should be of both methodological and policy interest.

Clustered survival data also provide us with an opportunity to estimate the association among children belonging to the same family or residing in the same community that persists even after controlling for observed covariates. This association is not a result of sample design—rather, it is a consequence of unmeasured or unmeasurable genetic, behavioral, and environmental factors that are related to mortality risks and are common to siblings or to children residing in the same community. We exploit the fact that children in the same cluster share the same unmeasured factors at the family or the community level to estimate parameters that describe the distribution of the unobservables in the population.

Research on the statistical analysis of survival data from related individuals began in the mid-1970s with papers by Clayton (1978) and Holt and Prentice (1974). Subsequent work by Clayton and Cuzick (1985), Cox and Oakes (1984), Hougaard (1986), and Oakes (1982) was reviewed by Hougaard (1987). This stream of research is closely related to Vaupel's (1989, 1990) work on the analysis of frailty effects in demographic models of kindred or related lifetimes. Cox and Oakes (1984) and Kalbfleish and Prentice (1980) provided details introductions to standard hazard models, developed initially by Cox (1972). Hazard models have the important advantage over other regression methods of being able to accommodate censored observations and time-varying covariates in analyses of duration or event-history data.

A number of works on the statistical analysis of correlated mortality risks have appeared recently in the de-

Narayan Sastry is Associate Social Scientist, RAND Corporation, Santa Monica, CA 90407. The author gratefully acknowledges the guidance and advice of Germán Rodríguez and thanks Noreen Goldman and Lorenzo Moreno for useful comments and suggestions and Dean Hyslop for generous help with computing.


Journal of the American Statistical Association
June 1997, Vol. 92, No. 438, Applications and Case Studies

*Table 1. Summary Statistics for Variables Used in Hazard Model Analysis: Births in Northeast Brazil 1976–1986*

| Variable | Mean or percent in category |
|---|---|
| Sex | |
| Female | 49.42% |
| Male | 50.58 |
| Maternal age (years) | |
| Mean | 26.02 |
| Standard deviation | 6.01 |
| Birth order/preceding birth interval[a] | |
| First birth | 21.11% |
| Order 2–4/short | 9.40 |
| Order 2–4/medium | 20.20 |
| Order 2–4/long | 13.27 |
| Order 5+/short | 9.00 |
| Order 5+/medium | 16.84 |
| Order 5+/long | 10.18 |
| Duration of breastfeeding (at age 1–5 months)[b] | |
| <1 month | 8.42% |
| Following birth interval | |
| <12 months | 7.09% |
| Previous child dead[c] | 9.67% |
| Number of births | 2,946 |
| Number of deaths | 430 |

[a] Previous birth interval lengths: short, <15 months; medium, 15–29 months; long, >30 months.
[b] For children weaned before death and born in the 5 years preceding the survey.
[c] Previous child dead prior to conception of reference child.

mographic literature. Guo (1993) and Guo and Rodríguez (1992) presented a multivariate proportional hazards model with a single random effect and discussed estimation of their model using the EM algorithm. Curtis, Diamond, and McDonald (1993) and Zenger (1993) have applied random-effects logistic models to the study of family effects on child mortality. These studies have been restricted to the analysis of clustering at the family level and have not considered the correlation among children living in the same community. The magnitude and importance of clustering at the community level is not known, and there has been no previous work on community-level clustering or on multilevel survival models (Rodríguez 1994).

This article is organized into seven sections. In Section 2 we describe the data, and in Section 3 we present the nested frailty model for hierarchically clustered survival data. In Sections 4 and 5 we discuss estimation of the model and calculation of standard errors. We present our results in Section 6 and a discussion and conclusions in Section 7.

## 2. DATA

The child survival data analyzed in this study come from the *Pesquisa Nacional sobre Saúde Materno-Infantil e Planejamento Familiar—Brasil* (Arruda, Rutenberg, Morris, and Ferraz 1987), a household survey of Brazil conducted as part of the Demographic and Health Survey program. Retrospective maternity histories were collected from 5,892 women age 15–44 over a 3-month period in mid-1986. The survey was based on a multistage, clustered sampling scheme.

A total of 12,356 births were recorded in the survey. Of these, 2,946 were singleton births that occurred within 10 years of the survey and in the northeast region; it is the mortality experience of these children that we analyze in this article. The northeast is the high-mortality region of Brazil and is poorer and less-developed than other parts of the country—particularly the south/southeast region, which includes highly industrialized areas in the states of São Paulo and Rio de Janeiro. For the period 1976–1986, the infant mortality rate was 143.7 in the northeast, compared to 53.0 in the rest of Brazil, and the child mortality rate was 164.4 in the northeast and 58.6 in the rest of Brazil. Relationships between covariates and child mortality risks are substantially different in the two regions (Sastry, Goldman, and Moreno 1993). This, and the difference in mortality rates between the two regions, lead us to the decision to stratify the sample and focus on the northeast.

Covariates in our analysis consist of the child's age and sex, birth order, birth spacing, the survival status of the preceding child, breastfeeding status, and maternal age (i.e., mother's age at the birth of the child). This is a standard set of proximate determinants of child mortality that are typical of the covariates that have appeared in previous demographic studies of child mortality (Hobcraft, McDonald, and Rutstein 1985; Miller, Trussell, Pebley, and Vaughan 1992; Palloni and Millman 1986; Pebley and Stupp 1987). Summary statistics for these variables are presented in Table 1. (See Sastry 1995 for further details about the sample and variable construction.)

The 2,946 children in the sample come from 1,051 families, which are distributed in turn among 101 sampling clusters. The sampling clusters are census tracts that should correspond fairly closely to neighborhoods or villages. These units provide our measure of a community.

The distribution of children by family is presented in Table 2. More than 90% of the children belong to families that contribute two or more children to the sample; the nearly 18% of families that contribute five or more children together account for almost 40% of all children. The mean number of children per family is 2.8. Of the 430 deaths in the sample, 57% come from the 9% of families with more than one death. Approximately 3.5% of the families in the sample contribute three or more deaths; together, these families account for 29% of the total number of deaths.

The distribution of children by community is shown in Table 3. The number of children per community ranges from 4 to 77. The proportion of children who have died increases with the size of the community. The mean number

*Table 2. Distribution of Children by Family*

| Children per family | Deaths per family | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 255 | 12 | | | | | | 267 |
| 2 | 239 | 44 | 2 | | | | | 285 |
| 3 | 143 | 41 | 15 | 3 | | | | 202 |
| 4 | 69 | 30 | 9 | 2 | 0 | | | 110 |
| 5 | 43 | 34 | 15 | 9 | 3 | 0 | | 104 |
| 6 | 15 | 18 | 8 | 5 | 3 | 0 | 1 | 50 |
| 7 | 4 | 4 | 7 | 4 | 2 | 0 | 0 | 21 |
| 8 | 1 | 2 | 4 | 3 | 1 | 1 | 0 | 12 |
| Total | 769 | 185 | 60 | 26 | 9 | 1 | 1 | 1,051 |

Table 3. Distribution of Children by Community

| Children per community | Deaths per community | | | | | | | | Total | Percent of | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6–9 | 10–20 | | Children | Deaths |
| 4–17 | 8 | 3 | 2 | 2 | 2 | 1 | 0 | 0 | 18 | 7.9 | 6.0 |
| 18–24 | 5 | 6 | 3 | 3 | 2 | 1 | 2 | 0 | 22 | 15.7 | 10.9 |
| 25–29 | 0 | 6 | 1 | 5 | 2 | 3 | 1 | 1 | 19 | 17.3 | 14.7 |
| 30–39 | 0 | 1 | 4 | 3 | 0 | 4 | 6 | 3 | 21 | 24.3 | 27.0 |
| 40–77 | 0 | 1 | 1 | 1 | 1 | 3 | 6 | 8 | 21 | 34.7 | 41.4 |
| Total | 13 | 17 | 11 | 14 | 7 | 12 | 15 | 12 | 101 | 100.0 | 100.0 |
| % children | 6.7 | 13.9 | 10.0 | 12.7 | 6.0 | 12.9 | 20.0 | 17.8 | 100.0 | | |
| % deaths | 0 | 3.4 | 4.9 | 7.3 | 7.0 | 15.5 | 26.8 | 35.1 | 100.0 | | |

of children per community is 29.2, and the mean number of deaths is 4.3. Just under one-third of the communities contribute 6 or more deaths each, and these communities together account for almost two-thirds of all deaths. Clustering of births and deaths by family and by community thus appears to be of some importance in northeast Brazil. Our analysis based on the nested frailty model provides an indication of the importance of this clustering *after* controlling for observed covariates.

## 3. THE MODEL

The survival times for a hierarchically clustered sample of individuals are assumed to be conditional on two independent, cluster-specific random effects, $v_i$ and $w_{ij}$. We assign one $v_i$ random effect to each of the $i = 1, \ldots, I$ clusters and one $w_{ij}$ random effect to each of the $j = 1, \ldots, J_i$ subclusters of cluster $v_i$. Let $\mathbf{t}'_{ij} = (t_{ij1}, t_{ij2}, \ldots, t_{ijK_{ij}})$ represent the survival times for the $K_{ij}$ individuals in the $i-j$th cluster and let $\mathbf{x}_{ijk}$ represent a vector of covariates for the $k$th member of this cluster.

If the random effects are assumed to operate multiplicatively on the baseline hazard, then they are interpreted as relative risks, and the model is written as

$$\begin{aligned} h_{ijk}(t_{ijk}|v_i, w_{ij}) &= v_i w_{ij} \lambda_{ijk}(t_{ijk}) \\ &= v_i w_{ij} \lambda_0(t_{ijk}) \exp(\boldsymbol{\beta}' \mathbf{x}_{ijk}), \end{aligned}$$

where $\lambda_0(t_{ijk})$ represents the baseline hazard, which we assume to be piecewise constant, and $\exp(\boldsymbol{\beta}' \mathbf{x}_{ijk})$ is the relative risk associated with covariates $\mathbf{x}_{ijk}$.

Individuals who have died ($\delta_{ijk} = 1$) contribute to the conditional likelihood function the product of the conditional hazard and the conditional survival functions, whereas individuals whose survival times are censored ($\delta_{ijk} = 0$) contribute the conditional survival function

$$f_{ijk}(t_{ijk}, \delta_{ijk}|v_i, w_{ij}) = [v_i w_{ij} \lambda_{ijk}(t_{ijk})]^{\delta_{ijk}} S_{ijk}(t_{ijk})^{v_i w_{ij}}. \tag{1}$$

Here $S_{ijk}(t_{ijk}) = \exp[-\Lambda_{ijk}(t_{ijk})]$, and $\Lambda_{ijk}(t_{ijk})$ is the integrated hazard corresponding to $\lambda_{ijk}(t_{ijk})$. We omit the time arguments when referring to the hazard and the integrated hazard and denote these quantities simply as $\lambda_{ijk}$ and $\Lambda_{ijk}$. We can write the $i$th cluster's contribution to the conditional likelihood function as

$$\begin{aligned} &L_i(\alpha, \eta, \boldsymbol{\theta}; v_i, w_{i1}, \ldots, w_{iJ_i}) \\ &\quad = f_v(v_i; \alpha) \prod_{j=1}^{J_i} f_w(w_{ij}; \eta) \\ &\qquad \times \prod_{k=1}^{K_{ij}} f_{ijk}(t_{ijk}, \delta_{ijk}|v_i, w_{ij}; \boldsymbol{\theta}). \end{aligned} \tag{2}$$

Once we condition on the unobserved community- and family-level random or *frailty* effects, $v_i$ and $w_{ij}$, the individual observations in the sample are independent. We assume that $v_i$ and $w_{ij}$ are mutually independent, are gamma distributed with variances $1/\alpha$ and $1/\eta$, and, for identifiability, have means equal to 1 at birth (i.e., at duration $\mathbf{t}_{ijk} = 0$). The assumption that the random effects are gamma distributed follows previous research on unobserved heterogeneity that makes wide use of this distribution (Clayton 1978; Oakes 1982; Vaupel, Manton, and Stallard 1979). The advantages of the gamma distribution are its flexible shape and its analytical tractability.

## 4. PARAMETER ESTIMATION

One method for obtaining maximum likelihood estimates of $\alpha$, $\eta$, and $\boldsymbol{\theta}$ when the frailties $v_i$ and $w_{ij}$ are unobserved is to apply the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977). The structure of the EM maximization problem in this application permits us to implement a straightforward acceleration step that overcomes the most important drawback of the EM algorithm: its slow convergence speed.

The EM algorithm requires $Q(\alpha, \eta, \boldsymbol{\theta})$, which is the expectation of the log of Equation (2) conditional on the incomplete data $\mathbf{t}$ and $\boldsymbol{\delta}$ and the current parameter estimates:

$$\begin{aligned} &Q(\alpha, \eta, \boldsymbol{\theta}) \\ &\quad = \mathrm{E}\left[\log \prod_{i=1}^{I} L_i(\alpha, \eta, \boldsymbol{\theta}; v_i, w_{i1}, \ldots, w_{iJ_i})\right] \\ &\quad = \mathrm{E}\left[\sum_{i=1}^{I} \log f_v(v_i) + \sum_{i=1}^{I} \sum_{j=1}^{J_i} \log f_w(w_{ij}) \right. \\ &\qquad \left. + \sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \log f_{ijk}(t_{ijk}, \delta_{ijk}|v_i, w_{ij})\right]. \end{aligned} \tag{3}$$

At each iteration, $Q$ is evaluated at the current parameter estimates and then maximized with respect to the model parameters $(\alpha, \eta, \boldsymbol{\theta})$. Usually, dividing the algorithm into separate E and M steps is impractical (Beale 1977; Dempster et al. 1977). Because the maximization problem for this model breaks into three separate parts (Guo and Rodríguez 1992), the E and M steps can be divided into a set of smaller steps and reordered.

To obtain $Q$ at each E step, we require

$$\mathrm{E}[v_i|\mathbf{t}_i, \boldsymbol{\delta}_i, \alpha, \eta, \boldsymbol{\theta}] = \hat{\mu}_i,$$

$$\mathrm{E}[\log v_i|\mathbf{t}_i, \boldsymbol{\delta}_i, \alpha, \eta, \boldsymbol{\theta}] = \hat{\omega}_i,$$

$$\mathrm{E}[w_{ij}|t_j, \delta_j, \alpha, \eta, \boldsymbol{\theta}] = \hat{\gamma}_{ij},$$

$$\mathrm{E}[\log w_{ij}|t_j, \delta_j, \alpha, \eta, \boldsymbol{\theta}] = \hat{\pi}_{ij},$$

and

$$\mathrm{E}[v_i w_{ij}|\mathbf{t}_i, \boldsymbol{\delta}_i, \alpha, \eta, \boldsymbol{\theta}] = \hat{\varphi}_{ij}. \tag{4}$$

These conditional expectations are derived in the Appendix. The result of the E step, obtained by substituting the conditional expectations (4) into the log-likelihood function (3), is

$$\begin{aligned} Q(\alpha, \eta, \boldsymbol{\theta}) &= \sum_{i=1}^{I} [\alpha \log \alpha + (\alpha - 1)\hat{\omega}_i - \alpha\hat{\mu}_i - \log \Gamma(\alpha)] \\ &\quad + \sum_{i=1}^{I} \sum_{j=1}^{J_i} [\eta \log \eta + (\eta - 1)\hat{\pi}_{ij} - \eta\hat{\gamma}_{ij} - \log \Gamma(\eta)] \\ &\quad + \sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} [\delta_{ijk} \log \lambda_{ijk} - \hat{\varphi}_{ij}\Lambda_{ijk}] \\ &= Q_\alpha(\alpha) + Q_\eta(\eta) + Q_\theta(\boldsymbol{\theta}) + \text{const}, \end{aligned} \tag{5}$$

where const is a constant that depends on the data but not on the unknown parameter values. The maximization problem splits neatly into three parts. The first two parts—$Q_\alpha(\alpha)$ and $Q_\eta(\eta)$—are one-dimensional maximizations that are straightforward to solve using a Newton–Raphson procedure. The final part—$Q_\theta(\boldsymbol{\theta})$—involves maximization with respect to $\boldsymbol{\theta}$ alone. It turns out that this expression differs by a constant term from the log-likelihood function for a standard hazard model if we treat $\hat{\varphi}_{ij}$, the conditional expectation of the product of $v_i$ and $w_{ij}$, as fixed, and thus as a known extra relative risk (Guo and Rodríguez 1992). This allows us to use standard statistical routines to maximize $Q_\theta(\boldsymbol{\theta})$. Because our baseline hazard is assumed to be piecewise exponential, we introduce $\log \hat{\varphi}_{ij}$ as a fixed offset in the associated Poisson regression problem (Holford 1980; Laird and Olivier 1981).

The three parts of the maximization procedure are performed sequentially, and the entire process is repeated to convergence. However, Equation (5) shows that the three parts of $Q$ differ in the degree of information loss due to incomplete data. The only terms that appear in $Q_\alpha(\alpha)$ and $Q_\eta(\eta)$ besides the unknown parameters and missing conditional expectations are the total number of clusters $(I)$ and subclusters $(J = \sum_{i=1}^{I} J_i)$. In contrast, the information loss due to incomplete data in $Q_\theta(\boldsymbol{\theta})$ is small. The EM algorithm converges rapidly when information loss is small (Dempster et al. 1977), suggesting that more iterations are necessary for $Q_\alpha(\alpha)$ and $Q_\eta(\eta)$.

Thus our approach is to mix the E and M operations in the maximization of $Q_\alpha(\alpha)$ and $Q_\eta(\eta)$ to maximize the overall speed and efficiency of the algorithm. This increases the number of iterations required for maximizing $Q_\alpha(\alpha)$ and $Q_\eta(\eta)$, which are simple to perform, while reducing the number of iterations required for maximizing $Q_\theta(\boldsymbol{\theta})$, which can be difficult. In particular, after obtaining new estimates of $\alpha$ and $\eta$, we reevaluate the conditional expectations $\hat{\mu}_i$ and $\hat{\omega}_i$ using the new estimates of $\alpha$ and $\eta$, and the conditional expectations $\hat{\gamma}_{ij}$ and $\hat{\pi}_{ij}$ using the new estimate of $\eta$. We then repeat the maximization procedures for $Q_\alpha(\alpha)$ and $Q_\eta(\eta)$ using the new conditional expectations. The process continues until each of the two sequences of estimates for $\alpha$ and $\eta$ converge, at which time the next cycle through the EM algorithm begins.

This approach essentially applies the EM algorithm independently to each of these two components of $Q$ (at every global EM Iteration). As Dempster et al. (1977) noted, the EM algorithm applied to factors of the likelihood separately should result in more rapid convergence. Our experience using this approach indicates that in the present application it provides an enormous improvement over the standard EM algorithm in terms of speed of convergence. We implemented estimation of the nested frailty model with the EM algorithm using routines written in the statistical language S-Plus (Becker, Chambers, and Wilks 1988; Chambers and Hastie 1992) and in Fortran, and we tested the programs using simulated data. In light of the rapid rate of convergence we obtained, and the analytical and computational complexity associated with acceleration routines based on the results of Louis (1982), Meilijson (1989), and others, we did not incorporate an acceleration procedure into the estimation routine.

## 5. STANDARD ERROR ESTIMATION

The remaining problem is to calculate the standard errors of the parameter estimates. The EM algorithm does not provide correct standard errors; those based on the information matrix obtained from the M step do not take the missing data into account. The standard errors that we report are based on an estimate of the asymptotic covariance matrix constructed from the first derivatives of the incomplete-data log-likelihood function.

A potential problem with this approach, pointed out by a referee, is that with a piecewise baseline hazard, the standard errors based on the score may not be consistent as the number of intervals increase (although the parameter estimates are consistent). We investigate the sensitivity of the standard errors to changes in the number and length of in-

*Table 4. Results for Standard, Family Effects, Community Effects, and Multilevel Hazards Models (With z Statistics in Parentheses)*

| *Variable* | *Model I* | | *Model II* | | *Model III* | | *Model IV* | |
|---|---|---|---|---|---|---|---|---|
| Age (Months) | | | | | | | | |
| 0 | 2.069 | (2.27) | 2.620 | (2.58) | 1.881 | (1.99) | 2.051 | (2.08) |
| 1–5 | .825 | (.91) | 1.410 | (1.39) | .651 | (.69) | .829 | (.87) |
| 6–11 | .164 | (.18) | .789 | (.77) | .007 | (.01) | .194 | (.20) |
| 12–23 | −1.768 | (1.90) | −1.120 | (1.08) | −1.916 | (1.99) | −1.724 | (1.70) |
| 24–59 | −2.953 | (3.13) | −2.298 | (2.19) | −3.102 | (3.19) | −2.907 | (3.14) |
| Sex | | | | | | | | |
| Female[a] | | | | | | | | |
| Male | .249 | (2.56) | .244 | (2.40) | .249 | (2.53) | .248 | (2.26) |
| Birth order/preceding interval | | | | | | | | |
| First | −.417 | (2.33) | −.488 | (2.63) | −.410 | (2.27) | −.429 | (2.14) |
| 2–4/short | .640 | (3.85) | .582 | (3.32) | .584 | (3.46) | .577 | (3.24) |
| 2–4/medium[a] | | | | | | | | |
| 2–4/long | −.612 | (2.71) | −.611 | (2.64) | −.594 | (2.61) | −.596 | (2.38) |
| 5+/short | .916 | (5.27) | .843 | (4.47) | .825 | (4.58) | .816 | (3.82) |
| 5+/medium | .467 | (2.77) | .476 | (2.65) | .389 | (2.23) | .402 | (2.12) |
| 5+/long | −.369 | (1.53) | −.355 | (1.41) | −.403 | (1.64) | −.396 | (1.43) |
| Breastfeeding duration (at age >1 month) | | | | | | | | |
| ≥1 month[a] | | | | | | | | |
| <1 month | .656 | (3.73) | .600 | (3.21) | .621 | (3.46) | .611 | (3.16) |
| Succeeding birth interval | | | | | | | | |
| ≥12 months[a] | | | | | | | | |
| <12 months | .898 | (2.04) | .800 | (1.80) | .815 | (1.85) | .800 | (1.65) |
| Previous child dead | | | | | | | | |
| No[a] | | | | | | | | |
| Yes | .463 | (3.36) | .062 | (.36) | .312 | (2.18) | .216 | (1.45) |
| Maternal age | | | | | | | | |
| Linear effect | −.225 | (3.32) | −.260 | (3.47) | −.209 | (2.99) | −.219 | (3.10) |
| Squared effect | .004 | (3.17) | .004 | (3.31) | .004 | (2.86) | .004 | (2.93) |
| Variance of random effects[b] | | | | | | | | |
| Family | | | .529 | (2.92) | | | .115 | (.96) |
| Community | | | | | .200 | (2.84) | .170 | (1.98) |

[a] Omitted category.
[b] The *z* test for the variance of the random effects is equivalent to a one-sided test with a critical value of 1.645 at the 5% significance level.

tervals in the baseline hazard; we report the results from this analysis in the next section. Note that this type of sensitivity analysis is usually an integral part of any hazard model analysis.

## 6. RESULTS

We estimate four different model specifications; the results are presented in Table 4. Throughout our discussion, results described as "statistically significant" are significant at the 5% level. Model I in Table 4 is a standard hazard model with no correction for clustering. Models II and III include a single frailty effect to allow for clustering by family and by community. Finally, Model IV is based on our nested frailty model that allows for clustering by both family and community. Model I passes a standard $\chi^2$ goodness-of-fit test, and Models II–IV each offer significant improvements over Model I.

The parameters of the gamma distribution estimated in Models II–IV describe the shape of the frailty distribution and can be interpreted in several different ways. The most direct interpretation is simply as the variance of the distribution of the random effects. If the variance is 0, then observations from the same group are independent. A larger variance implies greater heterogeneity in frailty across groups and greater correlation among individuals belonging to the same group. A second way to describe the results is to construct a risk ratio that compares the conditional expectation of the frailty effect for a high-risk group to that for a low-risk group, where for concreteness, high-risk refers to a cluster at the 90th percentile and low risk to a cluster at the 10th percentile of the frailty distribution. A risk ratio of 1 indicates that observations are independent. A significance test for the variance of the random-effects distribution occurs on the boundary of the parameter space; the necessary modification to the usual $z$ test simplifies to a comparison of the test statistic with a critical value from a one-sided test (Self and Liang 1987).

We undertook a sensitivity analysis of the nested frailty model (Model IV) to assess whether standard errors and inferences changed as the number and length of intervals in the baseline hazard were altered. We found that the standard errors were not sensitive to moderate increases in the number of intervals or to changes in their lengths. We estimated models with up to 10 intervals and found that although standard errors tended to increase slightly with the number of intervals, the majority of the $z$ statistics changed only at the second decimal place, if at all. The largest change in a $z$ statistic was a decline of one-tenth of point; however, neither in this nor in any other case did the change come close to altering our inferences. The sensitivity analysis confirmed that a baseline hazard with five intervals was sufficient to provide a good fit to the data.

In Model I nearly all of the covariates are statistically significant, and the estimated effects for the covariates are in close agreement with findings from previous research. The hazard rate declines steadily with child age and is lower for girls. Childbearing at younger and older maternal ages is associated with higher mortality. High parity births and short interbirth intervals are associated with lower survival chances, whereas births of order 2–4 spaced far apart have the lowest mortality risk. The combined effects of high birth order and a short preceding birth interval result in a relative risk of death more than two-and-one-half times higher than the baseline case of medium birth order and spacing. Short breastfeeding durations are associated with significantly higher mortality risks for children age 1–6 months in northeast Brazil. Short following birth intervals (under 12 months) are associated with very high mortality risks. The death of the previous child also has a large and significant effect on a child's survival chances.

The results for Model II indicate that unobserved family effects have a sizeable impact on child mortality risks. The variance of the random effect in this model is .53 and the risk ratio is 7.6; that is, the effect of unobserved family factors for high-risk families, defined as families in the 90th percentile, is more than seven times as large as the effect for low-risk families, defined as families in the 10th percentile.

The effects of breastfeeding, following birth interval, and the preceding child's survival status are all reduced in Model II. Most notable is the change in the effect of the preceding child's death. In the standard hazard model the relative risk is 60% higher for a child whose preceding sibling died and is highly significant. After controlling for unobserved family effects, the relative risk is 1. No clear pattern emerges among the other covariates in this model.

The variance of the community random effect in Model III is .20 and is statistically significant. The risk ratio of high-risk communities to low-risk communities is 2.9. The pattern of changes in parameter estimates in the community frailty model is quite similar to that found in the family frailty model. However, the magnitude of the changes is generally much smaller.

The standard errors are consistently larger when we account for the clustering among observations, and corresponding values are in all cases larger in the family frailty model than in the community frailty model. This provides another indication that family-level clustering is of greater magnitude and importance than community-level clustering in this sample. Of course, the changes in the $z$ statistics depend on the size and direction of the changes in parameter estimates as well as on the increases in the standard errors. It turns out that the $z$ statistics for several of the covariates are actually larger in the random-effects models than in the standard model, although in general they are smaller.

Adding a community random effect to the family random-effects model reduces by almost 80% our previous estimate of the variance of the family effect (see Model IV). The variance of the family random effect is now .12, with a risk ratio of 2.6. The net effect of unmeasured community factors is also smaller than the gross effect, but by a much more modest 15%. The community random effect in Model IV has a variance of .17 and a risk ratio of 2.7. The community effect is statistically significant ($p$ value $= .024$), although the family effect is not ($p$ value $= .169$).

The pattern of changes in parameter estimates found in the family and community frailty models persists in the nested frailty model. Compared to Model I, the effects of breastfeeding, following birth interval, and the previous child's survival status are all substantially smaller. Parameter estimates generally lie in between the values found in Models II and III. The effect of the preceding child's survival status is insignificant, although it is substantially larger than in the model that controls only for family frailty. A final pattern to emerge is that the effects of short and medium preceding birth intervals are overstated in the standard model.

Several of the covariates are insignificant in the final model. Recall that in the original model all but one of the covariates were statistically significant. In particular, the covariate effects for succeeding birth interval and survival status of the preceding child become insignificant in Model IV. For each of the other covariates, in almost every case the $z$ statistic in the nested frailty model is the smallest of the four. This result clearly supports the notion that a sample of correlated observations contains less information than an independent sample.

## 7. DISCUSSION

### 7.1 Family and Community Frailty Effects

The clustering of children by family and by community in demographic survey data provides researchers with an opportunity to estimate measures of familial and community association. However, in models that do not control for the hierarchical nature of clustering (Models II and III), estimates of the variance of family frailty and community frailty tend to be overstated. The variance of the family random effect is biased upward, because it incorporates the effects of environmental factors that are common not only to all children in the same family, but also to all children residing in the same community. The variance of the community random effect is biased upward, because it includes omitted family-level frailty effects common to groups of siblings in each community.

Our results for the northeast region of Brazil indicate that the variance of the family random effect is overstated by a factor of *four-and-one-half* (.529/.115) when community effects are ignored. This variance is not statistically significant in the nested frailty model, although it is highly significant in the model that includes only the family frailty effect. This result suggests that estimates of family frailty that do not control for observed or unobserved community effects must be interpreted extremely cautiously. Not only are these effects likely to be greatly overstated, but they also combine family-specific effects together with community effects and thus are difficult to interpret.

The correlation among sibling survival times has previously been interpreted as the effect of shared genetic factors (Guo 1993) and of parental competence (Das Gupta 1990).

Because we control for community-level heterogeneity, our estimate of the correlation among siblings provides a closer approximation to the true upper bound of the unobserved family effect. Based on our nested frailty model results, the importance of unobserved family effects appears to be quite modest in northeast Brazil: neither parental competence nor shared genetic factors are important in influencing child mortality outcomes in this setting. Further study—especially for countries outside of Latin America, where research to date has focused—is needed to determine the generalizability and robustness of our findings.

The association among children residing in the same community—hypothesized to be the consequence of the shared physical, disease, cultural, and socioeconomic environment—is of modest magnitude in the nested frailty model. The variance of the community random effect is overstated by 18% (.200 vs. .170) in the model that does not account for family clustering. However, the community frailty effect is statistically significant in both models. Unobserved community heterogeneity is important, because no community-level covariates are included in the model and because covariates that are included are correlated strongly among siblings, but only weakly among children residing in the same community.

### 7.2 Parameter Estimates and Standard Errors

Correlation among observations in a clustered sampling scheme can lead to substantial bias in the estimated parameter effects when the variance of the distribution of unobservables is large and the analysis considers survival over an interval in which a large proportion of the initial population dies (Guo and Rodríguez 1992). The extent to which parameter values are affected also depends on a number of other factors, including the distributions assumed for the random effects. To understand the conditions under which our results are likely to be sensitive to the assumption of gamma-distributed frailty, we examine the divergence of the mean hazard from the baseline hazard under the assumptions that the random effect follows three different parametric distributions: gamma, inverse Gaussian, and lognormal (see Vaupel and Yashin 1985). This information is reported in Table 5 by the proportion of the initial population surviving.

*Table 5. Divergence of the Mean Hazard From the Baseline Hazard at Different Levels of Survival and for Different Frailty Distributions*

| S(a) | Gamma | Inverse Gaussian | Log-normal |
|---|---|---|---|
| Variance = .1 | | | |
| 1.00 | 1.000 | 1.000 | 1.000 |
| .90 | 1.011 | 1.011 | 1.011 |
| .75 | 1.029 | 1.029 | 1.029 |
| .50 | 1.072 | 1.069 | 1.069 |
| .10 | 1.259 | 1.230 | 1.230 |
| Variance = .5 | | | |
| 1.00 | 1.000 | 1.000 | 1.000 |
| .90 | 1.054 | 1.053 | 1.053 |
| .75 | 1.155 | 1.144 | 1.140 |
| .50 | 1.414 | 1.346 | 1.334 |
| .10 | 3.163 | 2.151 | 2.145 |
| Variance = 1.0 | | | |
| 1.00 | 1.000 | 1.000 | 1.000 |
| .90 | 1.111 | 1.105 | 1.102 |
| .75 | 1.333 | 1.288 | 1.268 |
| .50 | 2.000 | 1.693 | 1.634 |
| .10 | 10.000 | 3.303 | 3.293 |
| Variance = 2.0 | | | |
| 1.00 | 1.000 | 1.000 | 1.000 |
| .90 | 1.234 | 1.210 | 1.190 |
| .75 | 1.778 | 1.575 | 1.487 |
| .50 | 4.000 | 2.387 | 2.156 |
| .10 | 100.000 | 5.605 | 5.598 |

We find that results are unlikely to be sensitive to the choice of frailty distributions when the proportion of the population surviving the period of analysis is high, *unless* the variance of the frailty distribution is large and the distribution thus highly skewed. Consequently, the effect of distributional assumptions is likely to be relatively minor in the present analysis. At the level of under-age 5 mortality prevailing in our dataset—which is relatively high compared to other countries and regions—the simulation results indicate that the choice of frailty distribution would matter only if the variance was 2 or greater. Results from our study and others to date have found the variances of frailty to be fairly modest—certainly much smaller than the magnitude above which results may be sensitive to distributional assumptions. Our conclusion that the choice of frailty distribution is relatively unimportant is supported by Pickles and Crouchley's (1995) simulation results and by Guo and Rodríguez's (1992) finding that hazard model results were essentially identical when frailty was assumed to be either gamma-distributed or nonparametric in an analysis of child mortality in Guatemala.

The effects of covariates that are indicative of high-risk families and have similar values among siblings tend to be overstated in the standard model that includes no correction for clustering. The coefficients for short preceding birth interval, breastfeeding duration, preceding child's survival status, and succeeding birth interval are all attenuated in the nested frailty model. The most substantial change for the parameter estimates occurs for the preceding child's survival status, which is 53% smaller in the nested frailty model. This covariate clearly plays an important role as a proxy for unmeasured family mortality risk. Note that these changes occur because these covariates reflect the *consequences* of family frailty. Families with high frailty have more deaths, which result, for instance, in shorter interbirth intervals through well-known biological mechanisms. Covariates that are essentially independent among siblings—most notably child sex and maternal age—change the least between the standard model and the nested frailty model.

Our results indicate that standard errors are overstated when we use estimation procedures that assume a simple random sample of observations. On average, the standard errors are 13% larger in the nested frailty model than in the standard model, although the magnitudes of the changes are distinct for different sets of covariates.

The effects of controlling for unobserved family and community heterogeneity on parameter values and standard errors, although fairly modest, appear to be important, because they reveal a systematic pattern in the direction of bias in estimates that assume an independent

sample. Specifically, parameter values are overstated and standard errors are understated in standard models that do not account for the hierarchically clustered sampling scheme. These biases are compounded in $z$ statistics—which can be overstated substantially—resulting in possibly misleading statistical inferences.

## APPENDIX: CONDITIONAL DISTRIBUTION OF RANDOM EFFECTS

Here we derive the joint conditional distribution of $v_i$ and the random effect for the $j'$th subcluster, $w_{ij'}$, which is used to calculate the conditional expectations of $v_i$, $\log v_i$, $w_{ij}$, $\log w_{ij}$, and $v_i w_{ij}$ that appear in the E step of the estimation procedure.

The joint distribution of $v_i$ and $w_{ij'}$, conditional on the data and the current parameter estimates, is

$$
\begin{aligned}
&f(v_i, w_{ij'} | \mathbf{t}_i, \boldsymbol{\delta}_i) \\
&\quad = \frac{f(v_i, w_{ij'}, \mathbf{t}_i, \boldsymbol{\delta}_i)}{f(\mathbf{t}_i, \boldsymbol{\delta}_i)} \\
&\quad = \frac{f(v_i, w_{ij}, \mathbf{t}_i, \boldsymbol{\delta}_i)}{\int_0^\infty \int_0^\infty f(v_i, w_{ij'}, \mathbf{t}_i, \boldsymbol{\delta}_i)\, dw_{ij'}\, dv_i} \\
&\quad = \frac{\int_0^\infty \cdots \int_0^\infty f(v_i, w_{i1}, \ldots, w_{iJ_i}, \mathbf{t}_i, \boldsymbol{\delta}_i) \times dw_{i1} \ldots dw_{ij'-1}\, dw_{ij+1} \ldots dw_{iJ_i}}{\int_0^\infty \int_0^\infty \cdots \int_0^\infty f(v_i, w_{i1}, \ldots, w_{iJ_i}, \mathbf{t}_i, \boldsymbol{\delta}_i) \times dw_{i1} \ldots dw_{iJ_i}\, dv_i}. \qquad \text{(A.1)}
\end{aligned}
$$

The joint distribution $f(v_i, w_{i1}, \ldots, w_{iJ_i}, \mathbf{t}_i, \boldsymbol{\delta}_i)$ is the product of the conditional distribution of survival times, $f(\mathbf{t}_i, \boldsymbol{\delta}_i | v_i, w_{i1}, \ldots, w_{iJ_i})$, with the joint distribution of the random effects, $f(v_i, w_{i1}, \ldots, w_{iJ_i})$. The conditional distribution of survival times for the $i$th cluster is the product of the individual contributions given in Equation (1):

$$
\begin{aligned}
&f(\mathbf{t}_i, \boldsymbol{\delta}_i | v_i, w_{i1}, \ldots, w_{iJ_i}) \\
&\quad = \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} f_{ijk}(t_{ijk}, \delta_{ijk} | v_i, w_{ij}) \\
&\quad = \prod_{j=1}^{J_1} \prod_{k=1}^{K_{ij}} [v_i w_{ij} \lambda_{ijk}]^{\delta_{ijk}} \exp[-v_i w_{ij} \Lambda_{ijk}] \\
&\quad = \left[ \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} \lambda_{ijk}^{\delta_{ijk}} \right] v_i^{\Sigma_j \Sigma_k \delta_{ijk}} \prod_{j=1}^{J_i} w_{ij}^{\Sigma_k \delta_{ijk}} \exp[-v_i w_{ij} \Sigma_k \Lambda_{ijk}].
\end{aligned} \qquad \text{(A.2)}
$$

Because the random effects are assumed to be independent, their joint distribution is simply the product of the marginals,

$$
\begin{aligned}
&f(v_i, w_{i1}, \ldots, w_{iJ_i}) \\
&\quad = f_v(v_i) \prod_{j=1}^{J_i} f_w(w_{ij}) \\
&\quad = \frac{\alpha^\alpha}{\Gamma(\alpha)} v_i^{\alpha-1} \exp(-\alpha v_i) \prod_{j=1}^{J_i} \frac{\eta^\eta}{\Gamma(\eta)} w_{ij}^{\eta-1} \exp(-\eta w_{ij}) \\
&\quad = \frac{\alpha^\alpha}{\Gamma(\alpha)} \left[ \frac{\eta^\eta}{\Gamma(\eta)} \right]^{J_i} v_i^{\alpha-1} \exp(-\alpha v_i) \prod_{j=1}^{J_i} w_{ij}^{\eta-1} \exp(-\eta w_{ij}).
\end{aligned} \qquad \text{(A.3)}
$$

Multiplying (A.2) and (A.3) gives us the full joint distribution of survival times and random effects for the $i$th cluster:

$$
\begin{aligned}
&f(v_i, w_{i1}, \ldots, w_{iJ_i}, \mathbf{t}_i, \boldsymbol{\delta}_i) \\
&\quad = f(\mathbf{t}_i, \boldsymbol{\delta}_i | v_i, w_{i1}, \ldots, w_{iJ_i}) f(v_i, w_{i1}, \ldots, w_{iJ_i}) \\
&\quad = C_i v_i^{\alpha-1+\Sigma_j \Sigma_k \delta_{ijk}} \exp(-\alpha v_i) \\
&\qquad \times \prod_{j=1}^{J_i} w_{ij}^{\eta-1+\Sigma_k \delta_{ijk}} \exp[-\eta w_{ij} - v_i w_{ij} \Sigma_k \Lambda_{ijk}], \qquad \text{(A.4)}
\end{aligned}
$$

where

$$
C_i = \frac{\alpha^\alpha}{\Gamma(\alpha)} \left[ \frac{\eta^\eta}{\Gamma(\eta)} \right]^{J_i} \prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} \lambda_{ijk}^{\delta_{ijk}}.
$$

We obtain $f(v_i, w_{ij'}, \mathbf{t}_i, \boldsymbol{\delta}_i)$, the joint distribution of survival times and the two random effects $v_i$ and $w_{ij'}$ from the full joint distribution in Equation (A.4), by integrating out the random effects for the subclusters that are not of interest; that is, $w_{ij}$, for all $j \neq j'$:

$$
\begin{aligned}
&f(v_i, w_{ij'}, \mathbf{t}_i, \boldsymbol{\delta}_i) \\
&\quad = \int_0^\infty \cdots \int_0^\infty f(v_i, w_{i1}, \ldots, w_{iJ_i}, \mathbf{t}_i, \boldsymbol{\delta}_i) \\
&\qquad \times dw_{i1} \ldots dw_{ij'-1}\, dw_{ij'+1} \ldots dw_{iJ_i} \\
&\quad = C_i v_i^{\alpha-1+\Sigma_j \Sigma_k \delta_{ijk}} \\
&\qquad \times \exp(-\alpha v_i) w_{ij'}^{\eta-1+\Sigma_k \delta_{ij'k}} \exp[-\eta w_{ij'} - v_i w_{ij'} \Sigma_k \Lambda_{ij'k}] \\
&\qquad \times \prod_{j \neq j'} \int_0^\infty w_{ij}^{\eta-1+\Sigma_k \delta ijk} \exp[-\eta w_{ij} - v_i w_{ij} \Sigma_k \Lambda_{ijk}]\, dw_{ij} \\
&\quad = C_i v_i^{\alpha-1+\Sigma_j \Sigma_k \delta_{ijk}} \exp(-\alpha v_i) w_{ij'}^{\eta-1+\Sigma_k \delta_{ij'k}} \\
&\qquad \times \exp[-\eta w_{ij'} - v_i w_{ij'} \Sigma_k \Lambda_{ij'k}] \\
&\qquad \times \prod_{j \neq j'} \Gamma(\eta + \Sigma_k \delta_{ijk}) [\eta + v_i \Sigma_k \Lambda_{ijk}]^{-\eta-\Sigma_k \delta_{ijk}}. \qquad \text{(A.5)}
\end{aligned}
$$

The denominator of the conditional distribution is the joint marginal distribution of survival times, $f(t_i, \delta_i)$, which can be obtained by integrating out $v_i$ and $w_{ij'}$ from Equation (A.5). Using

(A.1) and (A.5) and simplifying, we find the conditional distribution of $v_i$ and $w_{ij'}$ to be

$$f(v_i, w_{ij'}|\mathbf{t}_i, \boldsymbol{\delta}_i)$$

$$= \frac{f(v_i, w_{ij'}, \mathbf{t}_i, \boldsymbol{\delta}_i)}{f(\mathbf{t}_i, \boldsymbol{\delta}_i)} = \frac{f(v_i, w_{ij}, \mathbf{t}_i, \boldsymbol{\delta}_i)}{\int_0^\infty f(v_i, \mathbf{t}_i, \boldsymbol{\delta}_i)\, dv_i}$$

$$= \frac{\begin{array}{c} v_i^{\alpha-1+\Sigma_j\Sigma_k\delta_{ijk}} \exp(-\alpha v_i) w_{ij'}^{-\eta-1+\Sigma_k\delta_{ij'k}} \\ \times \exp[-\eta w_{ij'} - v_i w_{ij'}\Sigma_k\Lambda_{ijk}] \\ \times \prod_{j\neq j'}[\eta + v_i\Sigma_k\Lambda_{ijk}]^{-\eta-\Sigma_k\delta_{ijk}} \end{array}}{\begin{array}{c} \Gamma(\eta + \Sigma_k\boldsymbol{\delta}_{ij'k}) \int_0^\infty v_i^{\alpha-1+\Sigma_j\Sigma_k\delta_{ijk}} \\ \times \exp(-\alpha v_i) \prod_{j=1}^{J_i}[\eta + v_i\Sigma_k\Lambda_{ijk}]^{-\eta-\Sigma_k\delta_{ijk}}\, dv_i \end{array}}.$$

This distribution is used to calculate the conditional expectation of the product of $v_i$ and $w_{ij'}$ (denoted as $\hat{\varphi}_{ij'}$). It is also used to derive the marginal conditional distribution of $v_i$, $f(v_i|t_i, \delta_i)$, and the marginal conditional distribution of $w_{ij'}$, $f(w_{ij'}|\mathbf{t}_i, \boldsymbol{\delta}_i)$. These distributions are used to calculate the conditional expectations of $v_i$ (denoted as $\hat{\mu}_i$) and $\log v_i$ (denoted as $\hat{\omega}_i$) and the conditional expectations of $w_{ij'}$ (denoted as $\hat{\gamma}_{ij'}$) and $\log w_{ij'}$ (denoted as $\hat{\pi}_{ij'}$).

*[Received June 1995. Revised May 1996.]*

## REFERENCES


Arruda, J. M., Rutenberg, N., Morris, L., and Ferraz, E. (1987), *Pesquisa Nacional Sobre Saúde Materno-Infantil e Planejamento Familiar—Brasil, 1986*, Rio de Janeiro: BEMFAM/IRD.

Beale, E. M. L. (1977), Discussion of "Maximum Likelihood Estimation From Incomplete Data via the EM Algorithm," by A. P. Dempster, N. M. Laird, and D. B. Rubin, *Journal of the Royal Statistical Society*, Ser. B, 39, 22–23.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language*, Pacific Grove: Wadsworth & Brooks/Cole.

Blank, R. (1989), "Analyzing the Length of Welfare Spells," *Journal of Public Economics*, 39, 245–273.

Chambers, J. M., and Hastie, T. J. (1992), *Statistical Models in S*, Pacific Grove, CA: Wadsworth & Brooks/Cole.

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Clayton, D. G., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, 34, 187–220.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, New York: Chapman and Hall.

Curtis, S. L., Diamond, I., and McDonald, J. W. (1993), "Birth Interval and Family Effects on Postneonatal Mortality in Brazil." *Demography*, 30, 33–43.

Das Gupta, M. (1990), "Death Clustering, Mothers' Education and the Determinants of Child Mortality in Rural Punjab, India," *Population Studies*, 44, 489–505.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Goldstein, H. (1995), *Multilevel Statistical Models*, New York: Halstead Press.

Gönül, F., and Srinivasan, K. (1993), "Consumer Purchase Behavior in a Frequently Bought Product Category: Estimation Issues and Managerial Insights From a Hazard Function Model With Heterogeneity," *Journal of the American Statistical Association*, 88, 1219–1227.

Guo, G. (1993), "Use of Sibling Data to Estimate Family Mortality Effects in Guatemala," *Demography*, 30, 15–32.

Guo, G., and Rodríguez, G. (1992), "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the EM Algorithm, With an Application to Child Survival in Guatemala," *Journal of the American Statistical Association*, 87, 969–976.

Hobcraft, J. N., McDonald, J. W., and Rutstein, S. O. (1985), "Demographic Determinants of Infant and Early Child Mortality: A Comparative Analysis," *Population Studies*, 39, 363–385.

Holford, T. R. (1980), "The Analysis of Rates and Survivorship Using Log-Linear Models," *Biometrics*, 36, 299–306.

Holt, J. D., and Prentice, R. L. (1974), "Survival Analysis in Twin Studies and Matched Pair Experiments," *Biometrika*, 61, 17–30.

Hougaard, P. (1986), "Survival Models for Heterogenous Populations Derived From Stable Distributions," *Biometrika*, 73, 387–396.

——— (1987), "Modelling Multivariate Survival," *Scandinavian Journal of Statistics*, 14, 291–304.

Kalbfleish, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Laird, N., and Olivier, D. (1981), "Covariate Analysis of Censored Survival Data Using Log-Linear Analysis Techniques," *Journal of the American Statistical Association*, 76, 231–240.

Larsen, U., and Vaupel, J. W. (1993), "Hutterite Fecundability by Age and Parity: Strategies for Frailty Modeling of Event Histories," *Demography*, 30, 81–102.

Lillard, L. A., Brien, M. J., and Waite, L. J. (1995), "Premarital Cohabitation and Subsequent Marital Dissolution: A Matter of Self Selection?" *Demography*, 32, 437–457.

Lindstrom, D. P. (1996), "Economic Opportunities in Mexico and Return Migration to the United States," *Demography*, 33, 357–374.

Louis, T. A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

McCall, B. P. (1994), "Testing the Proportional Hazards Assumption in the Presence of Unmeasured Heterogeneity: An Application to the Unemployment Durations of Displaced Workers," *Journal of Applied Econometrics*, 9, 321–334.

Meilijson, I. (1989), "A Fast Improvement to the EM Algorithm on Its Own Terms," *Journal of the Royal Statistical Society*, Ser. B, 51, 127–138.

Miller, J. E., Trussell, J., Pebley, A. R., and Vaughan, B. (1992), "Birth Spacing and Child Mortality in Bangladesh and the Philippines: How Much is Explained by Gestation?" *Demography*, 29, 305–818.

Oakes, D. (1982), "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society*, Ser. B, 44, 414–422.

Palloni, A., and Millman, S. (1986), "Effects of Inter-Birth Intervals and Breastfeeding on Infant and Early Childhood Mortality," *Population Studies*, 40, 215–236.

Pebley, A. R., and Stupp, P. W. (1987), "Reproductive Patterns and Child Mortality in Guatemala," *Demography*, 24, 43–60.

Pickles, A., and Crouchley, R. (1995), "A Comparison of Frailty Models for Multivariate Survival Data," *Statistics in Medicine*, 14, 1447–1461.

Rodríguez, G. (1994), "Statistical Issues in the Analysis of Reproductive Histories Using Hazard Models," *Annals of the New York Academy of Sciences*, 709, 266–279.

Sargent, D. J. (1996), "A General Framework for Random Effects Survival Analysis in the Cox Proportional Hazards Setting," unpublished manuscript, University of Minnesota, School of Public Health, Division of Biostatistics.

Sastry, N. (1995), "Community Characteristics, Individual Attributes and Child Survival in Brazil: A Multilevel Hazard Model Analysis," Unpublished Ph.D. dissertation, Princeton University.

Sastry, N., Goldman, N., and Moreno, L. (1993), "The Relationship Between Place of Residence and Child Survival in Brazil," in *International Population Conference, Montreal, 1993*, Vol. 3, Liège, Belgium: IUSSP, pp. 293–322.

Self, S. G., and Liang, K. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association*, 82, 605–610.

Vaupel, J. W. (1989), "Kindred Lifetimes: Frailty Models in Population Genetics," pp. 155–170 in *Convergent Issues in Genetics and Demog-*

*raphy*, eds. J. Adams, D. Lam, and P. E. Smouse, New York: Oxford University Press.

——— (1990), "Relatives' Risks: Frailty Models of Life History Data," *Theoretical Population Biology*, 37, 220–234.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1987), "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography*, 16, 439–454.

Vaupel, J. W., and Yashin, A. I. (1985), "The Deviant Dynamics of Death in Heterogenous Populations," pp. 179–211 in *Sociological Methodology 1985*, ed. N. Tuma, San Francisco: Jossey-Bass.

Zenger, E. (1993), "Siblings' Neonatal Mortality Risks and Birth Spacing in Bangladesh," *Demography*, 30, 477–488.