

Estimation and Comparison of Changes in the Presence of Informative Right  
Censoring By Modeling the Censoring Process

Margaret C. Wu

Biometrics Research Branch  
National Heart, Lung, and Blood Institute  
Bethesda, Maryland 20892

Raymond J. Carroll<sup>1</sup>

Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27514

<sup>1</sup>Research supported by Air Force Office of Scientific Research contract  
AFOSR-F-49620-85-C-0144.

## SUMMARY

In estimating and comparing the rates of change of a continuous variable between two groups, the unweighted averages of individual simple least squares estimates from each group are often used. Under a linear random effects model, when all individuals have complete observations at identical time points these statistics are maximum likelihood estimates for the expected rates of change. However, with censored or missing data, these estimates are no longer efficient when compared to generalized least squares estimates. When, in addition, the right censoring process is dependent upon the individual rates of change (i.e., informative right censoring), the generalized least squares estimates will be biased. Likelihood ratio tests for informativeness of the censoring process and maximum likelihood estimates for the expected rates of change and the parameters of the right censoring process are developed under a linear random effect models with a probit model for the right censoring process. In realistic situations, we illustrate that the bias in estimating group rate of change and the reduction of power in comparing group difference could be substantial when strong dependency of the right censoring process on individual rates of change is ignored.

Some Key Words: Informative right censoring; Linear random effect; Probit right censoring; Rate of change.

SECTION 1 : Introduction

In clinical trials and longitudinal studies it is often of interest to estimate and compare the rates of change of one or more variables between groups, in e.g., lung function or tumor growth. Furthermore, comparing the rates of change of a continuous response variable between two treatment groups is often the primary objective. Death or withdrawal may cause some observations of the primary variable to be right censored.

Growth curve methods for comparing rates of change have been studied extensively, see Rao (1965), Fearn (1975) and Schlesselman (1973). Most of these analyses assume that there are no right censored or missing observations. Maximum likelihood and generalized weighted least squares provide alternative approaches to simple least squares for the analysis of series measurements when some observations are right censored or missing. Koziol, et al. (1981) proposed a distribution-free test for the comparison of growth curves with incomplete data. In order to be valid, these procedures require that the probabilities of right censoring or missing do not depend on the parameter values of the response under investigation, i.e., they are non-informative with respect to the response parameters.

In this paper we are primarily interested in right censoring caused by the participant's death or withdrawal, to be referred to as the primary right censoring process. The primary-right-censoring process could be informative with respect to the response parameters. In our development, staggered entry and other missing value processes, if incorporated, are assumed to be non-informative and independent of the primary right censoring process.

Under a linear random effects model, we propose a model which can depend both on the individual's initial value and slope. A likelihood ratio test for informativeness and maximum likelihood estimates for the response parameters and

the primary right censoring process coefficients are derived under a probit model for the probability of primary right censoring.

The right censoring is considered to be non-informative with respect to the response parameters if the likelihood function can be factored into two independent parts, one corresponding to the response parameter and the other corresponding to censoring parameters.

We show that when the primary right censoring is non-informative, the maximum likelihood estimates for the average linear regression coefficients of the response are weighted linear combinations of the simple least squares estimates. In the case of complete observations at identical time points among all individuals, these estimates are just the unweighted averages of the individual simple least squares estimates.

The proposed method is applied to data on patients with PiZ phenotype, gathered by the workshop on Natural History of PiZ Emphysema (1983). To illustrate the effect of informative right censoring, maximum likelihood and the weighted and unweighted least squares procedures are applied to a set of simulated clinical trials with primary right censoring generated from a non-informative probit process and then to another set of simulated trials with primary right censoring generated from an informative probit process. Mean squared error and power comparisons are made among the different statistical procedures and between these two sets.

## SECTION 2 : Linear Random Effects and Informative Right Censoring

We assume that the participants of a longitudinal study are divided into two treatment groups of sample sizes  $n_k$ , for  $k = 1, 2$ . The combined sample size is  $n = n_1 + n_2$ . Let there be  $J$  identical mortality (and withdrawal status) follow-up

time points,  $t_j$ , with  $t_1 = 0$  and  $t_J =$  the length of the study. Each participant can have at most  $R$  measurements of the response during the study. The measurement time need not be identical among individuals. Let  $v_i =$  total number of measurements made for the  $i$ th individual. Let  $Y_{iv}$  and  $t_{iv}$  be the  $v$ th response and the corresponding measurement time for the  $i$ th participant in the combined sample for  $v=1,2,\dots,v_i$  and  $i=1,2,\dots,n$ . With  $t_{i1}=0$ , let  $t_{iv_i} \leq t_j$  if death, withdrawal, or right censoring due to staggered entry occurred for the  $i$ th participant between time  $t_j$  and  $t_{j+1}$ ; otherwise  $t_{iv_i} = t_j$  if the  $i$ th participant was not right censored and  $t_{iv_i} = t_{iR} = t_j$  if the  $i$ th participant had complete observations.

It is assumed that the serial measurements of the primary variable follow a linear function of time. Let  $\beta_i^t = (\beta_{i1}, \beta_{i2})^t$  be the unobservable vector representing the true initial value and slope of the primary variable for the  $i$ th individual in the combined sample. For  $i \in k$  and  $k = 1, 2$ :

$$Y_i = X_i \beta_i + \epsilon_i, \text{ where } Y_i^t = (Y_{i1}, \dots, Y_{iv_i})^t \quad (2.1)$$

$$\beta_i \sim N(B_k, \Sigma_\beta) \text{ and } \epsilon_i \sim N(0, \sigma_\epsilon^2 I);$$

$$X_i = \begin{bmatrix} 1 & \dots & 1 \\ t_{i1} & \dots & t_{iv_i} \end{bmatrix}, \quad \Sigma_\beta = \begin{bmatrix} \sigma_{\beta_1}^2 & \sigma_{\beta_1\beta_2} \\ \sigma_{\beta_1\beta_2} & \sigma_{\beta_2}^2 \end{bmatrix}, \quad B_K^t = (B_{k1}, B_{k2}).$$

The notation,  $i \in k$ , is used to denote that the  $i$ th participant in the combined sample belonged to the  $k$ th treatment group.

We further suppose that the probability of being primarily right censored due to death or withdrawal during a specified time interval  $(0, t_j)$ , given  $\beta_i$ , is  $M(\alpha^t \beta_i, t_j)$ . Here  $\alpha^t = (\alpha_1, \alpha_2)$  is the vector of "regression parameters" relating

this probability to the primary variables  $\beta_{\sim i}$ . Examples of logical choices for M are proportional hazards regression (Cox 1972), logistic regression (Walker and Duncan (1967)) and probit regression (Halperin, Wu and Gordon (1979)). For instance, under probit regression  $M(\alpha^t \beta_{\sim i}, t_j) = \Phi(\alpha^t \beta_{\sim i} + \alpha_{0j})$ , where  $\Phi$  is the cumulative probability of a standard normal variate.

Since for each  $\beta_{\sim i}$ , (i) the simple least squares estimates  $\hat{\beta}_{\sim i} = (X_i^t X_i)^{-1} (X_i^t Y_i)$  (ii) censoring time and (iii) survival time are sufficient statistics for  $\beta_{\sim i}$ , it suffices to consider the joint distribution of  $\hat{\beta}_{\sim i}$ ,  $\beta_{\sim i}$  and the primary right censoring process. The marginal likelihood for  $B_{\sim k}$  and  $\alpha$  for the  $i$ th individual can be expressed as

$$L_i = D \int \phi_2(\hat{\beta}_{\sim i}, \beta_{\sim i}, C_{1i}) \phi_2(\beta_{\sim i}, B_{\sim k}, \Sigma) \prod_{j=2}^J \{(1-M_{j-1})^{C(i,j-1)} (M_j - M_{j-1})^{Z(i,j-1)} [1-M_j]^{(1-m(i))}\} d\beta_{\sim i} \quad (2.2)$$

where  $M_j = M(\alpha^t \beta_{\sim i}, t_j)$  for  $i \in k, k=1,2, j=1, \dots, J; M_1 = 0$ . Here  $C(i,j)$  is the indicator function that the  $i$ th individual was censored in the  $j$ th interval because of staggered entry,  $Z(i,j)$  is the indicator function that death or withdrawal occurred in the  $j$ th interval for the  $i$ th individual,  $D$  is constant with respect to  $\beta_{\sim i}, \alpha$  and  $B_{\sim k}$  and

$$C_{1i} = \sigma_e^2 (X_i^t X_i)^{-1}, \quad m(i) = \sum_j \{C(i,j) + Z(i,j)\} .$$

The notation  $\phi_2(Y, \beta, \Sigma)$  represents the bivariate normal density with mean vector  $\beta$  and covariance matrix  $\Sigma$ . On the right hand-side of equation (2.2), under the integration sign, the first factor represents the conditional probability

distribution of  $\hat{\beta}_{\sim i}$  given  $\beta_{\sim i}$ ,  $C(i,j)$  and  $Z(i,j)$  for  $j=1, \dots, J-1$ . The second factor is the probability distribution of  $\beta_{\sim i}$ . The third factor of products corresponds to the conditional probabilities that the  $i$ th participant survived the  $(j-1)$ th time point and then was censored by staggered entry or death (or withdrawal) between the  $(j-1)$ th and the  $j$ th time points, respectively for  $j=2, \dots, J$  given  $\beta_{\sim i}$ . The last factor represents the conditional probability that the  $i$ th participant survived the entire study, given  $\beta_{\sim i}$ . Therefore the product of these four factors is proportional to the joint distribution of  $\hat{\beta}_{\sim i}$ ,  $\beta_{\sim i}$ ,  $Z(i,j)$  and  $C(i,j)$ , because the staggered entry process and the missing value process are assumed to be non-informative and independent of the primary right censoring process. Hence, integration with respect to the vector  $\beta_{\sim i}$  provides the marginal likelihood of  $\hat{\beta}_{\sim i}$ ,  $Z(i,j)$  and  $C(i,j)$  with respect to  $B_{\sim k}$  and  $\alpha$ .

This notation can be used for those measured only at baseline. Equating all elements except the (1,1)th of  $C_{1i}$  and  $\Sigma_{\beta}$  to zero and letting the (1,1)th element of  $C_{1i}$  equal  $\sigma_{\epsilon}^2$  and  $\hat{\beta}_{i2} = \beta_{i2} = 0$ . The marginal likelihood for all  $n$  individuals is the product of the individual likelihoods.

Joint estimation of the parameters depends on the ability to evaluate (2.2) and its derivatives. For this section we assume that  $\Sigma_{\beta}$  and  $\sigma_{\epsilon}^2$  are known. The more realistic case will be discussed in the next two sections. In principle (2.2) can be evaluated by numerical integration. When the primary right censoring process is a probit model, (2.2) can be evaluated explicitly: for  $i \in k$  and  $k = 1, 2$ ,

$$\ln\{L_1\} = \ln\{D\} + \ln\{A_1\} - 0.5(\hat{\beta}_{\sim i} - B_{\sim k})^t C_{2i}^{-1} (\hat{\beta}_{\sim i} - B_{\sim k}) + T_i, \quad (2.3)$$

where

$$A_i = (2\pi |C_{2i}|^{\frac{1}{2}})^{-1}, \quad C_{2i} = C_{1i} + \Sigma_{\beta}.$$

$$T_i = \sum_{j=2}^J \{C(i,j-1) \ln[1-\Phi(U_{ij-1})] + Z(i,j-1) \ln[\Phi(U_{ij}) - \Phi(U_{ij-1})]\} \\ + (1 - \sum_j Z(i,j) - \sum_j C(i,j)) \ln[1-\Phi(U_{iJ})].$$

$$U_{ij} = (\alpha_{0j} + d_{ik}^t C_{3i} \alpha) (1 + \alpha^t C_{3i} \alpha)^{-\frac{1}{2}}, \\ d_{ik} = C_{1i}^{-1} \hat{\beta}_i + \Sigma_{\beta}^{-1} B_{k}, \quad C_{3i} = (C_{1i}^{-1} + \Sigma_{\beta}^{-1})^{-1}.$$

When there are right censored or missing observations,  $C_{3i}$  will differ among individuals. Hence, the primary right censoring contribution to the likelihood,  $T_i$ , is in the form of a non-linear probit model. Maximum likelihood estimation of the parameters can be made in principle provided that the number of time intervals is small. Otherwise, some constraints could be imposed on the  $\alpha_0$ 's to reduce the number of parameters.

Likelihood ratio tests for the hypothesis ( $H_0: \alpha_1 = \alpha_2 = 0$ ) versus ( $H_1: \alpha_2 = 0$  and  $\alpha_1 \neq 0$ ) and the hypothesis  $H_1$  versus ( $H_2: \alpha_1 \neq 0$  and  $\alpha_2 \neq 0$ ) can be conducted. When  $H_0$  is true, the primary right censoring will be non-informative with respect to  $B_{k}$  for  $k=1$  and  $2$ . However, when  $H_1$  is true, it can be shown that the coefficient of  $B_{2k}$  in  $U_{ij}$  of (2.3) is non-zero even when  $\sigma_{\beta_1\beta_2} = 0$ . Hence the primary right censoring will be informative with respect to  $B_{k2}$  for  $k=1,2$ . When  $H_1$  is true, and  $\sigma_{\beta_1}^2 = \sigma_{\beta_1\beta_2} = 0$ , the primary right censoring is non-informative with respect to  $B_{k2}$ .

**SECTION 3 : Estimation and Testing for Noninformative Censoring**

When  $H_0$  is true, the maximum likelihood estimate of  $B_{k}$  is



$$\hat{\tilde{B}}_{GL,k} = \left[ \sum_{i \in k} C_{2i}^{-1} \right]^{-1} \sum_{i \in k} (C_{2i}^{-1} \hat{\tilde{\beta}}_i), \quad (3.1)$$

the generalized least squares estimate (GLSE). When all individuals have complete observations measured at identical time points,  $C_{2i}$  will be the same among individuals, in which case (3.1) reduces to

$$\hat{\tilde{B}}_{UW,k} = \sum_{i \in k} \hat{\tilde{\beta}}_i / n_k,$$

the unweighted least squares estimate (UWLE). The covariance matrices are,

$$C_{GL,k} = \left[ \sum_{i \in k} C_{2i}^{-1} \right]^{-1} \quad \text{and} \quad C_{UW,k} = \left[ \sum_{i \in k} C_{2i} \right] n_k^{-2}. \quad (3.2)$$

When  $\Sigma_{\beta}$  and  $\sigma_{\epsilon}^2$  are unknown, the following unbiased estimators can be substituted,

$$\hat{\sigma}_{\epsilon}^2 = s_{\epsilon}^2 / \left[ \sum_{i=1}^n (v_i - 2) \right], \quad \hat{\Sigma}_{\beta} = s_{\beta} / (n-1) = \sum_{i=1}^n C_{1i} / n, \quad (3.3)$$

$$s_{\beta} = \sum_{k=1}^2 \sum_{i \in k} (\hat{\tilde{\beta}}_i - \hat{\tilde{B}}_{UW,k}) (\hat{\tilde{\beta}}_i - \hat{\tilde{B}}_{UW,k})^t \quad \text{and} \quad s_{\epsilon}^2 = \sum_{i=1}^n (Y_{i1}^t Y_{i1} - Y_{i1}^t X_{i1} \hat{\tilde{\beta}}_i).$$

However,  $\hat{\Sigma}_{\beta}$  has the disadvantage that it is not necessarily positive definite. The procedure given by Bock and Peterson (1975) for constructing an estimate that is at least semi-definite will be used.

When the goal of a study is to compare differences in rate of change between two groups, we wish to test the null hypothesis,  $H_N : B_{12} = B_{22}$  against the alternative hypothesis,  $H_A : B_{12} < B_{22}$ . The test statistic is of the form,

$$(\hat{B}_{12} - \hat{B}_{22}) / (\hat{\sigma}_{B_{12}}^2 + \hat{\sigma}_{B_{22}}^2)^{\frac{1}{2}},$$

with  $\hat{B}_{k2} = \hat{B}_{GL,k2}$  and  $\hat{B}_{UW,k2}$  respectively. For shifted alternatives, sample

size, power and significance level of the test can be related according to the approximate formula,

$$(\hat{\sigma}_{B_{12}}^2 + \hat{\sigma}_{B_{22}}^2)(Z_\alpha + Z_\beta)^2 = \Delta^2, \quad (3.4)$$

where  $\Delta$  is the difference in expected rates of change we wish to detect,  $\alpha$  and  $\beta$  are the Type I and Type II errors of the test with  $Z_\alpha$  and  $Z_\beta$  the unit normal deviates corresponding to  $\alpha$  and  $\beta$ .

REMARKS : We have by assumption that an individual's coefficient estimate is unbiased, i.e.,

$$E\left[\hat{\beta}_i \mid \beta_i\right] = \beta_i.$$

Thus the unweighted least squares estimate is unbiased for  $B_k$ . There are two cases. When the primary right censoring is non-informative, the distribution of  $C_{2i}$  in (3.1) does not depend on  $\beta_i$ , so that the GLSE and UWLE are both consistent and unbiased estimators of  $B_k$ , although of course the UWLE is less efficient.

Furthermore, the relative differences between the variances and hence the required sample sizes of the UWLE and the GLSE for the slope or initial value are a function of  $(\sigma_\epsilon^2/\sigma_{\beta_2}^2)$  or  $(\sigma_\epsilon^2/\sigma_{\beta_1}^2)$  respectively. When the primary right censoring process is informative, the unweighted least squares estimate is still unbiased, although the GLSE is not because  $C_{2i}$  and  $\beta_i$  are dependent.

#### SECTION 4 : Examples and Simulations

This paper was motivated by design and analysis problems encountered in many clinical trials concerning lung diseases, e.g., the Intermittent positive pressure breathing trial (IPPB 1983) for chronic pulmonary diseases. One

specific example was the feasibility study of an anti-proteolytic replacement therapy trial among individuals with PiZ phenotype, conducted by the Workshop on the Natural History of PiZ emphysema. The association between severe  $\alpha_1$ -antitrypsin deficiency and lung diseases, particularly pulmonary emphysema, has been observed since the early 1960's (Laurell and Eriksson (1963)). Individuals with PiZ phenotype tend to develop severe  $\alpha_1$  antitrypsin deficiency and hence pulmonary emphysema and more rapid decline in lung function. The planned trial was designed to detect differences in rates of decline of a one second forced expiratory volume ( $FEV_1$ ) between a control and a therapeutic group. Retrospective data on PiZ individuals were gathered from the ten participating institutions (see Workshop on Natural History of PiZ Emphysema (1983)) to provide crude estimates of parameter values required for sample size calculations.

#### 4.1 : Estimation and Testing

A Fortran program was developed for estimation when there is no staggered entry. The method of pseudo maximum likelihood estimation (PMLE - see Gong and Samaniego 1981) was used. Estimates of  $\sigma_e^2$  and  $\Sigma_\beta$  were made according to (3.3) and the Bock and Peterson (1975) procedure and substituted into (2.3), which was then maximized by the Newton-Raphson method. The algorithm first calculates the simple least squares intercept and slope for each individual and estimates  $\sigma_e^2$  and  $\Sigma_\beta$ . The UWLE of  $B_{\sim k}$  is used as initial value for  $B_{\sim k}$  in calculating  $d_{\sim ik}$  and  $C_{3i}$  for each individual according to (2.3). Partial derivatives of the log likelihood for the Newton-Raphson iterative procedure are then calculated using initial values for  $\alpha_1, \alpha_2, \alpha_{02}, \dots, \alpha_{0J}$ . Formulae for these partial derivatives are presented in the Appendix. Note that the initial values for the  $\alpha$ 's can be chosen arbitrarily with the constraint  $\alpha_{02} < \alpha_{03} < \dots < \alpha_{0J}$ .

This algorithm was applied to the PiZ emphysema data. Among the data

gathered for 294 PiZ individuals by the ten U.S. institutions, initial and follow-up FEV<sub>1</sub> values (with the initial and the last measurements at least 6 months apart) were available on 117 individuals. The number of FEV<sub>1</sub> measurements ranged from 2 to 12 (mean number of measurements = 3.8). The duration between the initial and the last measurement ranged from 6 to 227 months (mean duration = 52 months). Since the proposed trial duration was between three and six years, an analysis, corresponding to a three year follow-up study, was first made using the initial and all follow-up FEV<sub>1</sub> measurements made within three years of the initial measurement. Since many did not have reported follow-up FEV<sub>1</sub>'s within three years of the initial measurement, only 81 individuals with 8 deaths were included in this analysis. A second analysis, corresponding to a six year follow-up study, was also made among those with a minimum follow-up of six years or a reported death within the first six years. Follow-up FEV<sub>1</sub>'s within six years of the initial measurement were used. This analysis included 65 individuals with 19 deaths. Because of the small number of deaths, mortality follow-ups were grouped into two equal length intervals for both analyses. The average number of FEV<sub>1</sub> measurements were 2.9 and 3.6 and the average duration between the initial and the last FEV<sub>1</sub>'s were 28 and 48 months for the 3 and 6 year follow-ups, respectively. Those individuals with only one FEV<sub>1</sub> measurement were not included in these analyses. This has the effect of causing a slight bias in the unweighted least squares analysis and a slight loss of efficiency in the informative censoring analysis.

\*\*\*\*\* Table 1 about here \*\*\*\*\*

The purpose of these analyses was to test for informativeness of the right censoring caused by participant's death with respect to FEV<sub>1</sub> initial value and slope, obtained from 3 and 6 year follow-ups, respectively; and to derive crude estimates of the primary right censoring coefficients. The initial values used for the iterative procedure were  $\alpha_{02} = -1.35$ ,  $\alpha_{03} = -0.90$  and  $\alpha_1 = \alpha_2 = 0$ . The

algorithm converged after 12 and 10 iterations for the first and second analyses, respectively. The results are presented in Table 1. The estimated probit right censoring coefficients for FEV<sub>1</sub> initial value and slope ( $\alpha_1$  and  $\alpha_2$ ) were -3.8, -11.3 and -4.6, -13.8 for the two analyses, respectively. Likelihood ratio tests indicated that the coefficients for FEV<sub>1</sub> initial value ( $\alpha_1$ ) were statistically significantly different from zero in both analyses. Although the chi-squared statistic (with one degree of freedom) of 2.8 for the slope coefficient ( $\alpha_2$ ) of the first analysis was not statistically significant at a 5% level, the chi-squared statistic of 7.1 for the slope coefficient of the second analysis was statistically significant. The significance of the initial value coefficients and the large negative slope coefficients obtained from both analyses, the significance of the slope coefficient from the second analysis indicated that the right censoring by participants' deaths could be informative with respect to both FEV<sub>1</sub> initial value and slope.

Survival probability distributions estimated by the product limit method (Kaplan and Meier, 1958), for the entire data set of 294 individuals, for those individuals included in the first and second analyses of Table 1, respectively; and for the 117 individuals with two or more FEV<sub>1</sub> measurements are displayed in Figure 1. Since these data were collected retrospectively, mortality follow-ups were not as complete and rigorous as one would like them to be for the proposed prospective study. Hence, survival probabilities in Figure 1 could be optimistic.

The estimates we have proposed are of course sensitive to model misspecification. When using the estimation and test procedures derived under the probit model, goodness-of-fit to the data should be checked. One approach is to note that the estimated probability for the  $i$ th individual being primarily right censored in the  $j$ th time interval, for given  $\hat{\beta}_i$ , is  $\hat{P}_{ij} = \Phi(\hat{U}_{ij+1}) - \Phi(\hat{U}_{ij})$ , for  $j=1, \dots, J-1$ ; where  $\hat{U}_{ij}$  is  $U_{ij}$  of (2.3) with  $\alpha$  and  $\alpha_{0j}$  and the

expected intercept and slope for the primary variable being replaced by their MLE's. Therefore, group the  $\Phi(\hat{U}_{ij})$  into groups and compute for each group,  $E_j = \sum_i \hat{P}_{ij}$ . Then compare  $E_j$  with the observed deaths and dropouts between the  $j$ th and  $(j+1)$ th time points.

For the PiZ six year follow-up data of Table 1, the observed number of deaths among those whose estimated probabilities of death in six years were above the 85th percentile, between the 70th and 85th percentiles and below the 70th percentile (for the entire 65 individuals) were 4, 2, 2 and 3, 4, 4 for the first and second three year intervals, respectively. The corresponding expected numbers of death were 4.47, 1.82, 0.85 and 2.98, 3.70, 3.88 for the two time intervals, respectively. Hence, the probit model seems to fit the data reasonably well.

Graphical comparison of the actual versus expected cumulative numbers of death by the estimated probability of death in six years for the six year follow-up data is displayed in Figure 2A. Comparisons of the actual versus expected cumulative numbers of death in the first and second three year intervals by the estimated probability of death in the corresponding time intervals, for the same six year follow-up data, are shown in Figure 2B. The overall fits of the data from both figures were again reasonably good.

#### 4.2 : The effect of informative censoring

The UWLE, GLSE, and the PMLE were compared in simulated experiments based on the model (2.2) with the following primary right censoring processes: (1) probit non-informative censoring with  $\alpha_1 = \alpha_2 = 0$ ; (2) probit informative censoring with coefficients  $\alpha_1 = -3.8$ ,  $\alpha_2 = 11.3$ ; (3) probit informative censoring with coefficients  $\alpha_1 = -4.6$ ,  $\alpha_2 = -13.8$ , corresponding to the two analyses of Section 4.1; and (4) probit informative censoring with  $\alpha_1 = -3.8$  and

$\alpha_2 = 0$ . Similar to the IPPB trial (1983), the study duration was assumed to be three years with four  $FEV_1$  measurements per year. The expected  $FEV_1$  slope and initial value in the control group and the within and between individual variances used were all estimated from the PiZ data. A 50% reduction in  $FEV_1$  rate of decline was assumed in the treatment group. Equal sample sizes of 100 each were generated for the two groups. In the IPPB trial, similar to the proposed trial, patients were required to have their  $FEV_1$  values less than 65% predicted at entry and the comparison of  $FEV_1$  annual rates of decline between two randomized treatment groups was the primary objective of the trial. The primary right censoring rate for the IPPB trial was more than 12% per year. For these illustrations the probability of primary right censoring was assumed to be 16% each year for all individuals under the non-informative right censoring process. When the informative probit model was used, this probability was assumed to be 16% for an individual whose initial value and slope were equal to the expected values for the control group. It was further assumed that there is no correlation between the slope and the initial value ( $\sigma_{\beta_1\beta_2} = 0$ ). The decision value used for rejecting the null hypothesis of no difference was  $(\hat{B}_{12} - B_{22}) / (\hat{\sigma}_{B_{12}}^2 + \hat{\sigma}_{B_{22}}^2)^{\frac{1}{2}} < -1.645$ . Normal random numbers were generated by the IMSL routine GNPM. The experiments were repeated 600 times.

The results in Table 2 indicate that the UWLE procedure remained relatively unbiased in estimating the mean  $FEV_1$  slope for each group and the between group difference in slopes. However, the PMLF clearly had much smaller mean squared errors in estimating the individual group mean slopes and the between group differences, and much higher statistical power in detecting the between group differences in all four censoring processes considered. The GLSE, although most efficient under non-informative censoring, resulted in large under-estimations of individual group mean  $FEV_1$  rates of decline (24-46%), under the two probit

informative censoring processes with non-zero coefficients for  $FEV_1$  slope. The under-estimation for the between group differences were much smaller (11-13%), because under the shifted alternative of equation (3.4), the biases in the two group estimates tend to cancel with each other. The GLSE had smaller mean squared errors in estimating the between group differences and higher statistical power to detect these differences than the UWLE in all four censoring processes considered. Compared to the PMLE, under the two probit informative censoring processes with non-zero slope coefficients, the GLSE had much larger mean squared errors (39-69%) in estimating the individual group mean slopes and (14-22%) in estimating the between group differences; and lower statistical power (10-15%) in detecting the between group differences. The expected power for the proposed study, calculated according to (3.4) using the assumed parameter values, was 0.85 for the GLSE under the non-informative censoring process. The simulated power for the GLSE under non-informative censoring, using the estimated within and between individual variances according to (3.3) and the Bock and Peterson (1975) procedure for constructing covariance matrices that were at least semi-definite was 0.81, and not very different from the expected power. Using the PMLE when the censoring process was non-informative or when the probit censoring slope coefficient was zero could result in larger mean squared errors than the GLSE, in estimating the group slopes. The simulated significance levels were not much different from the expected 5% level for all procedures in Table 2.

\*\*\*\*\* Table 2 about here \*\*\*\*\*

#### SECTION 5 : Discussion

The probit model used in Sections 2 and 4 is not necessarily meant to be biologically valid for describing the underlying right censoring process.



Indeed, the choice of the probit was made primarily on computational grounds, and because logistic and probit regressions give similar estimates of event probabilities (Halperin, et al. (1979)).

When using the estimation and test procedures derived under the probit model, goodness-of-fit to the data should be checked as suggested in Section 4.1. However the distribution of the chi-squared goodness-of-fit test statistic for this situation cannot be obtained from a straightforward application of the usual theory because (i) parameter estimates are determined using likelihood functions for ungrouped data; and (ii) random cell boundaries. Moore (1971) and Moore and Spruill (1975) derived large sample distribution of the usual chi-squared goodness-of-fit statistics under these two problems. Their basic result is that under appropriate regularity conditions the large sample distribution of the goodness-of-fit statistic is that of a central chi-squared with the usual reduction in degree of freedom due to estimated parameters plus a weighted sum of independent chi-squared random variables each with one degree of freedom. Application of their result to this problem is under investigation.

Although the estimation and test procedures of Sections 2 and 4 were developed for  $k=2$  groups, they could be extended easily to the case of  $k > 2$  groups. To test for equality or linear trend among the expected slopes of the  $k > 2$  groups, the likelihood ratio chi-squared statistic could be used.

The standard errors provided in Table 1 for estimates based on the probit right censoring model and those used in computing the test statistics for the PMLE in Table 2 were estimated from the sample information matrix based on the pseudo maximum likelihood, by assuming that the estimated between and within individual error variances were the true values, rather than based on the maximum likelihood. The bootstrap (Efron, 1979) could be used to improve these estimates. Alternatively, the maximum likelihood procedure, treating  $\sigma_\epsilon^2$ ,  $\sigma_{\beta_1}^2$  and  $\sigma_{\beta_2}^2$  as additional parameters, could also be used.

ACKNOWLEDGEMENT

The authors wish to thank the referees for many helpful suggestions.

References

- Bock, R.D. and Petersen, A.C. (1975). A multivariate correction for attenuation. Biometrika 62, 673-678.
- Carroll, R.J., Spiegelman, C.H., Lan G., Bailey, K. and Abbott, R.D. (1984). On errors-in-variables for binary regression models. Biometrika 71, 19-23.
- Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 34, 87-220.
- Efron, B. (1979). Bootstrap methods. Another look at the jackknife. Annals of Statistics 7, 1-26.
- Fearn, T. (1975). A Bayesian approach to growth curves (1975). Biometrika 62, 89-100.
- Gong, G. and Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: theory and application. Annals of Statistics 9, 861-869.
- Halperin, M., Wu, M. and Gordon, T. (1979). Genesis and Interpretation of differences in distributions of baseline characteristics between cases and non-cases in cohort studies. Journal of Chronic Diseases 32, 483-491.
- Intermittent Positive Pressure Breathing Trial Group (1983). Intermittent positive pressure breathing therapy of chronic obstructive pulmonary disease. Annals of Internal Medicine 99, 615-620.
- Kaplan, L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.
- Kannel, W.B. and Gordon, T. (1971). The Framingham Study Section 27. U.S. Department of Health, Education and Welfare, Public Health Service, National Institutes of Health.
- Koziol, J.A., Maxwell, D.A., Matsuro Fukushima, M.E.M. and Yosef, H.P. (1981). A distribution-free test for tumor-growth curve analyses with application to an animal immunotherapy experiment. Biometrics 37, 383-390.
- Laurell, B. and Eriksson, S. (1963). The electrophoretic  $\alpha_1$  - 3 lobulin pattern of serum in  $\alpha_1$  - trypsin deficiency. Scandinavian Journal of Clinical and Laboratory Investigation 5, 132.
- Laird, N.M. and Ware, H.H. (1982). Random effect models for longitudinal data.

Biometrics 38, 963-974.

Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika 52, 447-458.

Schlesselman, J.J. (1973). Planning a longitudinal study, II frequency of measurements and study duration. Journal of Chronic Diseases 26, 561-570.

Walker, R.H. and Duncan, D.B. (1967). Estimation of the probability of an event as a function of several independent variables. Biometrika 54, 167-179.

Workshop on Natural History of PiZ Emphysema (1983). The natural history of PiZ emphysema. Proteases and Inhibitors in the Lungs. American Review of Respiratory Diseases, 127 (Supplement) 543-545.

APPENDIX

Let

$$C_{2i} = \begin{bmatrix} \sigma_{2i1}^2 & \sigma_{2i12} \\ \sigma_{2i12} & \sigma_{2i2}^2 \end{bmatrix}, \quad C_{3i} = \begin{bmatrix} \sigma_{3i1}^2 & \sigma_{3i12} \\ \sigma_{3i12} & \sigma_{3i2}^2 \end{bmatrix},$$

$d_{ik}^t = (d_{ik1}, d_{ik2})$ ,  $\rho_i = \sigma_{2i12} / (\sigma_{2i1} \cdot \sigma_{2i2})$ ,  $\varphi_{ij} = \varphi(U_{ij})$ ,  $\phi_{ij} = \phi(U_{ij})$ , for  $j=2, \dots, J$  and  $\varphi_{i1} = \phi_{i1} = 0$ . From (2.3),  $U_{ij} = (\alpha_{0j} + \alpha_1 d_{ik1} + \alpha_2 d_{ik2}) / D^{\frac{1}{2}}$ , for  $j=2, \dots, J$ ; where  $D = (1 + \sigma_{3i1}^2 \alpha_1^2 + \sigma_{3i12}^2 \alpha_1 \alpha_2 + \sigma_{3i2}^2 \alpha_2^2)$ . The parameter to be estimated is  $\theta^t = (\theta_1, \dots, \theta_{J+5}) = (B_{11}, B_{12}, B_{21}, B_{22}, \alpha_1, \alpha_2, \alpha_{02}, \dots, \alpha_{0J})$ . The partial derivatives of the log likelihood (2.3) with respect to these parameters are as follows:

$$\frac{\partial \ln L_i}{\partial B_{k\ell}} = \begin{cases} \{(\hat{\beta}_{i\ell} - B_{k\ell}) / \sigma_{2i2}^2 - \rho_i (\hat{\beta}_{im} - B_{km}) / (\sigma_{2i1} \sigma_{2i2})\} / \\ (1 - \rho_i^2) + [\partial T_i / \partial \beta_{k\ell}], \text{ for } i \in k, k=1,2, \ell=1,2, m=3-\ell \\ 0 \text{ otherwise} \end{cases}$$

$$\frac{\partial \ln L_i}{\partial B_{k_1 \ell} \partial B_{k_2 m}} = \begin{cases} (-1)^{(\ell-m)} \rho_i^{|\ell-m|} / [\sigma_{2i\ell} \sigma_{2im} (1-\rho_i^2)] + \partial T_i / \partial B_{k_1 \ell} \partial B_{k_2 m}, \\ \text{for } i \in k, k_1=k_2=k, \ell=1,2 \text{ and } m=1,2; \\ 0 \text{ otherwise;} \end{cases}$$

$$\partial \ln L_i / \partial \theta_\ell = \partial T_i / \partial \theta_\ell, \text{ for } \ell = 5, \dots, J+5;$$

$$\partial^2 \ln(L_i) / \partial \theta_\ell \partial \theta_m = \partial T_i / \partial \theta_\ell \partial \theta_m, \text{ for } \ell=5, \dots, J+5, m=1, \dots, J+5.$$

When there is no staggered entry we have,

$$\partial \tau_i / \partial \theta_\ell = \sum_{j=2}^J \{Z(i, j-1) [\varphi_{ij} (\partial U_{ij} / \partial \theta_\ell) - \varphi_{ij-1} (\partial U_{ij-1} / \partial \theta_\ell)] / (\phi_{ij} - \phi_{ij-1})\} - (1 - \sum_{j=2}^J Z(i, j-1)) \varphi_{iJ} (\partial U_{iJ} / \partial \theta_\ell) / (1 - \phi_{iJ}) ;$$

$$\begin{aligned} \partial^2 \tau_i / \partial \theta_\ell \partial \theta_m &= \sum_{j=2}^J \{Z(i, j-1) \{(\phi_{ij} - \phi_{ij-1}) [-U_{ij} \varphi_{ij} (\partial U_{ij} / \partial \theta_\ell) (\partial U_{ij} / \partial \theta_m) \\ &+ U_{ij-1} \varphi_{ij-1} (\partial U_{ij-1} / \partial \theta_\ell) (\partial U_{ij-1} / \partial \theta_m) \\ &+ \varphi_{ij} (\partial U_{ij} / \partial \theta_\ell \partial \theta_m) - \varphi_{ij-1} (\partial U_{ij-1} / \partial \theta_\ell \partial \theta_m)] \\ &- [\varphi_{ij} (\partial U_{ij} / \partial \theta_\ell) - \varphi_{ij-1} (\partial U_{ij-1} / \partial \theta_\ell)] \\ &[\varphi_{ij} (\partial U_{ij} / \partial \theta_m) - \varphi_{ij-1} (\partial U_{ij-1} / \partial \theta_m)]\} / \\ &(\phi_{ij} - \phi_{ij-1})^2\} - (1 - \sum_{j=2}^J Z(i, j-1)) [(1 - \phi_{iJ}) U_{iJ} \varphi_{iJ} (\partial U_{iJ} / \partial \theta_\ell) \\ &(\partial U_{iJ} / \partial \theta_m) + \varphi_{iJ} (\partial U_{iJ} / \partial \theta_\ell \partial \theta_m) + \varphi_{iJ}^2 (\partial U_{iJ} / \partial \theta_\ell) \\ &(\partial U_{iJ} / \partial \theta_m)] / (1 - \phi_{iJ})^2 . \end{aligned}$$

for  $\ell = 1, \dots, J+5$  and  $m = 1, \dots, J+5$ ;

with

$$\partial U_{ij} / \partial B_{k\ell} = \begin{cases} \alpha_1 (\partial d_{ik1} / \partial B_{k\ell}) + \alpha_2 (\partial d_{ik2} / \partial B_{k\ell}) / D^{\frac{1}{2}} & \text{for } i \in k, \ell = 1, 2 \\ 0 & \text{otherwise;} \end{cases}$$

$$\partial U_{ij} / \partial \alpha_\ell = [d_{ik\ell} D^{\frac{1}{2}} - U_{ij} C_\ell] / D, \text{ for } i \in k, \ell = 1, 2;$$

$$\partial U_{ij} / \partial \alpha_{0\ell} = \begin{cases} D^{-\frac{1}{2}} & \text{for } \ell = j \\ 0 & \text{otherwise;} \end{cases}$$

$$\partial U_{ij} / \partial B_{k\ell} \partial \alpha_m = [D^{\frac{1}{2}} (\partial d_{ik\ell} / \partial B_{k\ell}) - C_m (\partial U_{ij} / \partial B_{k\ell})] / D .$$

for  $i \in k$ ,  $\ell = 1,2$  and  $m = 1,2$ ;

$$\partial U_{ij} / \partial \alpha_\ell \partial \alpha_m = \{D[d_{ik\ell} C_m D^{-\frac{1}{2}} - a_{\ell m} U_{ij} - C_\ell (\partial U_{ij} / \partial \alpha_m)] - 2[D^{-\frac{1}{2}} d_{ik\ell} - C_\ell U_{ij}]C_m\} / D^2 ;$$

$$\partial U_{ij} / \partial \alpha_\ell \partial \alpha_{0j} = -C_\ell / D^{\frac{3}{2}} \quad \text{for } \ell = 1,2 ;$$

$$\partial U_{ij} / \partial B_{k\ell} \partial B_{km} = \partial U_{ij} / \partial B_{k\ell} \partial \alpha_{0j} = \partial U_{ij} / \partial \alpha_{0j_1} \partial \alpha_{0j_2} = 0 ,$$

for  $\ell = 1,2$ ,  $m = 1,2$ ,  $i \in k$ ,  $j_1 = 2, \dots, J$ ,  $j_2 = 2, \dots, J$  and

$j = 2, \dots, J$  ;

where  $C_\gamma = \sigma_{3i\gamma}^2 \alpha_\gamma + \sigma_{3i12} \alpha_{3-\gamma}$  for  $\gamma = 1,2$ ;

$$a_{\ell m} = \begin{cases} \sigma_{3i\ell}^2 & \text{for } \ell = m \\ \sigma_{3i12} & \text{for } \ell \neq m. \end{cases}$$

(11)

Table 1

Estimation for the Expected FEV<sub>1</sub> Slope, the Missing Value Coefficients  
and Likelihood Ratio Tests Statistics for the Coefficients

Estimates and Test Statistics	<u>Three Year Mortality</u>		<u>Six Year Mortality</u>	
	Three Year FEV <sub>1</sub> Follow-Up		Six Year FEV <sub>1</sub> Follow-Up	
<u>Estimated FEV<sub>1</sub> change/year</u>				
Unweighted	-0.093	(.0164)*	-0.078	(0.0138)*
Weighted	-0.090	(.0151)	-0.076	(0.0136)
Probit Informative Missing	-0.095	(.0152)	-0.085	(0.0133)
<u>Estimated Missing Value Coefficients</u>				
FEV <sub>1</sub> Initial Value	-3.80	(2.01)	-4.61	(1.70)
FEV <sub>1</sub> Slope	-11.30	(7.46)	-13.80	(6.76)
$\alpha_{02}$	-0.53	(1.10)	1.25	(0.93)
$\alpha_{03}$	0.42	(1.10)	2.42	(1.05)
<u>L.R. Test Statistics</u>				
Initial Value, H <sub>1</sub> vs. H <sub>0</sub> (X <sup>2</sup> 1)	11.02		26.97	
Slope, H <sub>2</sub> vs. H <sub>1</sub> (X <sup>2</sup> 1)	2.81		7.13	
No. at risk at baseline	81		65	
No. of Deaths	8		19	

\*Numbers in parentheses are estimated standard errors.

**Table 2**

Comparison of Simulated Results Among Different Procedures Under a Linear Random Effect Model with Non-Informative Versus Probit Informative Missing Values with Parameter Values Estimated from PiZ Emphysema Data\*

Statistical Procedures & Treatment Groups	Non-Informative Censoring $\alpha_1 = \alpha_2 = 0$		Informative Censoring						
			$\alpha_1 = -3.8$		$\alpha_1 = -4.6$		$\alpha_1 = 3.8$		
	$\alpha_2 = -11.3$		$\alpha_2 = -13.8$		$\alpha_2 = -0.0$				
	FEV <sub>1</sub>		FEV <sub>1</sub>		FEV <sub>1</sub>		FEV <sub>1</sub>		
	Slope	MSE	Slope	MSE	Slope	MSE	Slope	MSE	
<u>Control Group</u>									
UWLE	-89.3	465.4	-88.9	645.7	-88.3	719.4	-89.2	570.4	
GLE	-90.3	156.3	-68.3	634.9	-63.7	883.3	90.7	164.6	
PMLE	-89.4	200.6	-83.6	455.0	-81.2	597.8	-90.2	210.6	
<u>Treatment Group</u>									
UWLE	-46.5	442.3	-45.9	444.9	-46.2	489.8	-46.9	444.7	
GLE	-45.9	148.2	-28.2	423.1	-24.4	576.0	-46.3	150.7	
PMLE	-44.6	194.4	-39.7	288.4	-38.8	340.6	-44.3	162.9	
<u>Between Group Differences</u>									
UWLE	-42.8	935.8	-43.0	1059.3	-42.11	1210.9	-42.3	1052.0	
GLE	-44.5	361.8	-40.1	339.0	-39.3	377.8	-44.4	313.0	
PMLE	-44.8	267.4	-43.9	297.1	-42.4	309.0	-44.1	297.3	
<u>Simulated Power &amp; Significance Level</u>									
	Power	Signif	Power	Signif	Power	Signif	Power	Signif	
UWLE	0.45	0.05	0.39	0.05	0.36	0.05	0.40	0.05	
GLE(0.85)**	0.81	0.06	0.72	0.07	0.67	0.07	0.77	0.06	
PMLE	0.80	0.05	0.80	0.06	0.79	0.06	0.78	0.07	

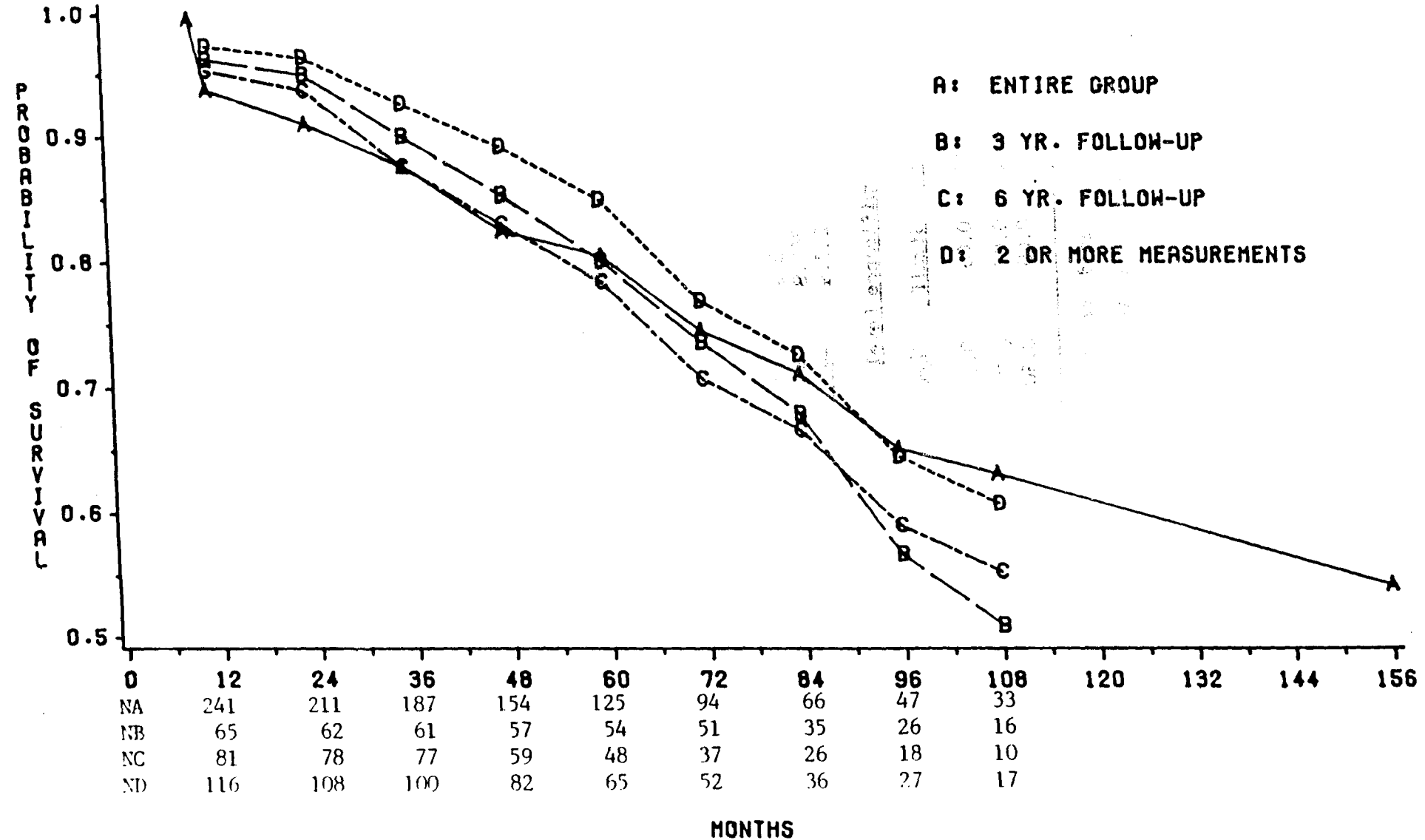
\*The parameter values used were: Measurement error standard deviation  $\sigma_e = 0.155$  FEV<sub>1</sub> initial value standard deviation  $\sigma_{\beta_1} = 0.39L$ , FEV<sub>1</sub> slope std. dev.  $\sigma_{\beta_2} = 0.091L/yr.$ ,  $\sigma_{\beta_2} = 0.091L/yr.$ ,  $\sigma_{\beta_1, \beta_2} = 0$ , expected FEV<sub>1</sub> initial value  $B_{11} = B_{21} = 0.96L$ , control and treatment group expected FEV<sub>1</sub> slopes  $B_{12} = -0.09L/yr.$  and  $B_{22} = -0.045/yr.$  For significance level,  $B_{21} = B_{22} = -0.09L/yr.$  The probability of missing, 16%/yr. and  $n_1 = n_2 = 100$ .

\*\*Expected power under non-informative missing process, calculated according to (3.4) using the actual parameter values.



FIGURE 1

SURVIVAL CURVES: PIZ EMPHYSEMA DATA



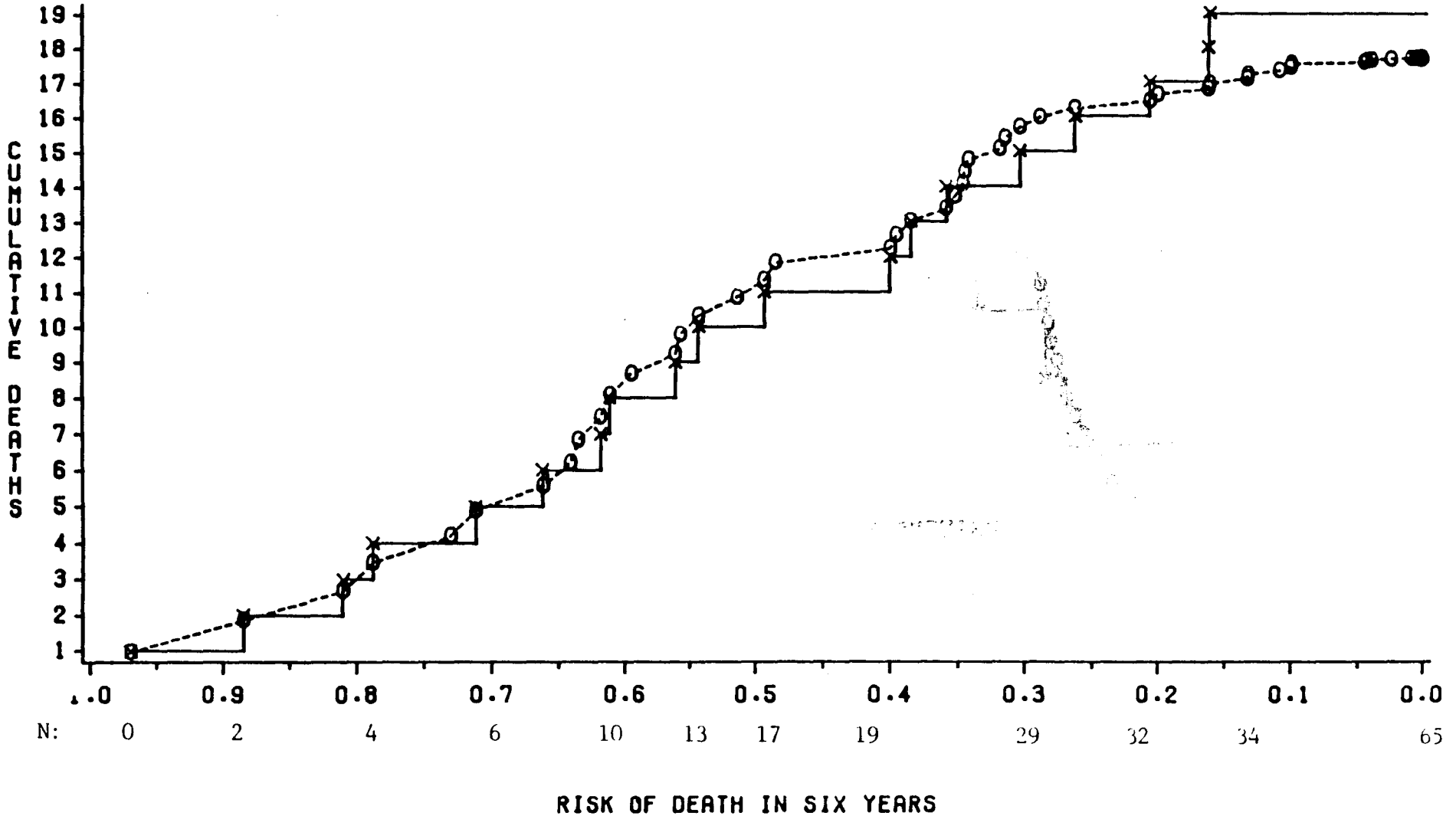
NA: NUMBER AT RISK FOR THE ENTIRE GROUP  
 NB: NUMBER AT RISK AMONG THOSE WITH THREE YEARS FOLLOW-UP  
 NC: NUMBER AT RISK AMONG THOSE WITH SIX YEARS FOLLOW-UP  
 ND: NUMBER AT RISK AMONG THOSE WITH TWO OR MORE FEV1 MEASUREMENTS

FIGURE 2A

# ACTUAL VS. EXPECTED DEATHS BY EXPECTED RISK OF DEATH IN SIX YEARS

X: ACTUAL NUMBER OF DEATHS AMONG THOSE WHOSE RISK OF DYING WERE AS INDICATED BY THE HORIZONTAL AXIS OR GREATER

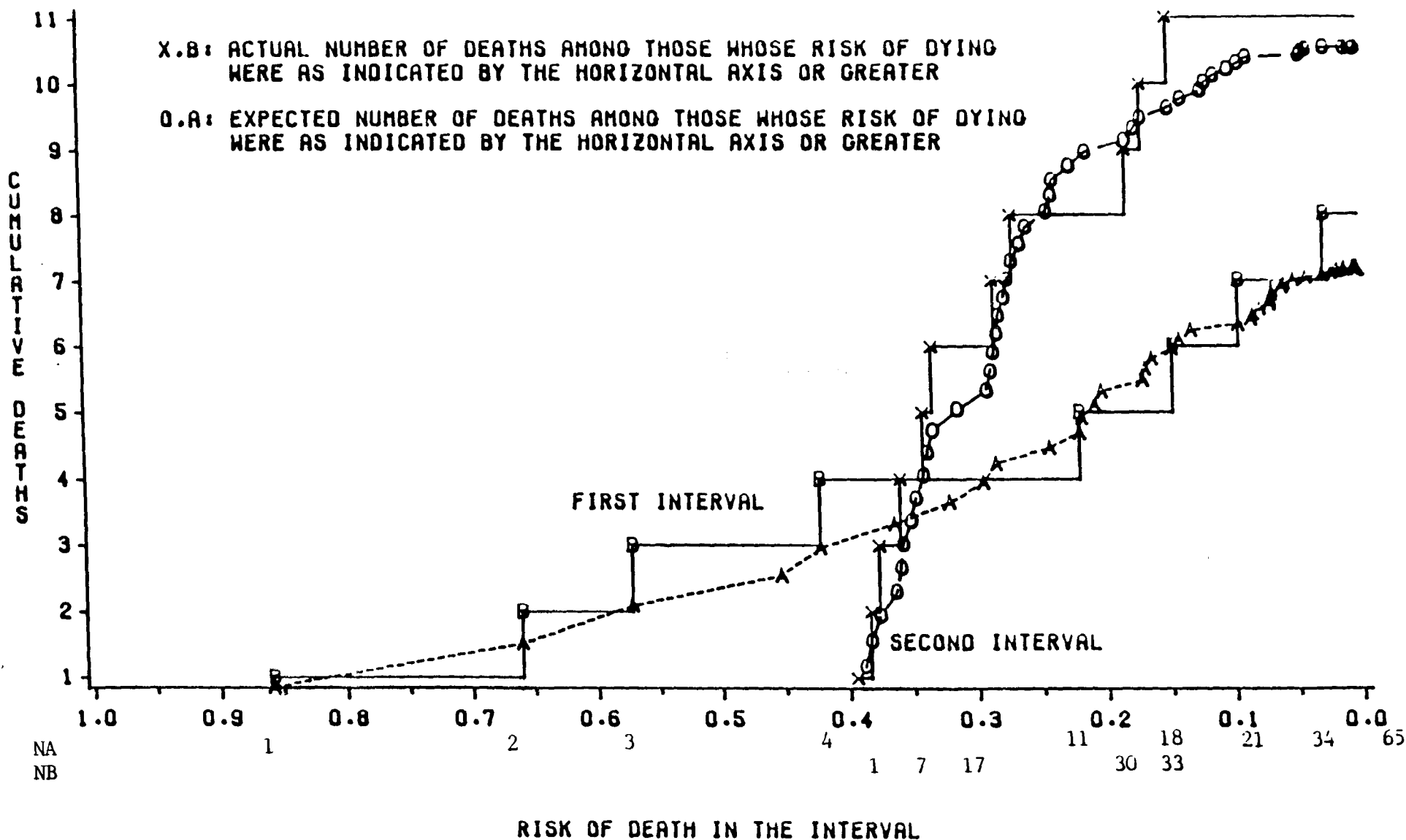
O: EXPECTED NUMBER OF DEATHS AMONG THOSE WHOSE RISK OF DYING WERE AS INDICATED BY THE HORIZONTAL AXIS OR GREATER



N: NUMBER OF INDIVIDUALS WHOSE RISK OF DYING WERE AS INDICATED BY THE HORIZONTAL AXIS OR GREATER

FIGURE 2B

ACTUAL VS. EXPECTED DEATHS  
 B. EXPECTED RISK OF DEATH IN EACH OF THE TWO THREE YEAR INTERVALS



NA: NUMBER OF INDIVIDUALS WHOSE RISK OF DYING IN THE FIRST INTERVAL WERE AS INDICATED BY THE HORIZONTAL AXIS OR GREATER  
 NB: NUMBER OF INDIVIDUALS WHOSE RISK OF DYING IN THE SECOND INTERVAL WERE AS INDICATED BY THE HORIZONTAL AXIS OR GREATER