Biostatistics (2003), **4**, 2, *pp.* 207–222
Printed in Great Britain

Waiting for organ transplantation: results of an analysis by an Institute of Medicine Committee

ROBERT D. GIBBONS (UNIVERSITY OF ILLINOIS, CHICAGO), NAIHUA DUAN (UNIVERSITY OF CALIFORNIA AT LOS ANGELES), DAVID MELTZER (UNIVERSITY OF CHICAGO), ANDREW POPE (INSTITUTE OF MEDICINE), EDWARD D. PENHOET (UNIVERSITY OF CALIFORNIA, BERKELEY; COMMITTEE CHAIR), NANCY N. DUBLER (ALBERT EINSTEIN), CHARLES FRANCIS (DREW UNIVERSITY), BARBARA GILL (ABILENE CARDIOTHORACIC AND VASCULAR SURGERY OF TEXAS), EVA GUINAN (HARVARD UNIVERSITY), MAUREEN HENDERSON (UNIVERSITY OF WASHINGTON), SUZANNE T. ILDSTAD (UNIVERSITY OF LOUISVILLE), PATRICIA A. KING (GEORGETOWN UNIVERSITY), MANUEL MARTINEZ-MALDONADO (OREGON HEALTH SCIENCES UNIVERSITY), GEORGE E. MCLAIN (MARTIN MEMORIAL MEDICAL CENTER), JOSEPH MURRAY (HARVARD UNIVERSITY), DOROTHY NELKIN (NEW YORK UNIVERSITY), MITCHELL W. SPELLMAN (HARVARD UNIVERSITY), SARAH PITLUCK (INSTITUTE OF MEDICINE),

Center for Health Statistics, University of Illinois at Chicago, 1601 W. Taylor, Chicago, IL 60612, USA

SUMMARY

One of the most visible and contentious issues regarding the fairness of the original system of organ procurement and allocation is the argument that it resulted in great disparities in the total amount of time a patient waited for an organ (i.e. the time from registration at a transplantation center to transplant), depending on where he or she lived. In an attempt to resolve this debate, Congress charged the National Academy of Sciences, Institute of Medicine to perform an independent study of the original system and proposed rule changes. In an analysis of approximately 68 000 transplant waiting list records, the committee developed several conclusions and recommendations largely specific to liver transplantation policies. The purpose of this paper is to describe both the results of the study and the statistical foundations of the mixed-effects multinomial logistic regression model that led to the committee's conclusions.

1. INTRODUCTION

Since the enactment of the National Organ Transplant Act of 1984, the number of people receiving organs has increased steadily over time. In 1998, more than 21 000 Americans—about 57 people a day—were transplanted with a kidney, liver, heart, lung, or other organ. On any given day, approximately 62 000 people are waiting for an organ and every 16 minutes a new name is added to the national waiting list (UNOS, 1999). Moreover, although the number of donors has increased steadily since 1988, donation rates are not growing as quickly as the demand for organs (US General Accounting Office, 1997). As a result, approximately 4000 Americans die each year (11 people per day) while waiting for a solid organ transplant (UNOS, 1999).

One of the most visible and contentious issues regarding the fairness of the original system of organ procurement and transplantation was the argument that it resulted in great disparities in the amount of time potential liver transplant patients wait for a transplant, depending on where the patient lived. (The term 'waiting time' is used to refer to the time from registration at a transplantation center to transplant, death, or removal from the waiting list for other reasons.) An additional concern was that minorities and the poor may have had less access to organ transplants than did whites of higher socioeconomic status.

In response to concerns expressed about possible inequities in the original system of organ procurement and transplantation, the US Department of Health and Human Services (DHHS) published a new regulation (Final Rule) in April 1998 (42 CFR Part 121) to 'assure that allocation of scarce organs will be based on common medical criteria, not accidents of geography' (DHHS, 1998). The stated principles underlying the Final Rule included increasing federal oversight, increasing public access to data, implementing consistent medical listing criteria, placing emphasis on medical need, and reducing disparities in waiting times for transplants among different areas of the country.

Issuance of the Final Rule generated considerable controversy in the transplant community. Concerns were expressed that its implementation would increase the cost of transplantation, force the closure of small transplant centers, adversely affect access to transplantation on the part of minorities and low-income patients, discourage organ donation, and result in fewer lives saved.

In October 1998, the US Congress suspended implementation of the Final Rule for one year to allow further study of its potential impact. During that time, Congress asked the Institute of Medicine (IOM) of the National Academy of Sciences (NAS) to conduct a study to review current Organ Procurement Transplantation Network (OPTN) policies and the potential impact of the Final Rule. The IOM study was completed in July of 1999 (IOM, 1999; Gibbons *et al.*, 2000a). In the following sections, we (the members of the IOM committee and IOM staff members who worked with the committee) provide an overview of the analysis of approximately 68 000 US liver transplant waiting list records that describe every transition made by every patient on the waiting list from 1995 through the first quarter of 1999. The committee focused on liver transplants because (1) disparities in median waiting times for liver transplants was a primary factor in DHHS's rationale for developing the Final Rule, (2) liver allocation policies have been especially contentious with the OPTN making several changes in the recent past, (3) the time that a liver can viably survive outside of the body (cold ischemic time) is much longer than for hearts and lungs, making changes to the allocation system possible, and (4) the medical urgency is greater than some other organs, such as kidneys, where medical alternatives such as dialysis are available. As will be shown, detailed analysis of these data revealed the strengths and weaknesses of the original system and indicated direction for change.

It is important to note that the analysis, results, and conclusions pertain to the US organ transplant system only. The IOM committee did not attempt to provide a detailed review of liver transplant allocation policies in other developed countries.

2. MEDICAL URGENCY

A fundamental issue in the liver allocation system is the classification of patients into 'status levels' based on the current medical severity of their illness. These status levels reflect the life expectancy of patients in the absence of transplantation, which vary from a few days for status 1 (i.e. the most severely ill patients with an average life expectancy of one week), to months in status 2 (recently divided into severe (2A) and less severe (2B) chronic conditions) to potentially years in status 3 (patients in need of transplantation but not at serious risk at this stage of their illness). Moreover, more than 50% of the patients on the list are in status 3.

It should be noted that at the beginning of 1998, the status categories were clarified to create more homogeneous and reliable patient listings and the older categories of 1–4 were replaced with categories

1, 2A, 2B and 3. All statistical analyses were performed on these more recent data. However, to provide a more complete view of the overall system, tabular displays of various summary statistics (e.g. mean waiting times within status categories, percentages of transplants and deaths) utilized all data from 1995–99 and used status categories 1, 2 (2, 2A and 2B) and 3 (3 and 4).

3. Conceptual and statistical issues

As is often the case with public interpretation of complex data, the content of the data and the way they are presented are often at odds. For example, the regional inequity that led to much of the debate in the first place, was based on median waiting time for patients of all status levels combined. Since, as previously noted, waiting times vary tremendously across status categories and the least severely ill status 3 patients comprise more than 50% of the waiting list, the overall median waiting time provides little information regarding the equity of the system since it describes the waiting time for patients who have the least serious need for transplantation. Since some centers routinely list such patients and others do not, gross differences in overall waiting time distributions can and do result from simple policy differences regarding when in the course of their illness status 3 patients are placed on the transplant waiting list.

There are 63 Organ Procurement Organizations (OPOs) in the United States. OPOs are the organizations responsible for obtaining and allocating organs for transplantation. The original allocation system was a local one in which organs were offered to patients in order of status level within an OPO and were only offered outside of the OPO if no patient, regardless of status level accepts the organ. As such a severely ill status 1 patient (life expectancy of approximately one week) in an adjacent OPO would not be offered an organ that was accepted by a status 3 patient within the OPO (life expectancy of a year or more).

In our analysis, we assessed regional difference as a random effect at the OPO level. Given the heterogeneity of the waiting times across status levels, we conducted stratified analyses for each status level (1, 2A, 2B, and 3) to control for the impact of status level on waiting time. The stratified analysis allows us to examine how the original organ allocation system worked for patients (patient-times) in each status level. Statistically, the validity of this approach rests on the underlying assumption of no frailty which allows us to break up a single record from a patient who changes status into multiple independent records. Future statistical work in this area should consider alternatives such as two-stage models (Duan *et al.*, 1984) which model both changes in status and time to transplant or death.

Conceptually, our objective was to evaluate the competing risks of transplant and mortality over time as a function of various predictors such as age, sex, blood type, and OPO size. In order to examine the concern about geographical inequity, we specified OPO as a random effect for both transplantation and mortality rates. In other words, we allowed each rate to vary across OPOs, and model the variations across OPOs as a random effect. The magnitudes of the OPO variance component describe geographic variability in transplantation and mortality rates.

In terms of time, we have used the alternative parametrization of the Cox proportional hazards model in terms of a ‘partial logistic regression’ model (Efron, 1988) or ‘person-time logistic regression’ model (Ingram and Kleinman, 1989). This approach to survival analysis involves the use of a series of sequential records from each subject for the period of time that they were observed in the study. For example, a patient who endured 4 days in status 1 and was transplanted would have four records with an outcome of 0 for days 1–3 and an outcome of 1 (transplant) on day 4. Efron (1988) and Ingram and Kleinman (1989) have shown that this approach to modeling time to event data provides excellent agreement with the traditional proportional hazards survival model and becomes identical as the time intervals go to zero (i.e. approach continuous time). The advantage of the approach in the present setting is that we can (a) simultaneously model both transplantation and mortality rates and (b) we can accommodate OPO-specific components of variability in those rates.

4. MIXED-EFFECTS MULTINOMIAL REGRESSION MODEL

The statistical development of the general model is described by Hedeker and Gibbons (1994) and Hedeker (1999). Note that as previously described, the unit of analysis is the patient-day (status 1 and 2A) or patient month (status 2B and 3) and not the patient. Following Efron (1988) we assume that days (or months in status 2B and 3) within patients are conditionally independent on the prior days (or months) as long as the competing risk outcomes of interest (i.e. death or mortality) can only occur on the final day (or month) for each subject. Using the terminology of multilevel analysis (Goldstein, 1995) let i denote the level 2 units (OPOs) and let j denote the level 1 units (patient-days (or months) within OPOs). Assume that there are $i = 1, \ldots N$ level 2 units (i.e. OPOs) and $j = 1, \ldots, n_i$ (unequal number of person-times for each OPO) level 1 patient-days (or months) nested within each OPO. The n_i patient-day (or month) measurements include the set of all available measurement days (or months) for all patients in OPO i (i.e. n_i is the total number of daily measurements in OPO i). Due to the local nature of the original allocation system, we did not incorporate inter-OPO (i.e. geographic) correlation in our statistical model. Let y_{ij} be the value of the categorical outcome variable associated with level 2 unit i and level 1 unit j. In our case, these represent transplant, death and other and we code the $K + 1$ response categories as 0, 1, 2 ($K = 2$).

Adding random effects to the multinomial logistic regression model of Bock (1970), Nerlove and Press (1973) and others, the probability for a given OPO i, and patient-day (or month) j, $y_{ij} = k$ (a response occurs in category k), conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ is

$$P_{ijk} = P(y_{ij} = k \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\exp(z_{ijk})}{1 + \sum_{h=1}^{K} \exp(z_{ijh})} \quad \text{for } k = 1, 2, \ldots K \tag{4.1}$$

$$P_{ij0} = P(y_{ij} = 0 \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{1 + \sum_{h=1}^{K} \exp(z_{ijh})} \tag{4.2}$$

where $z_{ijk} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_{ik} + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_k$. Here, $\boldsymbol{w}_{ij}$ is the $p \times 1$ covariate vector and $\boldsymbol{x}_{ij}$ is the design vector for the r random effects, both vectors being for the jth patient-day (or month) nested within OPO i. Correspondingly, $\boldsymbol{\alpha}_k$ is a $p \times 1$ vector of unknown fixed regression parameters, and $\boldsymbol{\beta}_{ik}$ is a $r \times 1$ vector of unknown random effects for OPO i. For the general case of multiple random effects, their distribution is assumed to be multivariate normal with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. Notice that the regression coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ carry the k subscript. Thus, for each of the p covariates and r random effects, there will be K parameters to be estimated. Additionally, the random effect variance–covariance matrix $\boldsymbol{\Sigma}_k$ is allowed to vary with k.

It is convenient to standardize the random effects by letting $\boldsymbol{\beta}_{ik} = \boldsymbol{T}_k\boldsymbol{\theta}_i + \boldsymbol{\mu}_k$, where $\boldsymbol{T}_k\boldsymbol{T}'_k = \boldsymbol{\Sigma}_k$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_k$. The model is now given as

$$z_{ijk} = \boldsymbol{x}'_{ij}(\boldsymbol{T}_k\boldsymbol{\theta}_i + \boldsymbol{\mu}_k) + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_k \ , \tag{4.3}$$

where $\boldsymbol{\theta}_i$ are mutually independent $N(0, 1)$ variates.

Parameter estimation Let $\boldsymbol{y}_i$ denote the vector of categorical responses from OPO i for all n_i patient-day (or month) measurements nested within. Then the probability of any $\boldsymbol{y}_i$, conditional on the random effects $\boldsymbol{\theta}$ and given $\boldsymbol{\alpha}_k$, $\boldsymbol{\mu}_k$, and $\boldsymbol{T}_k$, is equal to the product of the probabilities of the patient-day responses:

$$\ell(\boldsymbol{y}_i \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_k, \boldsymbol{\mu}_k, \boldsymbol{T}_k) = \prod_{j=1}^{n_i} \prod_{k=0}^{K} \left[P(y_{ij} = k \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_k, \boldsymbol{\mu}_k, \boldsymbol{T}_k)\right]^{d_{ijk}} \tag{4.4}$$

where $d_{ijk} = 1$ if $y_{ij} = k$, and 0 otherwise. Thus, associated with the response from a particular patient-day, $d_{ijk} = 1$ for only one of the $K + 1$ categories and zero for all others. The marginal density of the

response vector $\boldsymbol{y}_i$ in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the density $g(\cdot)$:

$$h(\boldsymbol{y}_i) = \int_{\boldsymbol{\theta}} \ell(\boldsymbol{y}_i \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_k, \boldsymbol{\mu}_k, \boldsymbol{T}_k)\, g(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} \tag{4.5}$$

where $g(\boldsymbol{\theta})$ represents the population distribution of the random effects.

For parameter estimation, the log-likelihood from the N OPOs can be written as: $\log L = \sum_i^N \log h(\boldsymbol{y}_i)$, which we maximize using Fisher scoring, using Gauss–Hermite quadrature (Stroud and Sechrest, 1966) to evaluate (5).

In general, the total number of parameters equals the $K \times p$ fixed regression coefficients ($\boldsymbol{\alpha}_k$; $k = 1, \ldots, K$), plus the $K \times r$ means of the random effects ($\boldsymbol{\mu}_k$; $k = 1, \ldots, K$), and the $K \times r \times (r-1)/2$ random effect variance–covariance terms (v[$\boldsymbol{T}_k$]; $k = 1, \ldots, K$). Notice that the parameter vector v($\boldsymbol{T}_k$), which indicates the degree of OPO population variance, is what distinguishes the mixed-effects model from the ordinary fixed-effects multinomial logistic regression model.

Comparison of alternative models can be performed using a Wald test (Wald, 1943), or likelihood ratio (LR) test (Cox and Hinkley, 1974). Under a number of regularity conditions, the test statistic $-2(\log LR)$ has a χ^2 distribution with degrees of freedom given by the difference in number of parameters between the null and alternative hypotheses. While the LR test as described above is appropriate for testing fixed-effects in the model, Stram and Lee (1994) have shown that this is not the case for variance components in mixed-effects regression models because the variance components lie on the boundary of the parameter space. Stram and Lee (1994) applied the method of Self and Liang (1987) to determine the correct asymptotic distribution of the LR test for a generalized mixed-model and found that for a test of a single random effect, it is a 50 : 50 mixture of χ^2 distributions but that use of a single χ^2 produces only a minimal degree of bias. More recently, Morrell (1998) studied a wide variety of cases for differing numbers of fixed and random effects, clusters and sample sizes and has confirmed earlier results that for a test of a single additional random effect, the usual LR test may be applied to obtain valid inferences about the number of variance components in the model. Alternatively, Morrell (1998) suggests that restricted maximum likelihood (REML) test statistic performs slightly better than the maximum likelihood test statistic, in the sense that the observed probability value for the REML test is on average closer to the nominal level.

Hazard rates and cumulative survival For a model with two random-effects (i.e. one for transplantation and one for pre-transplantation mortality) and three categories, we can estimate the probability of each outcome conditional on a particular covariate vector as

$$P_{ij2} = \frac{\exp(\sigma_2\theta_i + \mu_2 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_2)}{1 + \exp(\sigma_1\theta_i + \mu_1 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_2)} \tag{4.6}$$

$$P_{ij1} = \frac{\exp(\sigma_1\theta_i + \mu_1 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_1)}{1 + \exp(\sigma_1\theta_i + \mu_1 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_2)} \tag{4.7}$$

$$P_{ij0} = \frac{1}{1 + \exp(\sigma_1\theta_i + \mu_1 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_1) + \exp(\sigma_2\theta_i + \mu_2 + \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_2)}. \tag{4.8}$$

These are referred to as 'unit-specific' probabilities because they indicate response probabilities for particular values of the random unit (OPO) effect θ_i (Neuhaus *et al.*, 1991; Zeger *et al.*, 1988). Replacing the parameters with their estimates and denoting the resulting unit-specific probabilities as $\hat{P}_{ss}$, marginal

probabilities $\hat{P}_m$ are then obtained by integrating over the random-effect distribution, namely $\hat{P}_m = \int_\theta \hat{P}_{ss}\, g(\theta)\, \mathrm{d}\theta$. Numerical quadrature can be used for this integration as well. These marginal probabilities represent the hazard rate for a particular competing risk of interest (i.e. transplant, mortality or other) expressed as a daily rate for status 1 or monthly rate for status 2B and 3 patients. The cumulative survival rate is then computed by summing the daily risk for status 1, or monthly risk in the case of status 2B and 3, over time, adjusting for the number of subjects remaining on the list at that time point.

Intra-class correlation. The intra-class or intra-cluster correlation is a useful method of expressing the magnitude of the level 2 variance component (i.e. OPO variance). For the random-intercepts regression model used here, we denote the underlying response tendency associated with category k for person-time j in OPO i as Y_{ijk}. The random-intercepts regression model for the latent variable Y_{ijk} is therefore

$$Y_{ijk} = \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_k + \sigma_k\theta_i + \varepsilon_{ijk} \quad k = 0, \ldots, K. \tag{4.9}$$

Assuming $k = 0$ as the reference category, $\boldsymbol{\alpha}_0 = \sigma_0 = 0$, the model can be rewritten as

$$Y_{ijk} = \boldsymbol{w}'_{ij}\boldsymbol{\alpha}_k + \sigma_k\theta_i + (\varepsilon_{ijk} - \varepsilon_{ij0}) \quad k = 1, \ldots, K,$$

(see Hedeker, in press). The level 1 residuals ε_{ijk} are distributed according to a type I extreme-value distribution (Maddala, 1983, page 60). Furthermore, the standard logistic distribution is obtained as the difference of two independent type I extreme-value variates (McCullagh and Nelder, 1989, pages 20 and 142). As a result, the level 1 variance is given by $\pi^2/3$, which is the variance of the standard logistic distribution. The intra-class correlation is therefore estimated as $r_k = \hat{\sigma}_k^2/(\pi^2/3 + \hat{\sigma}_k^2)$, where $\hat{\sigma}_k^2$ is the estimated level 2 variance for category k assuming a normally distributed random-effect distribution.

All computations were performed using the MIXNO program developed under a grant from the National Institute of Mental Health and available at no charge at http://www.uic.edu/labs/biostat/.

5. MODEL SPECIFICATION

Stratified analyses were performed separately for the time spent in each status level, with the exception of 2A for which there were too few subjects. For status 1, time refers to days, whereas for status 2B and 3, time refers to months. In all analyses, the outcome measure is the nominal measure of transplant, death or other. Other can be shifting to another status level and never returning to the status level in question, being too sick to transplant, being delisted, being transplanted at another center, or still waiting. In terms of covariates, we have age (0–5, 6–17, 18 and over), sex (female = 0, male = 1), race (black = 1 else 0), blood type (O or B = 1 else 0) and OPO transplant volume (small, medium and large based on number of transplanted patients in 1995–99). For blood type, we selected a contrast between types O and B versus A and AB because the former two can only receive donation from a subset of donors whereas the latter can receive donation from almost all potential donors. Categorical predictors such as age and OPO transplant volume were dummy-coded in the analysis so individual groups could be compared without assuming a functional form for the relationship (e.g. linearity).

6. RESULTS

A summary of several statistics of interest that help characterize the sample and waiting time distributions is presented in Table 1. Maximum marginal likelihood estimates (MMLE), standard errors

Table 1. *Characteristics of liver transplant patients by status, 1995–99*

	Totals	Status 1	Status 2	Status 3
Total patients, 1995–99	33 286	5294	14 264	26 907
Percentage receiving a transplant	47.1	52.4	50.2	21.3
Percentage dying prior to transplantation	8.3	9.2	6.1	5.2
Percentage post-transplant mortality	5.4	11.1	5.0	1.9
Percentage male	58.7	54.1	59.9	58.7
Percentage with A or AB blood type	16.0	15.3	15.4	15.8
Percentage African American	7.7	11.2	8.3	6.9
Mean age (years)	45.0	36.3	44.9	46.1
Mean waiting time (days)	255.6	4.8	56.8	285.1

The 'Totals' column reflects number of unique listings and therefore does not equal the sum of the other three columns which count patients within status levels (a given patient may be counted in up to three status levels for a particular listing) Source: IOM, 1999.

(SE) and corresponding probabilities for Wald test statistics are presented in Table 2, separately for status levels 1, 2B and 3 respectively. In the following we provide an overview of the most important findings.

Geographic inequity. In terms of geographic inequity, systematic OPO-specific effects accounted for less than 5% of the total variance (i.e. intra-cluster correlation of 0.045) in transplantation rates for status 1 patients (see Table 2). The geographic distribution for the most severely ill patients is therefore reasonably equitable with mean waiting time of 4.8 days (see Table 1). In contrast, OPO-specific effects accounted for 13% of the variability in transplantation rates for status 2B patients (see Table 2) and 35% of the variability for status 3 patients (see Table 2). This implies that the systematic variation in waiting time across OPOs is almost completely determined by variations in waiting times for the less severely ill patients, with little variation for the most severely ill patients. This finding is further illustrated in Figure 1 where the Empirical Bayes (EB) estimates of the OPO-specific adjusted transplantation effects are displayed. The interpretation of the EB estimate is OPO i's deviation from the popualation rate, adjusted for covariates. As such, the y-axis in Figure 1 is in a log-odds scale, where values of 1, 2, and 3 represents increases in the likelihood of transplantation by factors of 2.7, 7.4 and 20.1 respectively. Negative values represent corresponding decreases in probability of transplantation relative to the overall population rate. Inspection of Figure 1 clearly illustrates that the greatest geographic variation in adjusted transplantation rates is for the least severely ill patients. Moreover, while transplantation rates of these less severely ill patients vary significantly, they have little relationship to mortality. In all cases, systematic OPO differences in pre-transplantation mortality rates accounted for less than 1% of the variation in overall adjusted mortality rates. No significant effects of race or gender were observed indicating that the system is equitable for woman and minorities once they are listed.

Table 2. *Mixed-effects competing risk survival models for patient time in status levels 1, 2B, and 3: maximum marginal likelihood estimates (standard errors)*

	Status 1		Status 2B		Status 3	
Transplant versus other						
Intercept	−1.829 (0.276)		−2.077 (0.129)		−3.593 (0.210)	
Day (1) Month (2B, 3)	0.016 (0.015)		−0.092 (0.016)	***	−0.220 (0.030)	***
Age 0–5 vs Adult	−0.907 (0.188)	***	0.470 (0.103)	***	1.156 (0.154)	***
Age 6–17 vs Adult	−0.362 (0.234)		0.135 (0.243)		0.844 (0.268)	**
Gender (1 = male)	−0.098 (0.198)		0.126 (0.087)		0.054 (0.186)	
Race (1 = black)	−0.275 (0.268)		0.134 (0.222)		0.158 (0.304)	
Blood Type (1 = B or O)	−0.076 (0.196)		−0.577 (0.062)	***	−0.477 (0.098)	***
OPO volume (M vs L)	−0.054 (0.319)		0.590 (0.157)	***	1.179 (0.149)	***
OPO volume (S vs L)	0.261 (0.336)		0.560 (0.187)	**	0.757 (0.228)	***
Random OPO effect SD	0.393 (0.144)	**	0.689 (0.064)	***	1.335 (0.162)	***
Mortality versus Other						
Intercept	−3.685 (0.482)		−3.313 (0.227)		−3.654 (0.172)	
Day (1), Month (2B, 3)	0.023 (0.047)		−0.213 (0.039)	***	−0.216 (0.041)	***
Age 0–5 vs Adult	−0.968 (0.378)	**	−0.195 (0.381)		−2.119 (2.099)	
Age 6–17 vs Adult	−1.001 (0.551)		−0.516 (0.641)		−1.193 (2.000)	
Gender (1 = male)	0.077 (0.371)		0.014 (0.191)		−0.063 (0.268)	
Race (1 = black)	0.162 (0.448)		−0.082 (0.359)		0.027 (0.544)	
Blood Type (1 = B or O)	0.003 (0.433)		−0.005 (0.164)		−0.017 (0.231)	
OPO volume (M vs L)	0.203 (0.491)		0.202 (0.126)		−0.526 (0.300)	
OPO volume (S vs L)	−0.230 (0.930)		0.355 (0.151)	**	−0.658 (0.358)	
Random OPO effect SD	0.042 (0.298)		0.116 (0.049)	**	0.137 (0.157)	

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

OPO volume and size. Smaller OPOs are more likely than larger OPOs to transplant status 2B and 3 patients (see Table 2). For status 1 patients, OPO size played no role in transplantation or mortality rates. In contrast, for status 2B and 3 patients, OPO size was significantly related to transplantation rates. For large OPOs (9+ million) the initial one month transplantation rates were 5% for status 2B patients and 3% for status 3 patients. By contrast in the smaller OPOs (four million or less), initial one-month transplantation rates were as high as 17% for status 2B patients and 9% for status 3 patients. Based on these results we recommended that at least nine million people be included in an organ allocation region to maximize the chance of transplantation for the most severely ill patients (for a detailed analysis of OPO size, see the original IOM report).

As further support for this recommendation, we (Gibbons *et al.*, 2000b) examined all new listings in 1998 ($n = 9585$). Of these, 731 were in status 1, 346 were in status 2A, 2257 were in status 2B, and 6251 were in status 3. Of the 731 status 1 patients, 417 were transplanted (57%). Based on the estimated marginal probabilities, we determined that in the first month after listing, an estimated excess of 155 status 2B patients and an estimated excess of 143 status 3 patients would have been transplanted in OPOs serving populations of less than nine million people, relative to the expected number for OPOs serving populations of nine million or more people. This finding indicates that had the minimum population size served by an OPO been nine million people, a minimum of 298 additional organs would have been available for status 1 patients. This is a minimum since the cumulative excess probability of transplantation of status 2B and 3 patients in small OPOs increases over time. In fact, 314 patients initially listed as status 1 were not transplanted under the existing system, and these organs would have likely made it to those patients in greatest need, had the IOM recommendations been followed in 1998.

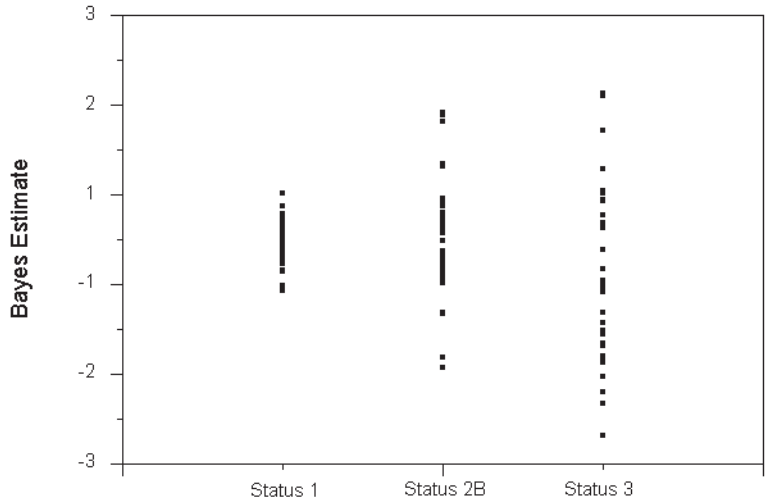

Fig. 1. OPO-specific Bayes estimates ($\sigma_1\theta_i$) of transplantation rates by status category adjusted for competing risk of mortality and model covariates (e.g. sex, race, blood type).

The effect of sharing. We further examined this finding by analyzing the results of several regional and state-wide sharing arrangements among two or more OPOs, most typically for status 1 patients. Our analysis of these 'natural experiments' revealed that sharing significantly increased the status 1 transplantation rate from 42% without sharing to 52% with sharing, lowered average status 1 waiting times from 4 to 3 days, and decreased status 1 pre-transplantation mortality from 9% to 7%. Not surprisingly, sharing significantly decreased the rate of transplantation for less severely ill patients. For example, among small OPOs that served a population of 2 million or less, the status 3 transplantation rate decreased from 31% for those OPOs that did not share to 6% for those that did share, making more organs available for more severely ill patients. Though sharing decreased status 3 transplantation rates, we did not find a concomitant increase in pre-transplantation mortality of status 3 patients.

Waiting times and need for transplant. Other interesting results of our analysis concerned the relationship of transplantation and pre-transplantation mortality to waiting times. For status 1 patients, the rates were constant over the first 12 days of listing at approximately 15% for transplantation per day and 3% for mortality (see Figure 2, panel A), but for status 2B and 3 patients both rates decreased rapidly over time (see Figure 2, panels B and C). For status 2B, transplantation rates decreased from 12% to 5% per month over a 12 month period while pre-transplant mortality rates decreased from 3% to 0.3% per month. For status 3 patients, transplantation rates decreased from 4% to 0.05% per month over a 12 month period and pre-transplant mortality rates decreased from 2% to 0.2% per month. These findings indicate that waiting time is inversely related to medical need in the less severely ill patients and should therefore not be used as a criterion for transplantation (as it had traditionally been) in status 2B and 3. Figure 3 displays estimated cumulative time-to-event distributions for status 1 (panel A), status 2B (panel B), and status 3 (panel C). Inspection of Figure 3 reveals that after 12 days 80% of the status 1 patients at risk are transplanted

whereas 10% die while waiting. For status 2B, 60% of patients at risk through 12 months are transplanted and 7% die while waiting. For status 3, 20% of patients at risk through 12 months are transplanted and 8% die while waiting. These estimates differ somewhat from the observed rates in Table 1 because subjects were not always observed for these period of times due to becoming too ill to transplant (and removed from the list) or changing status category.

Post-transplantation survival. The benefits of transplantation depend in large part on a patient's survival afterwards. Therefore we also examined post-transplant mortality data for all patients transplanted in 1998 and 1999 (see Table 3). Not surprisingly, mortality risk is highest immediately after transplantation and declines over time, and patients transplanted in status 2B and 3 both had substantially lower mortality than patients transplanted in status 1. As noted in Table 1, for the period of 1995–99 post-transplant mortality was 11.1% for status 1 patients but only 5% for status 2 and 1.9% for status 3. Of course, with few if any exceptions, all status 1 patients who are not transplanted will die, but pre-transplant mortality rates for status 2 and 3 patients were 6.1% and 5.2% respectively. As such, transplantation of status 1 patients first can save 88.9% of those patients (over the five-year time period studied), whereas transplantation of the less severely ill status 2 and 3 patients saved only 1.1% and 3.3% respectively. Additionally, patients in small OPOs had increased risk of mortality relative to those in larger OPOs. The reasons for increased mortality rates associated with smaller OPOs are not clear. A question of serious concern is whether this increased mortality is a consequence of the smaller number of procedures performed by the centers in the smaller OPOs.

Specification of the random-effect distribution. In examining the robustness of our results to model specification, we compared the MMLEs and SEs for the status 1 patient data for models with Gaussian and rectangular random-effect distributions. The results of this comparison are presented in Table 4. Table 4 reveals that there is virtually no effect of the specification of the random-effect distribution on the estimates and standard errors for the fixed effects in the model. As expected, the random-effect variance for the uniform distribution is different from that estimated for Gaussian random-effects. In either case (i.e. Gaussian or uniform random-effect distribution) the random-effects are given by $\sigma_{k_i} O_i$, however, in the Gaussian case the variance of O_i is unity, whereas in the uniform case the variance of O_i is $(a-b)^{2/12}$, where a and b are the maximum and minimum quadrature nodes respectively. In the current example, $a = 4.859$ and $b = -4.859$ (i.e. the extreme Gauss–Hermite quadrative nodes with 10 points). As such, the adjusted random-effect standard deviation for transplantation is $\sigma_1 = \sqrt{-131^2}(9.718^2/12) = 0.368$ which is similar to the value of $\sigma_1 = 0.393$ estimated for one case of Gaussian random effects (see Table 4).

7. DISCUSSION

From a statistical perspective, this case study is useful in that it illustrates the important role that statisticians play in the development of public policy. Clearly, without the analysis of these data, the IOM committee would not have reached these conclusions, and the life-saving changes in the regulations would not have occurred. This paper also illustrates that relatively complex statistical methods can, in fact, be used to evaluate important public policy questions and to communicate the results to a largely nonstatistical audience, including the US Congress. The notion that complex statistical methods will hinder the dissemination of research findings to policy makers and the public, was not borne out in our experience.

Our analysis revealed that the rate of liver transplantation of the most severely ill (status 1) patients is relatively homogeneous across OPOs. However, the allocation of organs is not efficient because many organs that could be used for transplantation of status 1 patients are being transplanted into less severely ill patients, especially in smaller OPOs. Our analysis revealed that this inefficiency resulting from small

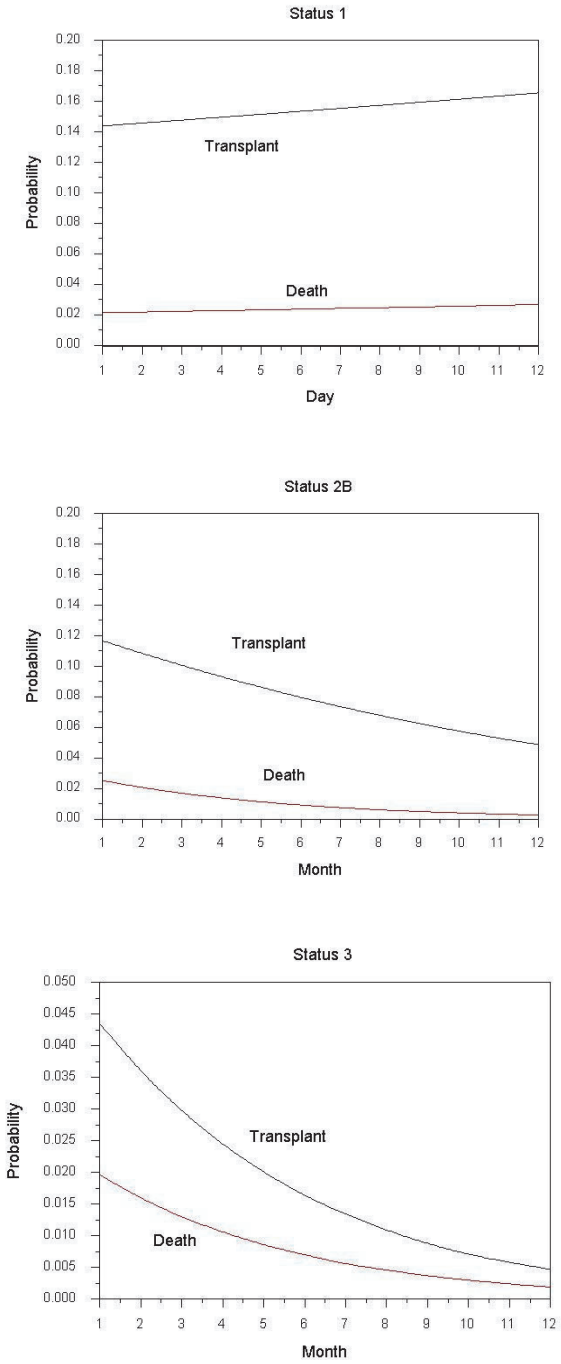

Fig. 2. Estimated hazard rates for (a) status 1, (b) for status 2B, and (c) status 3 patients awaiting liver transplantation. The hazard rate describes the likelihood of transplantation or mortality at a given point in time (using one whole day (status 1) or one whole month (status 2B and 3) as the unit) adjusted for the competing risks (i.e. transplantation or mortality) and the model covariates (e.g. sex, race, blood type).

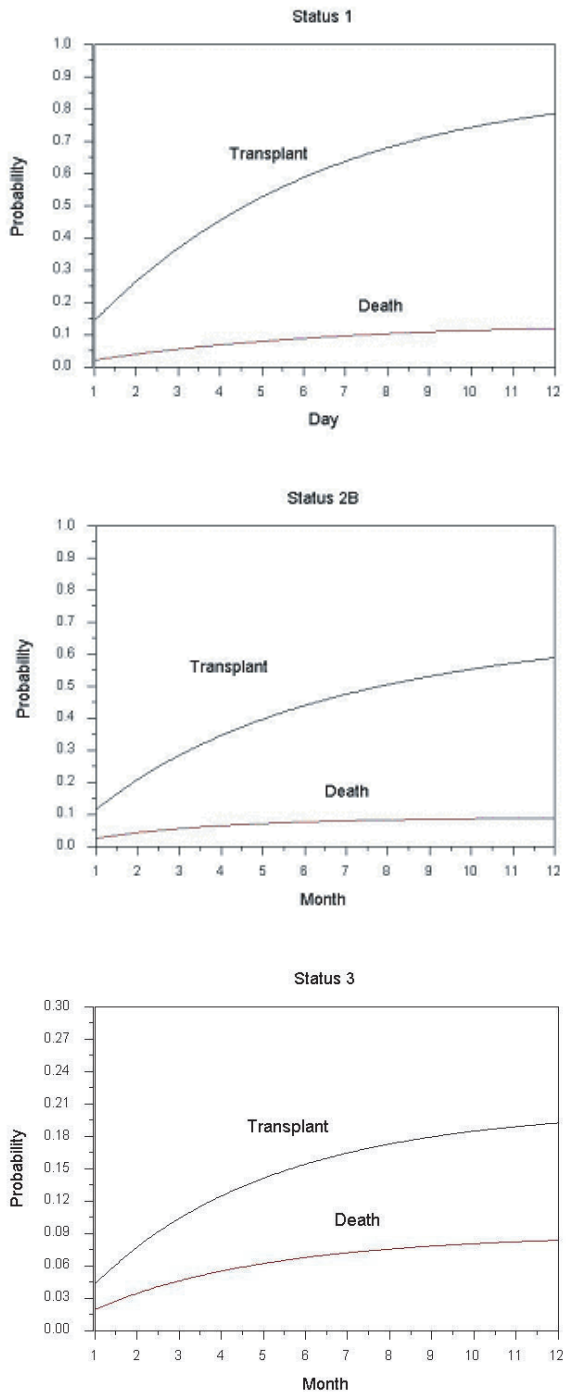

Fig. 3. Estimated cumulative time-to-event distributions (a) status 1, (b) for status 2B, and (c) status 3 patients awaiting liver transplantation. The cumulative time-to-event distribution describes the overall adjusted likelihood of transplantation or mortality up to a particular point in time.

Table 3. *Mixed-effects survival model post-transplant mortality: maximum marginal likelihood estimates (standard errors)*

	MMLE	SE	P
Intercept	−2.409	0.399	0.001
Month	−0.573	0.084	0.001
Age 0–5 vs Adult	0.150	0.370	0.685
Age 6–17 vs Adult	0.310	0.494	0.530
Gender (1 = male)	0.094	0.318	0.766
Race (1 = black)	−0.073	0.531	0.891
Blood Type (1 = B or O)	0.191	0.299	0.524
OPO volume (M vs L)	0.026	0.372	0.944
OPO volume (S vs L)	0.862	0.330	0.009
Status 2B vs 1	−0.705	0.312	0.024
Status 3 vs 1	−1.302	0.510	0.011
Random OPO Effect SD	0.307	0.427	0.236

Table 4. *Comparison of normal and rectangular random-effects distributions for patient time in status levels 1: maximum marginal likelihood estimates (standard errors)*

	Normal	Rectangular
Transplant versus other		
Intercept	−1.829 (0.276)	−1.837 (0.261)
Day (1) Month (2B, 3)	0.016 (0.015)	0.016 (0.015)
Age 0–5 vs Adult	−0.907 (0.188)	−0.894 (0.188)
Age 6–17 vs Adult	−0.362 (0.234)	−0.364 (0.236)
Gender (1 = male)	−0.098 (0.198)	−0.101 (0.196)
Race (1 = black)	−0.275 (0.268)	−0.279 (0.282)
Blood Type (1 = B or O)	−0.076 (0.196)	−0.070 (0.195)
OPO volume (M vs L)	−0.054 (0.319)	−0.070 (0.317)
OPO volume (S vs L)	0.261 (0.336)	0.263 (0.336)
Random OPO Effect SD	0.393 (0.144)	0.131 (0.040)
Mortality versus Other		
Intercept	−3.685 (0.482)	−3.693 (0.458)
Day (1), Month (2B, 3)	0.023 (0.047)	0.025 (0.046)
Age 0–5 vs Adult	−0.968 (0.378)	−0.970 (0.409)
Age 6–17 vs Adult	−1.001 (0.551)	−1.001 (0.552)
Gender (1 = male)	0.077 (0.371)	0.081 (0.344)
Race (1=black)	0.162 (0.448)	0.173 (0.493)
Blood Type (1=B or O)	0.003 (0.433)	−0.010 (0.441)
OPO volume (M vs L)	0.203 (0.491)	0.208 (0.470)
OPO volume (S vs L)	−0.230 (0.930)	−0.217 (0.908)
Random OPO Effect SD	0.042 (0.298)	0.014 (0.101)

OPOs is minimized when allocation areas comprise a minimum population of nine million people. Our analysis of existing broader sharing arrangements confirms this result by demonstrating that broader sharing of organs led to an overall increase in the rate at which the most severely ill patients were transplanted and a concomitant decrease in the excess transplantation of the least severely ill patients, without increasing pre-transplantation mortality.

Since the IOM report was released, several parties have cited the report, each claiming that it supports their preferred policy. For example, some members of Congress (proponents of the bill HR-2418) have claimed that we found the original system is 'fine as is'. Our analysis does not support such a claim. Some are concerned that our recommended allocation system might force smaller transplant centers to close. We found no evidence that this would occur. Nor did we find evidence that distributing organs across broader areas would drive down donation rates. In addition, we found no evidence to support the suggestion that minorities and economically disadvantaged patients would be adversely affected by broader sharing of organs. The evidence suggests that their obstacles to transplantation stem from access to health insurance and other socioeconomic factors. Additionally, our analysis revealed that once patients are listed for an organ transplant, there are no disparities by race or sex in how long it takes to receive an organ. Furthermore, as concluded in our report, despite objections and arguments from many in the transplant community, it is clear to us that the nation's transplant system needs more cohesive and attentive oversight from the federal government aided by an independent scientific review board. Such a board has now been established. Finally, patients, families, health care providers, and potential donors need much better information about organ transplantation. Timely and more readily available data, that have been independently reviewed for accuracy and relevance, would build confidence in the system's fairness and ensure that it continues to improve. Although a suitable donor organ cannot be provided for every person who needs one at this time, the improvements we recommended would help ensure that those patients who are in greatest need of a transplant receive the highest priority.

The revised regulation, which adopted all of these recommendations, became law on March 16, 2000. Despite this important regulatory advance, some members of Congress and some state governments have continued to block implementation of the new regulation in favor of continued use of a local allocation system. The debate continues.

REFERENCES

BOCK, R. D. (1970). Estimating multinomial response relations. In Bose, R. C. (ed.), *Contributions to Statistics and Probability*, Chapel Hill, NC: University of North Carolina Press, pp. 453–479.

BOCK, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37**, 29–51.

BOCK, R. D. AND AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika* **46**, 443–459.

BOCK, R. D., GIBBONS, R. D. AND MURAKI, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement* **12**, 261–280.

BRYK, A. S. AND RAUDENBUSH, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

COX, D. R. AND HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

DEJONG, W., DRACHMAN, J. AND GORTMANKER, S. L. (1995). Options for increasing organ donation: The potential role of financial incentives, standardized hospital procedures, and public education to promote family discussion. *Milbank Quarterly* **73**, 463–479.

DHHS (DEPARTMENT OF HEALTH AND HUMAN SERVICES) (1998). Organ procurement and transplantation

network, Final Rule (40 CFR Part 121). *Federal Register* **63**, 16296–16338.

DUAN, N., MANNING, W. G., MORRIS, C. N. AND NEWHOUSE, J. P. (1984). Choosing between the sample selection model and the multi-part model. *Journal of Business and Economic Statistics* **2**, 283–289.

EFRON, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association* **83**, 414–425.

US GENERAL ACCOUNTING OFFICE (1997). Report to the ranking minority member, Committee on Labor and Human Resources, US Senate. *Organ Procurement Organizations: Alternatives Being Developed to More Accurately Assess Performance*, GAO/HEHS-98-26. Washington, DC: GAO.

GIBBONS, R. D., MELTZER, D., DUAN, N., PENHOET, E. D., DUBLER, N. N., FRANCIS, C. K., GILL, B., GUINAN, E., HENDERSON, M., ILDSTAD, S. T., KING, P. A., MARTINEZ-MALDONADO, M., MCLAIN, G. E., MURRAY, J. E., NELKIN, D., SPELLMAN, M. W, POPE, A. AND PITLUCK, S. (2000a). Waiting for organ transplantation. *Science* **287**, 237–238.

GIBBONS, R. D., DUAN, N. AND MELTZER, D. (2000b). Inequities in liver transplant allocation. *Science* **289**, 549.

GOLDSTEIN, H. (1995). *Multilevel Statistical Models*. New York: Halsted Press.

HEDEKER, D. AND GIBBONS, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–944.

HEDEKER, D. (1999). A mixed-effects multinomial logistic regression model. *Journal of Statistical Software* **5**, 1–92.

HEDEKER, D. (in press). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*.

INGRAM, D. D. AND KLEINMAN, J. C. (1989). Empirical comparisons of proportional hazards and logistic regression models. *Statistical in Medicine* **8**, 525–538.

INSTITUTE OF MEDICINE (1999). *Organ Procurement and Transplantation: Assessing Current Policies and the Potential Impact of the DHHS Final Rule*. Washington D.C.: National Academy Press.

MADDALA, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

MAGNUS, J. R. (1988). *Linear Structures*. London: Charles Griffin.

MCCULLAGH, P. AND NELDER, J. A. (1989). *Generalized linear models*, 2nd edition. New York: Chapman and Hall.

MORRELL, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* **54**, 1560–1568.

NERLOVE, M. AND PRESS, S. J. (1973). Univariate and multivariate log-linear and logistic models. *Rand Corporation Technical Report*, R-1306-EDA/NIH. Santa Monica, CA.

NEUHAUS, J. M., KALBFLEISCH, J. D. AND HAUCK, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* **59**, 25–35.

SELF, S. G. AND LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

STRAM, D. O. AND LEE, J. W. (1994). Variance component testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.

STROUD, A. H. AND SECHREST, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall.

UNITED NETWORK FOR ORGAN SHARING (Accessed July 1, 1999). Available at: http://www.unos.org.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426–482.

ZEGER, S. L., LIANG, K. Y. AND ALBERT, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060.

[*Received June 28, 2001; first revision March 5, 2002; second revision May 17, 2002; accepted for publication May 20, 2002*]