FINAL REPORT

NIJ GRANT 89-IJ-CX-0010

# IMPROVING PREDICTIONS OF OFFENDER RECIDIVISM AND PATTERNS OF OFFENDER CRIME

National Bureau of Economic Research

Ann D. Witte, Principal Investigator

This report consists of three parts.

1. "Additional Results on Functional Form, Proportional Hazards Model, North Carolina Data"

2. "Results on Functional Form, Parametric (Lognormal) Models, North Carolina Data"

3. "Predictions from Proportional Hazards and Parametric Models, North Carolina Data"

# ADDITIONAL RESULTS ON FUNCTIONAL FORM
# PROPORTIONAL HAZARDS MODEL, NORTH CAROLINA DATA

Ching-Fan Chung, Peter Schmidt and Ann Witte

Report 1 of results under NIJ grant 89-IJ-CX-0010

## 1. Introduction

One of the tasks of grant 89-IJ-CX-0010 is to investigate appropriate ways to enter explanatory variables into models for the time until recidivism. This work is an extension of the analyses performed under our previous grant (84-IJ-CX-0021) and reported in P. Schmidt and A.D. Witte, Predicting Recidivism Using Survival Models, Springer-Verglag, 1989. In that work we tried many different models for time until recidivism, based on different probability distributions for the time until failure, with the most successful models being based on the lognormal distribution. We also tried a variety of explanatory variables, and found some that were significantly related to time until recidivism and some that were not. Our models predicted recidivism accurately for a random sample of individuals (our "validation sample") distinct from the sample used in estimation. However, they did not predict recidivism accurately for certain subgroups of the population, such as youthful offenders, drug users, and individuals without prior convictions. Our models also yielded individual predictions that were too imprecise to be of practical use.

A possible explanation for the failure of our models to predict accurately for subgroups of the population or for individuals is that the effects of the explanatory variables were not accurately captured by our specifications. We simply entered the explanatory variables <u>linearly</u> into our models (as is typical in such analyses), and this may be inadequate if their effects on time until recidivism are nonlinear. In the current project we intend to experiment much more intensely with the way in which the explanatory variables are entered into our models. In particular, we will consider quadratic terms, interactions between variables, redefinitions of some variables, and other types of nonlinearities. By doing so we hope to capture adequately the effects of these variables on time until recidivism, so that we can predict more accurately for subgroups of the population and for individuals.

In the research under our previous grant, the choice of statistical model did not have much effect on the estimates of the effects of explanatory variables on time until recidivism. We will therefore do most of our experimentation with explanatory variables in the context of the proportional hazards model, which is relatively easy and economical to estimate. Those results are reported in this document. Results for parametric models based on the lognormal distribution will be provided later, as will a description of the implications of our revised models for predictive accuracy.

As in our previous research, our results are obtained using data from samples of inmates released by the North Carolina Department of Correction in 1978 and 1980. These data are described in detail in Schmidt and Witte, Chapter 2. The estimates in this document are all based on the so-called "estimation samples" as described on p. 23 of Schmidt and Witte. In the next section of this document, we present our results for the 1978 estimation sample, while the results for 1980 are given in the section thereafter.

## 2. Results for the 1978 Estimation Sample

This section contains estimates of the proportional hazards model applied to the 1978 estimation sample, which consists of 1540 observations. All calculations were performed on 80386-based microcomputers using the GAUSS software package. Incidentally, many of the calculations of Schmidt and Witte (which had been done originally on a mainframe computer, in FORTRAN) were redone using GAUSS as a check on the accuracy both of the Schmidt and Witte calculations and on the use of GAUSS. In all cases the Schmidt and Witte calculations were accurately reproduced.

We begin with the "original specification" which corresponds to the final specification of Schmidt and Witte, Chapter 6, Table 6.1, p. 86. This is repeated in Table 1. This model contains the nine explanatory variables TSERVD, AGE, PRIORS, WHITE, FELON, ALCHY, JUNKY, PROPTY AND MALE. These variables are as defined by Schmidt and Witte, Chapter 2, pp. 24-25, with the following exceptions: TSERVD here corresponds to Schmidt and Witte's TSERVD/100; AGE here corresponds to Schmidt and Witte's AGE/1000; and PRIORS here corresponds to Schmidt and Witte's PRIORS/10. For each explanatory variable, we display its sample mean value and standard deviation; its estimated coefficient in the proportional hazards model; the standard error of the estimated coefficient; and the asymptotic t-ratio used to test the hypothesis that the variable's coefficient equals zero. The results in Table 1 reproduce quite closely the results in Schmidt and Witte, Table 6.1. They indicate "that the type of individual most likely to return to prison (and most likely to have a small time until recidivism) is a young, black male with a large number of previous incarcerations, who is a drug addict and/or alcoholic, and whose previous incarceration was lengthy and for a crime against property" (Schmidt and Witte, p. 87).

All of the explanatory variables in this model are binary (dummy) variables except TSERVD, AGE and PRIORS. We begin by considering possible transformations (redefinitions) of these three variables, and more specifically we begin by considering the variable AGE. With AGE simply measured in months (actually, months/1000), the model implies that the marginal effect of an extra year of age is the same whether the individual is twenty years old or forty, and this seems unreasonable. (The difference between a twenty year old and a twenty-one year old seems more important than the difference between a forty year old and a forty-one year old.) One way to put a heavier weight on age differences for younger individuals is to make a logarithmic transformation of age, which we did. That is, we replaced AGE by LN(AGE), where LN(AGE) = (natural) logarithm of AGE. All other variables are as before. The results for this specification are given in Table 2. They are quite similar to the results in Table 1. However, the modified model in Table 2 fits the data better than the original model in Table 1; the logarithm of the likelihood value is -3968.36 as opposed to -3970.70 in Table 2. We therefore will use LN(AGE) instead of AGE from this point on.

Following the same intuition, we next considered the same (logarithmic) transformation of TSERVD. That is, the specification in Table 2 is changed by replacing TSERVD by its natural logarithm, LN(TSERVD). The results for this specification are given in Table 3. They are quite similar to the results in Table 2. Even the log likelihood values are quite similar: -3968.36 for the model in Table 2, and -3969.69 for the model in Table 3. Since the likelihood value is lower when we use LN(TSERVD) in place of TSERVD, we decided against using TSERVD in logarithmic form.

We now proceeded to add variables to the model. The first variable we considered is a binary (dummy) variable for youthful offenders. Since it is not clear at what age one ceases to be youthful, in the present context, we tried nine different values of the age that defined our dummy variable YOUNG: AGE ≤ 17 years, AGE ≤ 18 years, ... , AGE ≤ 25 years. The definition that yielded the best-fitting model (highest log likelihood) was AGE ≤ 20 years. The likelihod values obtained by using alternative years of age to define YOUNG are given in Table 4, and the likelihood value of -3964.91 obtained by using AGE ≤ 20 years is clearly higher than the likelihood values obtained using other definitions of YOUNG.

The results for the proportional hazards model with YOUNG added to the specification of Table 2 (thus yielding a total of ten explanatory variables) are given in Table 5. Note that YOUNG has a coefficient that is statistically significant at commonly used significance levels as measured either by its asymptotic t-ratio (2.67) or by the likelihood ratio test statistic (6.90) based on the change in likelihoods from Table 2 to Table 5. This coefficient indicates a higher failure rate (smaller time until recidivism) for youthful releasees, as probably would have been expected. We also note that the statistical significance of the coefficient of LN(AGE) is reduced only slightly by the addition of YOUNG to the model.

We now return to consideration of the variable TSERVD. We found above that taking logarithms did not improve the fit of the model. We now try a different route to give more weight to small values of TSERVD; namely, adding a dummy variables for short time served. Since it is not clear in the present context what a short time served is, we tried 32 different values of the time served that defined our dummy variable SHORT: TSERVD ≤ 5 months, TSERVD ≤ 6 months, ... , TSERVD ≤ 30 months. The likelihood values obtained by using alternative months of time served to define SHORT are given in Table 6. The data do not indicate the best definition of SHORT as unambiguously as they indicated the best definition of YOUNG, but the highest likelihood value is obtained using SHORT defined as TSERVD ≤ 30 months, and that is the definition we will use.

The results for the proportional hazards model with SHORT added to the specification of Table 5 (thus yielding a total of eleven explanatory variables) are given in Table 7. Note that SHORT has a coefficient that is statistically significant at commonly used significance levels as meassured either by its asymptotic t-ratio (-3.70) or by the likelihood ratio test statistic (13.4) based on the change in likelihoods from Table 5 to Table 7. As expected, the coefficient of SHORT indicates a lower failure rate (longer time until recidivism) for individuals who had served short sentences. The addition of SHORT to the model has a strong effect on the statistical significance of the coefficient of TSERVD (which decreases from 8.59 to 4.05), but the coefficient of TSERVD is still statistically significant at usual significance levels. The results for other variables are not much affected by the addition of SHORT to the model.

We next consider adding to the specification a dummy variable indicating no prior incarcerations (other than the one resulting in the sample sentence), which we will call NOPRIOR.

This is motivated by the results of Schmidt and Witte, Chapter 8, who found that their final model did not predict well for subsamples defined by the presence or absence of prior incarcerations. The results for the specification that adds NOPRIOR to the set of explanatory variables (thus yielding a specification with twelve explanatory variables) are given in Table 8. The log likelihood for this model is -3946.91, a considerable increase from the level (-3958.21) for the model without NOPRIOR. The coefficient of NOPRIOR is statistically significant at any reasonable level, as indicated by its asymptotic t-ratio (-4.78) or the likelihood ratio test statistic (22.6). The addition of NOPRIOR to the model does not change the results for other explanatory variables very much, except that, as would be expected, it reduces the level of significance of PRIORS. The coefficient of NOPRIOR indicates that individuals with no prior incarceration have lower failure rates (longer times until recidivism) that individuals with prior incarcerations, as expected.

We now turn to consideration of interaction terms. The existence of such interaction terms is intuitively reasonable; for example, the effect of being a drug addict (JUNKY = 1) may be different for youthful releasees than for older ones, and this would suggest that an interaction between JUNKY and some measure of age (such as LN(AGE) or YOUNG) might be important. There are many possible interactions, however, and our approach therefore is simply to consider all possible interactions, adding to the specification those that have statistically significant coefficients and that result in the largest increase in the log likelihood value.

We begin by considering thirteen-variable specifications formed by adding a single interaction term to the twelve-variable specification of Table 8. There are 66 such interactions; however, the interaction of NOPRIOR and PRIORS is impossible since their product is always zero. We tried all of the 65 possible specifications, and seven yielded a coefficient for the interaction term that was significantly different from zero at the 1% level (according to the likelihood ratio test). We will restrict our attention to these seven very significant interactions: TSERVD*LN(AGE), TSERVD*YOUNG, TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*FELON, FELON*NOPRIOR and SHORT*NOPRIOR. The log likelihood values obtained using these interaction terms (individually, plus the twelve variables already in the specification of Table 8) are as follows: -3934.94, -3944.24, -3940.71, -3939.96, -3943.22, -3941.99 and -3940.87. Clearly the highest likelihood value is achieved by using the interaction term TSERVD*LN(AGE). The results for the specification that includes this interaction term in addition to the twelve variables previously considered are given in Table 9. Note that the interaction term is very highly significant, as judged either by its asymptotic t-ratio (-4.39) or the likelihood ratio test statistic (23.9); its addition improves the likelihood from -3946.91 to -3934.94. Adding the interaction term has some effects on the coefficients and significance levels of the other explanatory variables; as would be expected, these are most noticable for the variables involved in the interaction (LN(AGE) and TSERVD).

We next considered adding a second interaction term to the model. It was not feasible to try all possible pairs of the 65 interactions with which we started, but we did try all 21 possible pairs of the seven interactions which were most significant in the analysis reported in the previous paragraph. The highest log likelihood value reached was -3931.02, and it was achieved in the specification that included the two interactions TSERVD*LN(AGE) and SHORT*NOPRIOR. Thus this specification just amounts to adding the interaction SHORT*NOPRIOR to the specification of Table 9. The results for this specification are given in Table 10. We note two things. First, the interaction SHORT*NOPRIOR is significant by its asymptotic t-ratio (-2.87) or likelihood ratio test statistic (7.83). Second, including SHORT*NOPRIOR makes the coefficiients of the individual variables SHORT and NOPRIOR insignificant, based on asymptotic t-ratios of -0.43 and 0.09. (We will shortly report that they are also jointly insignificant.)

We now consider adding a third interaction term to the model, along with the two we have so far identified. We tried the five remaining interactions that had very significant coefficients when added individually to the specification of Table 8 (as reported above); these are TSERVD*YOUNG,

TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*FELON, and FELON*NOPRIOR. The resulting log likelihoods are -3930.98, -3931.00, -3930.89, -3930.78 and -3929.55, and the highest value clearly corresponds to the interaction FELON*NOPRIOR. (We also considered another eighteen combinations of three interactions, not necessarily including the two previously included in the specification, and did not find a log likelihood value that exceeded -3929.55.) Based on these log likelihood values we added the interaction FELON*NOPRIOR to the specification of Table 10. The results for this specification are given in Table 11. We note that the coefficient of the third interaction (FELON*NOPRIOR) is only marginally significant, based on its asymptotic t-ratio of 1.72 or the likelihood ratio test statistic of 2.94. (This corresponds to significance at about the 8% level, in both cases.) The inclusion of FELON*NOPRIOR does not change the results for other variables very much. In particular, SHORT and NOPRIOR remain individually insignificant.

The final specification that we consider is obtained by dropping the variables (SHORT, NOPRIOR and FELON*NOPRIOR) that were individually insignificant in the specification of Table 11. The results for this specification are given in Table 12. The log likelihood value attained is -3931.16. Comparing this to the log likelihood value (-3929.55) reported in Table 11, the likelihood ratio test for the joint significance of the coefficients of the three variables just dropped is only 3.22, which is far from the usual critical values for a chi-squared with three degrees of freedom. Thus the variables SHORT, NOPRIOR and FELON*NOPRIOR are insignificant jointly as well as individually. Similarly, if we compare the log likelihood value of -3931.16 to the value (-3931.02) reported in Table 10, the likelihood ratio test statistic for the joint significance of the variables SHORT and NOPRIOR in Table 10 is only 0.28. Thus in that specification as well SHORT and NOPRIOR are insignificant jointly as well as individually.

The specification in Table 12 is our final specification for the proportional hazards model applied to the 1978 estimation sample. It contains 12 variables. The original specification taken from Schmidt and Witte contained nine variables, and eight of them remain in our final specification: TSERVD, PRIORS, WHITE, FELON, ALCHY, JUNKY, PROPTY and MALE. The variable AGE has been replaced by LN(AGE), the logarithm of AGE. In addition, three variables have been added to the original specification: YOUNG, TSERVD*LN(AGE), and SHORT*NOPRIOR. YOUNG is a new variable (a dummy for age of less than or equal to twenty years). TSERVD*LN(AGE) is an interaction between two variables that also appear separately in the specification. Finally, SHORT*NOPRIOR is an interaction between two variables that do not appear separately in the specification. Since SHORT is a dummy variable, and so is NOPRIOR, we note that their product is also a dummy variable: SHORT*NOPRIOR equals one if the individual served a sentence of 30 months or less <u>and</u> had no previous incarcerations, and it equals zero if the individual served a sentence of more than 30 months or had previous incarcerations. Unsurprisingly, individuals with short time served and no prior incarcerations have lower failure rates than individuals with longer sentences or previous incarcerations. However, it is perhaps surprising that SHORT and NOPRIOR affect time until recidivism only in their interaction form. Apparently the effect of a short time served is significant only for individuals with no prior incarcerations; or, equivalently, the effect of no prior incarcerations is significant only for individuals with short time served.

## 3. Results for the 1980 Estimation Sample

This section contains estimates of the proportional hazards model applied to the 1980 estimation sample, which consists of 1435 observations. The basic structure of our analyses is much the same as for the 1978 estimation sample, as described in section 2, so we will present the results of this section more concisely than we did in the last section.

We begin with the "original specification" which corresponds to the final specification of Schmidt and Witte, Chapter 6, Table 6.2, p. 86. This is repeated in Table 13. The model contains the nine explanatory variables TSERVD, AGE, PRIORS, WHITE, MARRIED, ALCHY, JUNKY, PROPTY and MALE. All of these variables are exactly the same as in the previous section, except that MARRIED did not appear in the model for 1978; it is defined as in Schmidt and Witte, p. 25. Also, the variable FELON appeared in the 1978 model, but does not appear in the 1980 model. The results for the 1980 model are quite similar to those for the 1978 model. The only real difference is that the specification now implies that being married reduces the liklihood of return to prison (increases time until recidivism).

We begin by considering the variable LN(AGE) in place of AGE, as was done in section 2. The results for this specification are given in Table 14. Changing AGE to LN(AGE) results in an increase in the log likelihood value, from -3634.05 to -3632.24, and so from now on we will use LN(AGE) to represent age. This does not affect the results for the other variables enough to require comment.

Table 15 gives the results for the same specification as Table 14 except that TSERVD is replaced by its logarithm, LN(TSERVD). This results in a decrease in the log likelihood from -3632.24 to -3635.81. Since the likelihood value is lower when we use LN(TSERVD) in place of TSERVD, we decided against using TSERVD in logarithmic form.

We next consider adding to the specification a dummy variable for youthful releasees. As we did in the analysis of the 1978 data, we tried nine different values of the age that defined our dummy variable YOUNG: AGE ≤ 17 years, AGE ≤ 18 years, ..., AGE ≤ 25 years. The definition that yielded the best-fitting model (highest log likelihood value) was AGE ≤ 18 years. (Note that this makes the variable YOUNG slightly different than in the analysis of the 1978 data, where it was defined as AGE ≤ 20 years.) The likelihood values obtained by using alternative years of age to define YOUNG are given in Table 16. The results for the proportional hazards model with YOUNG added to the specification of Table 14 (thus yielding a total of ten explanatory variables) are given in Table 17. Adding YOUNG to the model does not change any of the results for other variables very much. It increases the log likelihood value from -3632.24 to -3631.19, yielding a likelihood ratio test statistic of 2.10. Neither this statistic nor the asymptotic t-ratio of 1.49 is statistically significant at usual critical levels. Thus YOUNG does not appear to belong in the model, in this specification. However, we will find it to be significant in other specifications to be reported shortly.

We next consider adding a dummy variable SHORT representing a short time served for the sample conviction. As in the analysis of the 1978 data, we tried 32 different values of the time served that defined SHORT: TSERVD ≤ 5 months, TSERVD ≤ 6 months, ... , TSERVD ≤ 30 months. The highest log likelihood value (-3627.41) is obtained using the definition TSERVD ≤ 22 months for SHORT. (This is different from the definition of SHORT in the 1978 analysis, which was TSERVD ≤ 30 months.) The results for the specification that includes SHORT plus the nine variables in the specification of Table 14 (note that YOUNG is not included) are given in Table 19. Adding SHORT decreases the significance of TSERVD, as would be expected, but has no notable effects on the results for other explanatory variables.

Table 20 gives the results for the specification of Table 19 but with YOUNG once again included in the model. The log likelihood value increases modestly, to -3625.67. The coefficient of YOUNG is significant at almost the 5% level as judged both by its asymptotic t-ratio of 1.93 and its likelihood ratio test statistic of 3.48, and we will continue to include it in the specification to the extent that its coefficient remains statistically significant (which it does).

We now increase the number of explanatory variables to twelve by adding the dummy variable NOPRIOR (equal to one for no previous incarcerations) to the specification. The results are given in Table 21. The log likelihood value is increased to -3818.30, a very significant increase, and the statistical significance of the coefficient of NOPRIOR is also confirmed by its large asymptotic t-ratio (-3.90). The addition of NOPRIOR to the specification reduces the significance level of the coefficient of MALE to less than 5%, but we will leave MALE in the specification until we are done considering interaction terms.

We now turn to the consideration of interaction terms. As in the analysis of the 1978 data, we have twelve variables in the specification prior to inclusion of interactions. There are 65 possible interactions (66 minus PRIORS*NOPRIOR, which is always zero) and we tried all 65 thirteen-variable specifications resulting from the addition of a single interaction term to the twelve-variable specification of Table 21. We will restrict our attention to the eight interactions that had very significant coefficients (significant at the 1% level by the likelihood ratio test) in these 65 trial specifications. These interactions are TSERVD*PRIORS, TSERVD*ALCHY, TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*WHITE, PRIORS*PROPTY, WHITE*NOPRIOR, and MALE*NOPRIOR. The corresponding log likelihood values are -3613.42, -3614.65, -3611.74, -3612.80, -3613.79, -3613.06, -3614.70 and -3614.31. The highest log likelihood value (-3611.74) corresponds to the inclusion of the interaction TSERVD*NOPRIOR, and the results for this specification are given in Table 22. Including this interaction reduces the level of significance of TSERVD somewhat but otherwise does not change the results very much.

We next considered adding a second interaction term to the model. We considered all 28 possible pairs of the eight interactions listed in the previous paragraph. The highest log likelihood value reached was -3606.85, and corresponds to the inclusion of the interactions TSERVD*NOPRIOR and PRIORS*WHITE. The results for this specification are given in Table 23. Both interactions are highly significant by any measure. The coefficient of MALE continues to be insignificant at the 5% level, and the significance of the coefficient of TSERVD now falls below 5% as well. However, as before, we leave these variables in the specification until we are done adding interactions to the model.

The highest log likelihood value achieved using three interaction terms is -3602.46, and corresponds to the inclusion of TSERVD*ALCHY, LN(AGE)*SHORT and PRIORS*WHITE. (TSERVD*NOPRIOR is not in this specification despite being in the last two specifications; it will return shortly.) The results for this specification are given in Table 24. All three interactions are highly significant. Note also that TSERVD now regains its significance, but ALCHY and MALE have coefficients that are statistically insignificant at the 5% level.

The highest log likelihood achieved by the use of four interaction terms is -3598.59, and corresponds to adding the interaction TSERVD*NOPRIOR to the specification of Table 24. The results for this specification are given in Table 25. All four interactions are highly significant, but the coefficient of TSERVD once again becomes insignificant at the 5% level.

Table 26 gives the results for the best specification with five interaction terms. These are the four included in the specification of the previous paragraph plus MALE*NOPRIOR. The log likelihood increases to -3595.02, which is a significant increase, and all five interaction terms are statistically significant. TSERVD becomes marginally significant but MALE and ALCHY are still statistically insignificant.

The best-fitting model with six interaction terms adds PRIORS*PROPTY and WHITE*NOPRIOR but deletes PRIORS*WHITE. Therefore the specification now includes the interactions TSERVD*ALCHY, TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*PROPTY, WHITE*NOPRIOR AND MALE*NOPRIOR. This yields a log likelihood value of -3593.23, and the results given in Table 27. The variables WHITE, ALCHY and MALE now have coefficients that are insignificant at the 5% level.

Adding a seventh interaction term fails to improve the model. Table 28 gives the results when the interaction MALE*NOPRIOR is added back into the specification. The log likelihood increases only slightly, to -3592.71, and several of the interaction terms (as well as WHITE, ALCHY and MALE) now have insignificant coefficients. We therefore terminate our attempts to add more interaction terms to the model, ending with the six listed in the previous paragraph (the specification of Table 27).

Since WHITE, ALCHY and MALE have insignificant coefficients in the specification of Table 27, we now drop them from the model. This yields a log likelihood of -3594.75, and the results in Table 29. Comparing this log likelihood value to the log likelihood value (-3594.75) in Table 27, we obtain a likelihood ratio test statistic of 1.52, which is far from being significant. Thus we conclude that the three variables just dropped are jointly as well as individually insignificant.

However, we also note that the coefficient of TSERVD is statistically insignificant at the 5% level in Table 29, based on its asymptotic t-ratio of 1.59. This is not very surprising since the significance level of TSERVD has been marginal at best since we started including interaction terms involving TSERVD. We therefore drop TSERVD from the model. This yields the specification and results given in Table 30. The log likelihhood value is -3595.91. Comparing this to the likelihood value in

Table 29 (-3594.75), we obtain a likelihood ratio test statistic of 2.33, confirming the statistical insignificance of TSERVD in the model. Comparing the likelihood values from Tables 28 and 30, we obtain a likelihood ratio test statistic of 4.40, confirming the joint insignificance of the four variables TSERVD, WHITE, ALCHY and MALE. Thus the inclusion of our six interaction terms and dropping of the four variables just listed is justified by the usual statistical criteria of goodness of fit and statistical significance of coefficients.

The specification of Table 30 is our final specification for the proportional hazards model applied to the 1980 estimation sample. It contains fourteen variables. The original specification of Schmidt and Witte contained nine variables, and only four of them remain in our final specification: PRIORS, MARRIED, JUNKY and PROPTY. The variable AGE has been replaced by its logarithm, LN(AGE). Three new variables have been added to the specification: YOUNG, SHORT and NOPRIOR. The variables TSERVD, WHITE, ALCHY and MALE no longer appear in the model in their original form, though all four appear in at least one of the interaction variables that are now in the model, and the variable SHORT is also based on the value of TSERVD. Finally, the specification now includes the six interaction terms TSERVD*ALCHY, TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*PROPTY, WHITE*NOPRIOR and MALE*NOPRIOR. This is a larger number of significant interactions than we found for the 1978 data, and it is a larger number than we expected to find. As in the 1978 analysis, the significance of interaction terms involving variables that are themselves not significant has interesting (and perhaps strange) implications; for example, being male versus female matters only for individuals with no prior incarcerations. Nevertheless, we have let the data drive our specification, and apparently this is what the data indicate. Furthermore, a verbal summary of the results of our final specification is really not very different from a summary of the resutls of the original specification: the type of individual most likely to return to prison is a young, unmarried drug and alcohol abuser with many previous incarcerations and a long previous time served, and whose previous sentence was for a crime against property. The biggest difference is that race and sex now matter only for individuals with no prior incarcerations; among such individuals, black males have are most likely to return to prison quickly.

## 4. Concluding Remarks and Further Research

We have now completed an investigation into appropriate ways to enter our explanatory variables into the proportional hazards model, for two data sets of North Carolina prison releasees. The final specifications for these models are more complex than the specifications with which we started (and which corresponded to the final specifications under our earlier grant), but they do not change the basic nature of our results. In other words, they are better regarded as elaborations of our earlier results than as substantial changes in them.

The next steps to be taken are as follows. First, we will apply these specifications to parameteric models based on the lognormal model, since in our previous grant lognormal models predicted better than the proportional hazards model and also better than other parametric models. We previously found that the estimated effects of explanatory variables did not depend much on the choice of a parametric or non-parametric model, or on the choice of distribution in a parametric model. If this is still so with our more elaborate treatments of explanatory variables, we will not have to do much experimentation on the ways in which to enter explanatory variables into our parametric models. Second, when we have formulated and estimated appropriate parametric models, we will investigate the improvements in predictive accuracy that we hope will result from more careful and detailed treatment of the explanatory variables in the models.

TABLE 1

PROPORTIONAL HAZARDS MODEL - 1978 DATA

9 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -3970.70

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 1.3712 | 0.1682 | 8.15 |
| AGE | 0.3461 | 0.1212 | -3.4969 | 0.4933 | -7.09 |
| PRIORS | 0.1377 | 0.2789 | 0.8988 | 0.1331 | 6.75 |
| WHITE | 0.5091 | 0.5001 | -0.4404 | 0.0869 | -5.07 |
| FELON | 0.3123 | 0.4636 | -0.5734 | 0.1399 | -4.10 |
| ALCHY | 0.2097 | 0.4073 | 0.4125 | 0.1035 | 3.98 |
| JUNKY | 0.2390 | 0.4266 | 0.3151 | 0.0962 | 3.28 |
| PROPTY | 0.2506 | 0.4335 | 0.4048 | 0.1342 | 3.02 |
| MALE | 0.9383 | 0.2407 | 0.7025 | 0.2407 | 2.92 |

TABLE 2

PROPORTIONAL HAZARDS MODEL - 1978 DATA

9 VARIABLES: CHANGE AGE TO LN(AGE)

LOG-LIKELIHOOD: -3968.36

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 1.3922 | 0.1659 | 8.39 |
| LN(AGE) | 5.7955 | 0.3078 | -1.3628 | 0.1800 | -7.57 |
| PRIORS | 0.1377 | 0.2789 | 0.8677 | 0.1280 | 6.78 |
| WHITE | 0.5091 | 0.5001 | -0.4528 | 0.0870 | -5.20 |
| FELON | 0.3123 | 0.4636 | -0.5511 | 0.1400 | -3.94 |
| ALCHY | 0.2097 | 0.4073 | 0.4340 | 0.1038 | 4.18 |
| JUNKY | 0.2390 | 0.4266 | 0.3253 | 0.0962 | 3.38 |
| PROPTY | 0.2506 | 0.4335 | 0.4055 | 0.1338 | 3.03 |
| MALE | 0.9383 | 0.2407 | 0.7023 | 0.2407 | 2.92 |

TABLE 3

PROPORTIONAL HAZARDS MODEL - 1978 DATA

9 VARIABLES: CHANGE TSERVD TO LN(TSERVD)

LOG-LIKELIHOOD: -3968.69

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| LN(TSERVD) | 2.5255 | 0.8894 | 0.4299 | 0.0600 | 7.16 |
| LN(AGE) | 5.7955 | 0.3078 | -1.3286 | 0.1812 | -7.33 |
| PRIORS | 0.1377 | 0.2789 | 0.8641 | 0.1278 | 6.76 |
| WHITE | 0.5091 | 0.5001 | -0.4089 | 0.0872 | -4.69 |
| FELON | 0.3123 | 0.4636 | -0.6337 | 0.1415 | -4.48 |
| ALCHY | 0.2097 | 0.4073 | 0.4706 | 0.1041 | 4.52 |
| JUNKY | 0.2390 | 0.4266 | 0.2689 | 0.0966 | 2.78 |
| PROPTY | 0.2506 | 0.4335 | 0.3307 | 0.1340 | 2.47 |
| MALE | 0.9383 | 0.2407 | 0.6738 | 0.2408 | 2.80 |

TABLE 4

LOG LIKELIHOOD VALUES FOR DIFFERENT DEFINTIONS OF YOUNG

| | | | |
|---|---|---|---|
| AGE ≤ 17 years: | -3968.34 | AGE ≤ 22 years: | -3967.74 |
| AGE ≤ 18 years: | -3968.35 | AGE ≤ 23 years: | -3968.36 |
| AGE ≤ 19 years: | -3968.34 | AGE ≤ 24 years: | -3968.13 |
| AGE ≤ 20 years: | -3964.91* | AGE ≤ 25 years: | -3968.00 |
| AGE ≤ 21 years: | -3967.78 | | |

TABLE 5

PROPORTIONAL HAZARDS MODEL - 1978 DATA

10 VARIABLES: ADD YOUNG

LOG-LIKELIHOOD: -3964.91

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 1.3980 | 0.1627 | 8.59 |
| LN(AGE) | 5.7955 | 0.3078 | -1.1136 | 0.2006 | -5.55 |
| PRIORS | 0.1377 | 0.2789 | 0.8269 | 0.1285 | 6.43 |
| WHITE | 0.5091 | 0.5001 | -0.4694 | 0.0873 | -5.38 |
| FELON | 0.3123 | 0.4636 | -0.5261 | 0.1409 | -3.73 |
| ALCHY | 0.2097 | 0.4073 | 0.4419 | 0.1039 | 4.25 |
| JUNKY | 0.2390 | 0.4266 | 0.3408 | 0.0965 | 3.53 |
| PROPTY | 0.2506 | 0.4335 | 0.4281 | 0.1343 | 3.19 |
| MALE | 0.9383 | 0.2407 | 0.7236 | 0.2408 | 3.01 |
| YOUNG | 0.1318 | 0.3384 | 0.3569 | 0.1336 | 2.67 |

## TABLE 6

### LOG LIKELIHOOD VALUES FOR DIFFERENT DEFINITIONS OF SHORT

| | | | |
|---|---|---|---|
| TSERVD ≤ 5 months: | -3962.53 | TSERVD ≤ 21 months: | -3963.75 |
| TSERVD ≤ 6 months: | -3964.06 | TSERVD ≤ 22 months: | -3962.90 |
| TSERVD ≤ 7 months: | -3963.16 | TSERVD ≤ 23 months: | -3961.34 |
| TSERVD ≤ 8 months: | -3963.75 | TSERVD ≤ 24 months: | -3959.77 |
| TSERVD ≤ 9 months: | -3963.50 | TSERVD ≤ 25 months: | -3960.60 |
| TSERVD ≤ 10 months: | -3964.12 | TSERVD ≤ 26 months: | -3960.41 |
| TSERVD ≤ 11 months: | -3964.07 | TSERVD ≤ 27 months: | -3961.53 |
| TSERVD ≤ 12 months: | -3963.69 | TSERVD ≤ 28 months: | -3960.42 |
| TSERVD ≤ 13 months: | -3961.81 | TSERVD ≤ 29 months: | -3959.70 |
| TSERVD ≤ 14 months: | -3961.06 | TSERVD ≤ 30 months: | -3958.21* |
| TSERVD ≤ 15 months: | -3961.56 | TSERVD ≤ 31 months: | -3958.26 |
| TSERVD ≤ 16 months: | -3960.63 | TSERVD ≤ 32 months: | -3959.22 |
| TSERVD ≤ 17 months: | -3961.37 | TSERVD ≤ 33 months: | -3960.69 |
| TSERVD ≤ 18 months: | -3961.65 | TSERVD ≤ 34 months: | -3960.73 |
| TSERVD ≤ 19 months: | -3963.47 | TSERVD ≤ 35 months: | -3960.18 |
| TSERVD ≤ 20 months: | -3963.32 | TSERVD ≤ 36 months: | -3960.56 |

## TABLE 7

### PROPORTIONAL HAZARDS MODEL - 1978 DATA

11 VARIABLES: ADD SHORT

LOG-LIKELIHOOD: -3958.21

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 0.9236 | 0.2281 | 4.05 |
| LN(AGE) | 5.7955 | 0.3078 | -1.0357 | 0.2004 | -5.17 |
| PRIORS | 0.1377 | 0.2789 | 0.7654 | 0.1323 | 5.79 |
| WHITE | 0.5091 | 0.5001 | -0.4638 | 0.0874 | -5.31 |
| FELON | 0.3123 | 0.4636 | -0.6027 | 0.1434 | -4.20 |
| ALCHY | 0.2097 | 0.4073 | 0.4301 | 0.1037 | 4.15 |
| JUNKY | 0.2390 | 0.4266 | 0.3323 | 0.0965 | 3.44 |
| PROPTY | 0.2506 | 0.4335 | 0.4077 | 0.1355 | 3.01 |
| MALE | 0.9383 | 0.2407 | 0.7175 | 0.2408 | 2.98 |
| YOUNG | 0.1318 | 0.3384 | 0.4117 | 0.1356 | 3.04 |
| SHORT | 0.8221 | 0.3826 | -0.5312 | 0.1437 | -3.70 |

## TABLE 8

PROPORTIONAL HAZARDS MODEL - 1978 DATA

12 VARIABLES: ADD NOPRIOR

LOG-LIKELIHOOD: -3946.91

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 0.8722 | 0.2329 | 3.74 |
| LN(AGE) | 5.7955 | 0.3078 | -1.1451 | 0.2014 | -5.68 |
| PRIORS | 0.1377 | 0.2789 | 0.4968 | 0.1584 | 3.14 |
| WHITE | 0.5091 | 0.5001 | -0.4650 | 0.0876 | -5.31 |
| FELON | 0.3123 | 0.4636 | -0.5905 | 0.1451 | -4.07 |
| ALCHY | 0.2097 | 0.4073 | 0.4067 | 0.1037 | 3.92 |
| JUNKY | 0.2390 | 0.4266 | 0.3488 | 0.0969 | 3.60 |
| PROPTY | 0.2506 | 0.4335 | 0.3740 | 0.1373 | 2.72 |
| MALE | 0.9383 | 0.2407 | 0.6900 | 0.2409 | 2.86 |
| YOUNG | 0.1318 | 0.3384 | 0.4938 | 0.1372 | 3.60 |
| SHORT | 0.8221 | 0.3826 | -0.4287 | 0.1465 | -2.93 |
| NOPRIOR | 0.5766 | 0.4943 | -0.5081 | 0.1063 | -4.78 |

## TABLE 9

PROPORTIONAL HAZARDS MODEL - 1978 DATA

13 VARIABLES: ADD TSERVD*LN(AGE)

LOG-LIKELIHOOD: -3934.94

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 19.5971 | 4.1748 | 4.69 |
| LN(AGE) | 5.7955 | 0.3078 | -0.5066 | 0.2375 | -2.13 |
| PRIORS | 0.1377 | 0.2789 | 0.5296 | 0.1508 | 3.51 |
| WHITE | 0.5091 | 0.5001 | -0.4179 | 0.0882 | -4.74 |
| FELON | 0.3123 | 0.4636 | -0.6280 | 0.1472 | -4.27 |
| ALCHY | 0.2097 | 0.4073 | 0.3557 | 0.1044 | 3.41 |
| JUNKY | 0.2390 | 0.4266 | 0.3319 | 0.0969 | 3.43 |
| PROPTY | 0.2506 | 0.4335 | 0.3077 | 0.1390 | 2.21 |
| MALE | 0.9383 | 0.2407 | 0.6467 | 0.2411 | 2.68 |
| YOUNG | 0.1318 | 0.3384 | 0.6051 | 0.1370 | 4.42 |
| SHORT | 0.8221 | 0.3826 | -0.2844 | 0.1563 | -1.82 |
| NOPRIOR | 0.5766 | 0.4943 | -0.4646 | 0.1055 | -4.40 |
| TSERVD*LN(AGE) | 1.0939 | 1.2229 | -3.1335 | 0.7140 | -4.39 |

TABLE 10

PROPORTIONAL HAZARDS MODEL - 1978 DATA

14 VARIABLES: ADD SHORT*NOPRIOR

LOG-LIKELIHOOD: -3931.02

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 17.9235 | 4.1698 | 4.30 |
| LN(AGE) | 5.7955 | 0.3078 | -0.5816 | 0.2393 | -2.43 |
| PRIORS | 0.1377 | 0.2789 | 0.5456 | 0.1491 | 3.66 |
| WHITE | 0.5091 | 0.5001 | -0.4220 | 0.0882 | -4.78 |
| FELON | 0.3123 | 0.4636 | -0.6294 | 0.1467 | -4.29 |
| ALCHY | 0.2097 | 0.4073 | 0.3477 | 0.1046 | 3.33 |
| JUNKY | 0.2390 | 0.4266 | 0.3279 | 0.0969 | 3.39 |
| PROPTY | 0.2506 | 0.4335 | 0.3242 | 0.1382 | 2.35 |
| MALE | 0.9383 | 0.2407 | 0.6434 | 0.2412 | 2.67 |
| YOUNG | 0.1318 | 0.3384 | 0.6507 | 0.1388 | 4.69 |
| SHORT | 0.8221 | 0.3826 | -0.0753 | 0.1733 | -0.43 |
| NOPRIOR | 0.5766 | 0.4943 | 0.0184 | 0.1935 | 0.09 |
| TSERVD*LN(AGE) | 1.0939 | 1.2229 | -2.8356 | 0.7118 | -3.98 |
| SHORT*NOPRIOR | 0.5266 | 0.4995 | -0.6212 | 0.2162 | - 2.87 |

TABLE 11

PROPORTIONAL HAZARDS MODEL - 1978 DATA

15 VARIABLES: ADD FELON*NOPRIOR

LOG-LIKELIHOOD: -3929.55

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 17.7422 | 4.1837 | 4.24 |
| LN(AGE) | 5.7955 | 0.3078 | -0.6001 | 0.2399 | -2.50 |
| PRIORS | 0.1377 | 0.2789 | 0.5404 | 0.1492 | 3.62 |
| WHITE | 0.5091 | 0.5001 | -0.4261 | 0.0882 | -4.83 |
| FELON | 0.3123 | 0.4636 | -0.8061 | 0.1795 | -4.49 |
| ALCHY | 0.2097 | 0.4073 | 0.3429 | 0.1047 | 3.28 |
| JUNKY | 0.2390 | 0.4266 | 0.3270 | 0.0968 | 3.38 |
| PROPTY | 0.2506 | 0.4335 | 0.3357 | 0.1379 | 2.43 |
| MALE | 0.9383 | 0.2407 | 0.6338 | 0.2413 | 2.63 |
| YOUNG | 0.1318 | 0.3384 | 0.6658 | 0.1389 | 4.79 |
| SHORT | 0.8221 | 0.3826 | -0.1654 | 0.1808 | -0.91 |
| NOPRIOR | 0.5766 | 0.4943 | -0.2300 | 0.2415 | -0.95 |
| TSERVD*LN(AGE) | 1.0939 | 1.2229 | -2.8047 | 0.7136 | -3.93 |
| SHORT*NOPRIOR | 0.5266 | 0.4995 | -0.4473 | 0.2388 | -1.87 |
| FELON*NOPRIOR | 0.1578 | 0.3647 | 0.3609 | 0.2094 | 1.72 |

TABLE 12

PROPORTIONAL HAZARDS MODEL - 1978 DATA

12 VARIABLES: DELETE SHORT, NOPRIOR, FELON*NOPRIOR

LOG-LIKELIHOOD: -3931.16

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1880 | 0.2054 | 18.1506 | 4.0904 | 4.44 |
| LN(AGE) | 5.7955 | 0.3078 | -0.5891 | 0.2379 | -2.48 |
| PRIORS | 0.1377 | 0.2789 | 0.5435 | 0.1454 | 3.74 |
| WHITE | 0.5091 | 0.5001 | -0.4207 | 0.0882 | -4.77 |
| FELON | 0.3123 | 0.4636 | -0.6229 | 0.1462 | -4.26 |
| ALCHY | 0.2097 | 0.4073 | 0.3469 | 0.1046 | 3.32 |
| JUNKY | 0.2390 | 0.4266 | 0.3276 | 0.0968 | 3.39 |
| PROPTY | 0.2506 | 0.4335 | 0.3256 | 0.1378 | 2.36 |
| MALE | 0.9383 | 0.2407 | 0.6427 | 0.2412 | 2.66 |
| YOUNG | 0.1318 | 0.3384 | 0.6474 | 0.1383 | 4.68 |
| TSERVD*LN(AGE) | 1.0939 | 1.2229 | -2.8601 | 0.6993 | -4.09 |
| SHORT*NOPRIOR | 0.5266 | 0.4995 | -0.6204 | 0.1094 | -5.67 |

TABLE 13

PROPORTIONAL HAZARDS MODEL - 1980 DATA

9 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -3634.05

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 1.1314 | 0.1515 | 7.47 |
| AGE | 0.3390 | 0.1158 | -3.8762 | 0.5924 | -6.54 |
| PRIORS | 0.1392 | 0.3062 | 0.9847 | 0.1100 | 8.95 |
| WHITE | 0.5101 | 0.5001 | -0.2378 | 0.0889 | -2.68 |
| MARRIED | 0.2341 | 0.4236 | -0.3938 | 0.1242 | -3.17 |
| ALCHY | 0.3568 | 0.4792 | 0.3635 | 0.0955 | 3.80 |
| JUNKY | 0.2181 | 0.4131 | 0.2390 | 0.1008 | 2.37 |
| PROPTY | 0.4474 | 0.4974 | 0.3216 | 0.0912 | 3.53 |
| MALE | 0.9463 | 0.2254 | 0.5622 | 0.2554 | 2.20 |

TABLE 14

PROPORTIONAL HAZARDS MODEL - 1980 DATA

9 VARIABLES: CHANGE AGE TO LN(AGE)

LOG-LIKELIHOOD: -3632.24

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| T | 0.1952 | 0.2386 | 1.1506 | 0.1475 | 7.80 |
| LN(AGE) | 5.7761 | 0.3056 | -1.4682 | 0.2093 | -7.01 |
| PRIORS | 0.1392 | 0.3062 | 0.9742 | 0.1063 | 9.17 |
| WHITE | 0.5101 | 0.5001 | -0.2481 | 0.0891 | -2.79 |
| MARRIED | 0.2341 | 0.4236 | -0.3524 | 0.1253 | -2.81 |
| ALCHY | 0.3568 | 0.4792 | 0.3783 | 0.0959 | 3.94 |
| JUNKY | 0.2181 | 0.4131 | 0.2407 | 0.1009 | 2.38 |
| PROPTY | 0.4474 | 0.4974 | 0.3032 | 0.0919 | 3.30 |
| MALE | 0.9463 | 0.2254 | 0.5510 | 0.2554 | 2.16 |

TABLE 15

PROPORTIONAL HAZARDS MODEL - 1980 DATA

9 VARIABLES: CHANGE TSERVD TO LN(TSERVD)

LOG-LIKELIHOOD: -3635.81

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| LN(TSERVD) | 2.5072 | 0.9461 | 0.3095 | 0.0512 | 6.04 |
| LN(AGE) | 5.7761 | 0.3056 | -1.3387 | 0.2037 | -6.57 |
| PRIORS | 0.1392 | 0.3062 | 0.9415 | 0.1066 | 8.83 |
| WHITE | 0.5101 | 0.5001 | -0.2553 | 0.0890 | -2.87 |
| MARRIED | 0.2341 | 0.4236 | -0.3513 | 0.1256 | -2.80 |
| ALCHY | 0.3568 | 0.4792 | 0.3743 | 0.0959 | 3.90 |
| JUNKY | 0.2181 | 0.4131 | 0.2155 | 0.1011 | 2.13 |
| PROPTY | 0.4474 | 0.4974 | 0.3022 | 0.0912 | 3.31 |
| MALE | 0.9463 | 0.2254 | 0.5412 | 0.2555 | 2.12 |

TABLE 16

LOG LIKELIHOOD VALUES FOR DIFFERENT DEFINTIONS OF YOUNG

| | | | |
|---|---|---|---|
| AGE ≤ 17 years: | -3631.89 | AGE ≤ 22 years: | -3632.19 |
| AGE ≤ 18 years: | -3631.19* | AGE ≤ 23 years: | -3632.17 |
| AGE ≤ 19 years: | -3632.15 | AGE ≤ 24 years: | -3632.13 |
| AGE ≤ 20 years: | -3631.86 | AGE ≤ 25 years: | -3632.23 |
| AGE ≤ 21 years: | -3632.19 | | |

## TABLE 17

### PROPORTIONAL HAZARDS MODEL - 1980 DATA

### 10 VARIABLES: ADD YOUNG

### LOG-LIKELIHOOD: -3631.19

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 1.1553 | 0.1459 | 7.92 |
| LN(AGE) | 5.7761 | 0.3056 | -1.3780 | 0.2170 | -6.35 |
| PRIORS | 0.1392 | 0.3062 | 0.9595 | 0.1068 | 8.98 |
| WHITE | 0.5101 | 0.5001 | -0.2536 | 0.0892 | -2.84 |
| FELON | 0.2341 | 0.4236 | -0.3467 | 0.1255 | -2.76 |
| ALCHY | 0.3568 | 0.4792 | 0.3928 | 0.0967 | 4.06 |
| JUNKY | 0.2181 | 0.4131 | 0.2510 | 0.1012 | 2.48 |
| PROPTY | 0.4474 | 0.4974 | 0.2945 | 0.0923 | 3.19 |
| MALE | 0.9463 | 0.2254 | 0.5353 | 0.2557 | 2.09 |
| YOUNG | 0.0516 | 0.2212 | 0.2612 | 0.1759 | 1.49 |

## TABLE 18

### LOG LIKELIHOOD VALUES FOR DIFFERENT DEFINITIONS OF SHORT

| | | | |
|---|---|---|---|
| TSERVD ≤ 5 months: | -3632.21 | TSERVD ≤ 21 months: | -3628.49 |
| TSERVD ≤ 6 months: | -3632.19 | TSERVD ≤ 22 months: | -3627.41* |
| TSERVD ≤ 7 months: | -3632.19 | TSERVD ≤ 23 months: | -3628.93 |
| TSERVD ≤ 8 months: | -3632.24 | TSERVD ≤ 24 months: | -3628.61 |
| TSERVD ≤ 9 months: | -3632.10 | TSERVD ≤ 25 months: | -3629.20 |
| TSERVD ≤ 10 months: | -3632.01 | TSERVD ≤ 26 months: | -3629.14 |
| TSERVD ≤ 11 months: | -3632.06 | TSERVD ≤ 27 months: | -3629.43 |
| TSERVD ≤ 12 months: | -3631.99 | TSERVD ≤ 28 months: | -3629.43 |
| TSERVD ≤ 13 months: | -3631.51 | TSERVD ≤ 29 months: | -3630.79 |
| TSERVD ≤ 14 months: | -3631.51 | TSERVD ≤ 30 months: | -3629.10 |
| TSERVD ≤ 15 months: | -3629.16 | TSERVD ≤ 31 months: | -3629.59 |
| TSERVD ≤ 16 months: | -3628.75 | TSERVD ≤ 32 months: | -3629.50 |
| TSERVD ≤ 17 months: | -3628.95 | TSERVD ≤ 33 months: | -3630.71 |
| TSERVD ≤ 18 months: | -3628.82 | TSERVD ≤ 34 months: | -3631.03 |
| TSERVD ≤ 19 months: | -3627.61 | TSERVD ≤ 35 months: | -3630.63 |
| TSERVD ≤ 20 months: | -3627.43 | TSERVD ≤ 36 months: | -3631.19 |

TABLE 19

PROPORTIONAL HAZARDS MODEL - 1980 DATA

10 VARIABLES: ADD SHORT

LOG-LIKELIHOOD: -3627.41

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.8083 | 0.2065 | 3.91 |
| LN(AGE) | 5.7761 | 0.3056 | -1.5062 | 0.2107 | -7.15 |
| PRIORS | 0.1392 | 0.3062 | 0.9621 | 0.1067 | 9.02 |
| WHITE | 0.5101 | 0.5001 | -0.2519 | 0.0891 | -2.83 |
| FELON | 0.2341 | 0.4236 | -0.3620 | 0.1256 | -2.88 |
| ALCHY | 0.3568 | 0.4792 | 0.4157 | 0.0965 | 4.31 |
| JUNKY | 0.2181 | 0.4131 | 0.2153 | 0.1011 | 2.13 |
| PROPTY | 0.4474 | 0.4974 | 0.2907 | 0.0915 | 3.18 |
| MALE | 0.9463 | 0.2254 | 0.5396 | 0.2553 | 2.11 |
| SHORT | 0.7178 | 0.4502 | -0.3760 | 0.1207 | -3.12 |

TABLE 20

PROPORTIONAL HAZARDS MODEL - 1980 DATA

11 VARIABLES: ADD YOUNG (AGAIN)

LOG-LIKELIHOOD: -3625.67

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.8166 | 0.2000 | 4.08 |
| LN(AGE) | 5.7761 | 0.3056 | -1.3928 | 0.2179 | -6.39 |
| PRIORS | 0.1392 | 0.3062 | 0.9445 | 0.1074 | 8.79 |
| WHITE | 0.5101 | 0.5001 | -0.2714 | 0.0894 | -3.04 |
| MARRIED | 0.2341 | 0.4236 | -0.3479 | 0.1257 | -2.77 |
| ALCHY | 0.3568 | 0.4792 | 0.4369 | 0.0974 | 4.49 |
| JUNKY | 0.2181 | 0.4131 | 0.2132 | 0.1018 | 2.09 |
| PROPTY | 0.4474 | 0.4974 | 0.2798 | 0.0920 | 3.04 |
| MALE | 0.9463 | 0.2254 | 0.5050 | 0.2557 | 1.98 |
| YOUNG | 0.0516 | 0.2212 | 0.3450 | 0.1784 | 1.93 |
| SHORT | 0.6829 | 0.4655 | -0.3875 | 0.1165 | -3.33 |

TABLE 21

PROPORTIONAL HAZARDS MODEL - 1980 DATA

12 VARIABLES: ADD NOPRIOR

LOG-LIKELIHOOD: -3618.30

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.7500 | 0.2076 | 3.61 |
| LN(AGE) | 5.7761 | 0.3056 | -1.6231 | 0.2291 | -7.08 |
| PRIORS | 0.1392 | 0.3062 | 0.8238 | 0.1226 | 6.72 |
| WHITE | 0.5101 | 0.5001 | -0.2686 | 0.0890 | -3.02 |
| FELON | 0.2341 | 0.4236 | -0.3344 | 0.1254 | -2.67 |
| ALCHY | 0.3568 | 0.4792 | 0.4290 | 0.0974 | 4.40 |
| JUNKY | 0.2181 | 0.4131 | 0.2326 | 0.1014 | 2.29 |
| PROPTY | 0.4474 | 0.4974 | 0.2705 | 0.0919 | 2.94 |
| MALE | 0.9463 | 0.2254 | 0.4652 | 0.2560 | 1.82 |
| YOUNG | 0.0516 | 0.2212 | 0.3787 | 0.1788 | 2.12 |
| SHORT | 0.7178 | 0.4502 | -0.3844 | 0.1217 | -3.16 |
| NOPRIOR | 0.5909 | 0.4918 | -0.4117 | 0.1054 | -3.90 |

TABLE 22

PROPORTIONAL HAZARDS MODEL - 1980 DATA

13 VARIABLES: ONE INTERACTION

LOG-LIKELIHOOD: -3611.74

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.5228 | 0.2465 | 2.12 |
| LN(AGE) | 5.7761 | 0.3056 | -1.6621 | 0.2288 | -7.26 |
| PRIORS | 0.1392 | 0.3062 | 0.8527 | 0.1227 | 6.95 |
| WHITE | 0.5101 | 0.5001 | -0.2437 | 0.0896 | -2.72 |
| FELON | 0.2341 | 0.4236 | -0.3287 | 0.1253 | -2.62 |
| ALCHY | 0.3568 | 0.4792 | 0.4144 | 0.0977 | 4.24 |
| JUNKY | 0.2181 | 0.4131 | 0.2427 | 0.1015 | 2.39 |
| PROPTY | 0.4474 | 0.4974 | 0.2753 | 0.0919 | 2.99 |
| MALE | 0.9463 | 0.2254 | 0.4422 | 0.2562 | 1.73 |
| YOUNG | 0.0516 | 0.2212 | 0.4405 | 0.1796 | 2.45 |
| SHORT | 0.7178 | 0.4502 | -0.2688 | 0.1266 | -2.12 |
| NOPRIOR | 0.5909 | 0.4918 | -0.7098 | 0.1326 | -5.35 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 1.3072 | 0.3410 | 3.83 |

TABLE 23

PROPORTIONAL HAZARDS MODEL - 1980 DATA

14 VARIABLES: TWO INTERACTIONS

LOG-LIKELIHOOD: -3606.85

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.4794 | 0.2510 | 1.91 |
| LN(AGE) | 5.7761 | 0.3056 | -1.7365 | 0.2318 | -7.49 |
| PRIORS | 0.1392 | 0.3062 | 0.7137 | 0.1515 | 4.71 |
| WHITE | 0.5101 | 0.5001 | -0.3792 | 0.1000 | -3.79 |
| MARRIED | 0.2341 | 0.4236 | -0.3195 | 0.1252 | -2.55 |
| ALCHY | 0.3568 | 0.4792 | 0.3991 | 0.0983 | 4.06 |
| JUNKY | 0.2181 | 0.4131 | 0.2577 | 0.1015 | 2.54 |
| PROPTY | 0.4474 | 0.4974 | 0.2862 | 0.0917 | 3.12 |
| MALE | 0.9463 | 0.2254 | 0.4322 | 0.2562 | 1.69 |
| YOUNG | 0.0516 | 0.2212 | 0.4211 | 0.1798 | 2.34 |
| SHORT | 0.7178 | 0.4502 | -0.2582 | 0.1271 | -2.03 |
| NOPRIOR | 0.5909 | 0.4918 | -0.6719 | 0.1341 | -5.01 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 1.3489 | 0.3424 | 3.94 |
| PRIORS*WHITE | 0.0705 | 0.2109 | 0.7288 | 0.2252 | 3.24 |

TABLE 24

PROPORTIONAL HAZARDS MODEL - 1980 DATA

15 VARIABLES: THREE INTERACTIONS

LOG-LIKELIHOOD: -3602.46

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.7914 | 0.2551 | 3.10 |
| LN(AGE) | 5.7761 | 0.3056 | -2.8397 | 0.3801 | -7.47 |
| PRIORS | 0.1392 | 0.3062 | 0.7695 | 0.1571 | 4.90 |
| WHITE | 0.5101 | 0.5001 | -0.4249 | 0.1001 | -4.25 |
| MARRIED | 0.2341 | 0.4236 | -0.2898 | 0.1255 | -2.31 |
| ALCHY | 0.3568 | 0.4792 | 0.1633 | 0.1227 | 1.33 |
| JUNKY | 0.2181 | 0.4131 | 0.2922 | 0.1021 | 2.86 |
| PROPTY | 0.4474 | 0.4974 | 0.3020 | 0.0919 | 3.29 |
| MALE | 0.9463 | 0.2254 | 0.4626 | 0.2567 | 1.80 |
| YOUNG | 0.0516 | 0.2212 | 0.4212 | 0.1805 | 2.33 |
| SHORT | 0.7178 | 0.4502 | -9.2334 | 2.2210 | -4.16 |
| NOPRIOR | 0.5909 | 0.4918 | -0.3552 | 0.1080 | -3.29 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.0608 | 0.3242 | 3.27 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.5652 | 0.3918 | 3.99 |
| PRIORS*WHITE | 0.0705 | 0.2109 | 0.7859 | 0.2307 | 3.41 |

TABLE 25

PROPORTIONAL HAZARDS MODEL - 1980 DATA

16 VARIABLES: FOUR INTERACTIONS

LOG-LIKELIHOOD: -3598.59

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.5531 | 0.2881 | 1.92 |
| LN(AGE) | 5.7761 | 0.3056 | -2.6361 | 0.3766 | -7.00 |
| PRIORS | 0.1392 | 0.3062 | 0.7695 | 0.1566 | 4.91 |
| WHITE | 0.5101 | 0.5001 | -0.4066 | 0.1004 | -4.05 |
| MARRIED | 0.2341 | 0.4236 | -0.2878 | 0.1254 | -2.29 |
| ALCHY | 0.3568 | 0.4792 | 0.1523 | 0.1247 | 1.22 |
| JUNKY | 0.2181 | 0.4131 | 0.2933 | 0.1021 | 2.87 |
| PROPTY | 0.4474 | 0.4974 | 0.3003 | 0.0919 | 3.27 |
| MALE | 0.9463 | 0.2254 | 0.4524 | 0.2567 | 1.76 |
| YOUNG | 0.0516 | 0.2212 | 0.4564 | 0.1811 | 2.52 |
| SHORT | 0.7178 | 0.4502 | -7.3934 | 2.2678 | -3.26 |
| NOPRIOR | 0.5909 | 0.4918 | -0.5981 | 0.1373 | -4.36 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.0736 | 0.3382 | 3.17 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 1.0608 | 0.3591 | 2.95 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.2583 | 0.3986 | 3.16 |
| PRIORS*WHITE | 0.0705 | 0.2109 | 0.7863 | 0.2299 | 3.42 |

TABLE 26

PROPORTIONAL HAZARDS MODEL - 1980 DATA

17 VARIABLES: FIVE INTERACTIONS

LOG-LIKELIHOOD: -3595.02

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.5814 | 0.2863 | 2.03 |
| LN(AGE) | 5.7761 | 0.3056 | -2.6263 | 0.3766 | -6.97 |
| PRIORS | 0.1392 | 0.3062 | 0.7651 | 0.1581 | 4.84 |
| WHITE | 0.5101 | 0.5001 | -0.3978 | 0.1005 | -3.96 |
| MARRIED | 0.2341 | 0.4236 | -0.2873 | 0.1255 | -2.29 |
| ALCHY | 0.3568 | 0.4792 | 0.1651 | 0.1245 | 1.33 |
| JUNKY | 0.2181 | 0.4131 | 0.2729 | 0.1022 | 2.67 |
| PROPTY | 0.4474 | 0.4974 | 0.3080 | 0.0920 | 3.35 |
| MALE | 0.9463 | 0.2254 | -0.2959 | 0.3288 | -0.90 |
| YOUNG | 0.0516 | 0.2212 | 0.4380 | 0.1812 | 2.42 |
| SHORT | 0.7178 | 0.4502 | -7.2660 | 2.2691 | -3.20 |
| NOPRIOR | 0.5909 | 0.4918 | -1.9283 | 0.5258 | -3.67 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.0480 | 0.3380 | 3.10 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 1.0234 | 0.3597 | 2.85 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.2364 | 0.3988 | 3.10 |
| PRIORS*WHITE | 0.0705 | 0.2109 | 0.8088 | 0.2306 | 3.51 |
| MALE*NOPRIOR | 0.5512 | 0.4975 | 1.3847 | 0.5279 | 2.62 |

TABLE 27

PROPORTIONAL HAZARDS MODEL - 1980 DATA

18 VARIABLES: SIX INTERACTIONS

LOG-LIKELIHOOD: -3593.23

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.6067 | 0.2881 | 2.11 |
| LN(AGE) | 5.7761 | 0.3056 | -2.5343 | 0.3785 | -6.70 |
| PRIORS | 0.1392 | 0.3062 | 1.4398 | 0.2079 | 6.92 |
| WHITE | 0.5101 | 0.5001 | -0.0329 | 0.1297 | -0.25 |
| MARRIED | 0.2341 | 0.4236 | -0.2874 | 0.1256 | -2.29 |
| ALCHY | 0.3568 | 0.4792 | 0.1819 | 0.1244 | 1.46 |
| JUNKY | 0.2181 | 0.4131 | 0.2719 | 0.1021 | 2.66 |
| PROPTY | 0.4474 | 0.4974 | 0.4002 | 0.1026 | 3.90 |
| MALE | 0.9463 | 0.2254 | -0.3586 | 0.3299 | -1.09 |
| YOUNG | 0.0516 | 0.2212 | 0.4069 | 0.1816 | 2.24 |
| SHORT | 0.7178 | 0.4502 | -6.6643 | 2.2729 | -2.93 |
| NOPRIOR | 0.5909 | 0.4918 | -1.7744 | 0.5289 | -3.36 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.0005 | 0.3352 | 2.98 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 0.9470 | 0.3616 | 2.62 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.1256 | 0.3992 | 2.82 |
| PRIORS*PROPTY | 0.0626 | 0.2412 | -0.6303 | 0.2273 | -2.77 |
| WHITE*NOPRIOR | 0.2983 | 0.4577 | -0.4448 | 0.1791 | -2.48 |
| MALE*NOPRIOR | 0.5512 | 0.4975 | 1.4550 | 0.5287 | 2.75 |

TABLE 28

PROPORTIONAL HAZARDS MODEL - 1980 DATA

19 VARIABLES: SEVEN INTERACTIONS

LOG-LIKELIHOOD: -3592.71

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.6047 | 0.2879 | 2.10 |
| LN(AGE) | 5.7761 | 0.3056 | -2.5823 | 0.3807 | -6.78 |
| PRIORS | 0.1392 | 0.3062 | 1.2471 | 0.2849 | 4.38 |
| WHITE | 0.5101 | 0.5001 | -0.1506 | 0.1742 | -0.86 |
| MARRIED | 0.2341 | 0.4236 | -0.2876 | 0.1256 | -2.29 |
| ALCHY | 0.3568 | 0.4792 | 0.1754 | 0.1248 | 1.41 |
| JUNKY | 0.2181 | 0.4131 | 0.2753 | 0.1022 | 2.69 |
| PROPTY | 0.4474 | 0.4974 | 0.3844 | 0.1041 | 3.69 |
| MALE | 0.9463 | 0.2254 | -0.3510 | 0.3299 | -1.06 |
| YOUNG | 0.0516 | 0.2212 | 0.4137 | 0.1817 | 2.28 |
| SHORT | 0.7178 | 0.4502 | -6.9742 | 2.2904 | -3.04 |
| NOPRIOR | 0.5909 | 0.4918 | -1.8176 | 0.5306 | -3.43 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.0086 | 0.3371 | 2.99 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 0.9661 | 0.3623 | 2.67 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.1819 | 0.4025 | 2.94 |
| PRIORS*WHITE | 0.0705 | 0.2109 | 0.3277 | 0.3217 | 1.02 |
| PRIORS*PROPTY | 0.0626 | 0.2412 | -0.4848 | 0.2712 | -1.79 |
| WHITE*NOPRIOR | 0.2983 | 0.4577 | -0.3289 | 0.2126 | -1.55 |
| MALE*NOPRIOR | 0.5512 | 0.4975 | 1.4458 | 0.5288 | 2.73 |

TABLE 29

PROPORTIONAL HAZARDS MODEL - 1980 DATA

15 VARIABLES: DROP WHITE, ALCHY, MALE, PRIORS*WHITE

LOG-LIKELIHOOD: -3594.75

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| TSERVD | 0.1952 | 0.2386 | 0.4505 | 0.2829 | 1.59 |
| LN(AGE) | 5.7761 | 0.3056 | -2.5440 | 0.3781 | -6.73 |
| PRIORS | 0.1392 | 0.3062 | 1.4468 | 0.2068 | 6.99 |
| MARRIED | 0.2341 | 0.4236 | -0.2920 | 0.1253 | -2.33 |
| JUNKY | 0.2181 | 0.4131 | 0.2982 | 0.1005 | 2.97 |
| PROPTY | 0.4474 | 0.4974 | 0.3953 | 0.1023 | 3.86 |
| YOUNG | 0.0516 | 0.2212 | 0.3774 | 0.1802 | 2.09 |
| SHORT | 0.7178 | 0.4502 | -7.0553 | 2.2526 | -3.13 |
| NOPRIOR | 0.5909 | 0.4918 | -1.4501 | 0.4280 | -3.39 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.2972 | 0.2573 | 5.04 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 0.9859 | 0.3579 | 2.75 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 1.1956 | 0.3956 | 3.02 |
| PRIORS*PROPTY | 0.0626 | 0.2412 | -0.6191 | 0.2245 | -2.76 |
| WHITE*NOPRIOR | 0.2983 | 0.4577 | -0.4685 | 0.1247 | -3.76 |
| MALE*NOPRIOR | 0.5512 | 0.4975 | 1.1200 | 0.4140 | 2.71 |

TABLE 30

PROPORTIONAL HAZARDS MODEL - 1980 DATA

14 VARIABLES: DROP TSERVD

LOG-LIKELIHOOD: -3595.91

| VARIABLE | MEAN | S.D. | COEFFICIENT | S.E. | T-RATIO |
|---|---|---|---|---|---|
| LN(AGE) | 5.7761 | 0.3056 | -2.3575 | 0.3518 | -6.70 |
| PRIORS | 0.1392 | 0.3062 | 1.4154 | 0.2045 | 6.92 |
| MARRIED | 0.2341 | 0.4236 | -0.2894 | 0.1253 | -2.31 |
| JUNKY | 0.2181 | 0.4131 | 0.2901 | 0.1004 | 2.89 |
| PROPTY | 0.4474 | 0.4974 | 0.3936 | 0.1022 | 3.85 |
| YOUNG | 0.0516 | 0.2212 | 0.3823 | 0.1803 | 2.12 |
| SHORT | 0.7178 | 0.4502 | -5.9580 | 2.1087 | -2.83 |
| NOPRIOR | 0.5909 | 0.4918 | -1.5110 | 0.4262 | -3.55 |
| TSERVD*ALCHY | 0.0541 | 0.1435 | 1.4657 | 0.2498 | 5.87 |
| TSERVD*NOPRIOR | 0.0963 | 0.1629 | 1.2161 | 0.3331 | 3.65 |
| LN(AGE)*SHORT | 4.1327 | 2.6062 | 0.9861 | 0.3661 | 2.69 |
| PRIORS*PROPTY | 0.0626 | 0.2412 | -0.5994 | 0.2230 | -2.69 |
| WHITE*NOPRIOR | 0.2983 | 0.4577 | -0.4745 | 0.1247 | -3.81 |
| MALE*NOPRIOR | 0.5512 | 0.4975 | 1.1266 | 0.4139 | 2.72 |

RESULTS ON FUNCTIONAL FORM

PARAMETRIC (LOGNORMAL) MODELS, NORTH CAROLINA DATA

Ching-Fan Chung, Peter Schmidt and Ann Witte

Report 2 of results under NIJ grant 89-IJ-CX-0010

## 1. Introduction

The work described in this report is a continuation of the work described in our previous report, "Additional Results on Functional Form, Proportional Hazards Model, North Carolina Data," September, 1989, hereafter called REPORT 1. This work is an extension of the analyses performed under our previous grant (84-IJ-CX-0021) and reported in P. Schmidt and A.D. Witte, Predicting Recidivism Using Survival Models, Springer-Verlag, 1989, hereafter called Schmidt and Witte, 1989.

REPORT 1 dealt with the proportional hazards model. It investigated ways to improve the models of Schmidt and Witte, 1989, by introducing explanatory variables into the proportional hazards model in nonlinear ways. Specifically, logarithmic transformations of certain variables and interactions between variables were found to improve the fit of the model significantly, for both the 1978 and 1980 estimation samples.

We now turn to parametric models based on the lognormal distribution, which were found in our previous work to predict better than the proportional hazards model. We wish to see whether these parametric models can also be improved by entering explanatory variables in nonlinear ways. As might be expected, we find that they can be improved in this way. However, the exact type of nonlinearities that the data support turn out to depend on the model used to a larger extent than we would have expected.

A further report will contain a comparison of the predictive accuracy of the models estimated under our previous grant and the improved models described in REPORT 1 and this report.

## 2. Results for the 1978 Estimation Sample

This section contains estimates of various lognormal models applied to the 1978 estimation sample, which consists of 1540 observations. We present estimates for four types of models based on the lognormal distribution; the lognormal model, the split lognormal model, the logit lognormal model, and the logit/individual lognormal model. These models are described in detail in Schmidt and Witte, 1989, chapter 7.

### Lognormal Model

The basic lognormal model assumes that every individual would eventually fail (return to prison), and that time until failure is lognormally distributed. More specifically, the logarithm of time until failure is assumed to be distributed normally with mean $\mu_i$ and variance $\sigma^2$. Here $\mu_i$ is the mean of the logarithm of failure time for person i, and it depends linearly on a set $X_i$ of individual characteristics: $\mu_i = X_i\beta$. We estimate the parameter vector $\beta$, which tells us how the individual characteristics in $X_i$ affect the individual's (mean) survival time.

We begin by considering the same specification (set of explanatory variables) as the final specification achieved for the proportional hazards model. This is the specification of Table 12 of REPORT 1, and includes 12 explanatory variables. These explanatory variables are TSERVD, LN(AGE), PRIORS, WHITE, FELON, ALCHY, JUNKY, PROPTY, MALE, YOUNG, TSERVD*LN(AGE), AND SHORT*NOPRIOR, and are defined in REPORT 1. (The lognormal model also includes an intercept and the variance parameter $\sigma^2$.)

Table 1 gives our results for this specification of the lognormal model. The results are qualitatively quite similar to the results for the proportional hazards model, as can be seen by comparing Table 1 to Table 12 of REPORT 1. The signs of all coefficients are reversed, as they should be (since a positive effect on the hazard rate corresponds to a negative effect on the time until failure), and the t-ratios are quite similar in magnitude for most variables. This is line with the results reported in Schmidt and Witte, 1989, in which the choice of model did not much affect measures of the influence of explanatory variables on time until failure.

The coefficients of all variables (except the intercept CNST) are statistically significant at usual confidence levels. Attempts to find additional variables with statistically significant variables were unsuccessful. Thus we consider the model presented in Table 1 as our final specification of the lognormal model for the 1978 estimation sample.

The log-likelihood value of -3238.4 compares favorably with the log-likelihood value of -3273.0 reported in Table 7.7 (p. 105) of Schmidt and Witte, 1989, for their final specification of the lognormal model. That specification included nine explanatory variables, eight of which remain in our specification in Table 1; the specification in Table 1 has replaced AGE by LN(AGE) and has added the variables YOUNG, TSERVD*LN(AGE) and SHORT*NOPRIOR. A formal likelihood-ratio test comparing these specifications is not appropriate, because neither specification contains the other as a special case, but an increase of 34.6 in the likelihood by adding only three parameters to the model is a considerable improvement. Thus we conclude that we have succeeded in significantly improving the fit of the lognormal model.

## Split Lognormal Model

The split lognormal model assumes that there is a probability $\delta$ that an individual would eventually fail, and that time until failure, for those individuals who would eventually fail, is lognormally distributed. As in the lognormal model we let $\mu_i$ be the mean of the logarithm of failure time for individual i, and we let $\mu_i = X_i\beta$. Note that $\delta$ does not depend on i; every individual has the same probability of eventual failure in this model. We estimate the parameters $\beta$ and $\delta$.

Table 2 gives our results for the split lognormal model applied to the 1978 estimation sample. The specification is the same as the specification of Table 1 for the lognormal model and the specification of Table 12 of REPORT 1 for the proportional hazards model. Attempts to find additional variables with statistically significant coefficients were unsuccessful, so this is our final specification for the split lognormal model for the 1978 estimation sample.

The results in Table 2 are quite similar to those in Table 1, in the sense that most coefficients and t-ratios are quite similar. The split lognormal model fits noticeably better than the lognormal model, however; the log likelihood value increases by 2.47, a statistically significant increase. The estimated value of $\delta$, the probability of eventual failure, is significantly different from one, either by the likelihood ratio test statistic of 4.9 or by its t-ratio of 2.80. The model with the specification in Table 2 also fits considerably better than the model reported in Table 7.8 (p. 106) of Schmidt and Witte, 1989, as indicated in the increase in log likelihood from -3265.1 to -3236.0. Thus we have succeeded in significantly improving the fit of the split lognormal model.

## Logit Lognormal Model

The logit lognormal model differs from the split lognormal model in that it parameterizes the probability of eventual failure instead of the mean time until failure. The probability of eventual failure is assumed to follow a logit model with coefficients $\alpha$, as in equation (7.2), p.93 of Schmidt and Witte, 1989. However, $\mu$, the mean of the logarithm of failure time for the eventual failures, is now assumed to be the same for all individuals.

We begin with the same specification as in Tables 1 and 2. The results for this specification are given in Table 3. In this specification, the coefficient of the variable PROPTY is only marginally significant (its t-ratio is -1.76), and we therefore dropped PROPTY from our specification. This leads to our final specification for the logit lognormal model, for which results are presented in Table 4. Incidentally, comparing the log likelihood values in Tables 3 and 4 yields a likelihood ratio test statistic of 3.54 for the restriction that PROPTY should not appear. Comparing this to the 5% critical value of the chi-squared distribution with one degree of freedom (3.84) again shows that PROPTY is not very significant in the logit lognormal model.

Those variables that remain in the specification of Table 4 have coefficients that are statistically significant, and the results are very similar, in terms of the signs of the coefficients and the size of the t-ratios, to the results of Tables 1 and 2 for the lognormal and split lognormal models. It may be noted that the logit lognormal model fits the data better than the split lognormal model, in the sense that its log likelihood value is higher (-3230.2 versus -3236.0) even though the final specification contains one less parameter.

The model in Table 4 also fits the data considerably better than the logit lognormal model reported in Table 7.9 (p. 107) of Schmidt and Witte, 1989. Comparing these models, the current model contains two additional parameters. It has substituted LN(AGE) for AGE, deleted PROPTY, and added the variables YOUNG, TSERVD*LN(AGE) and SHORT*NOPRIOR. This has resulted in an increase in the log likelihood value from -3256.5 to -3230.2, a considerable increase from the addition of only two parameters. We therefore conclude that we have succeeded in significantly improving the fit of the logit lognormal model.

Logit/Individual Lognormal Model

The logit/individual lognormal model allows both the probability of eventual failure and the distribution of the time until failure to depend on individual characteristics. Thus the probability of eventual failure follows a logit model with coefficients $\alpha$, as in the logit lognormal model, while the mean of the logarithm of the failure time is normally distributed with mean $\mu_i = X_i\beta$, as in the lognormal and split lognormal models. We estimate the parameters $\alpha$ and $\beta$.

We begin with a specification that allows all twelve variables found in the specifications of Tables 1 and 2 to affect both the probability of eventual failure and the mean failure time. The results for this specification are given in Table 5. These results are broadly similar in nature to the results presented in Table 7.11 (p. 108) of Schmidt and Witte, 1989, for the simpler (nine variable) specification that omits YOUNG, TSERVD*LN(AGE), and SHORT*NOPRIOR (and uses AGE instead of LN(AGE)). In both cases, many of the variables have coefficients that are insignificantly different from zero in either the logit portion of the model (for the probability of eventual failure) or the lognormal portion of the model (for time until failure). For example, in both cases WHITE, JUNKY and MALE have insignificant coefficients in the lognormal portion of the model and PROPTY has an insignificant coefficient in the logit portio of the model. We note in passing that a comparison of log likelihood values (-3203.2 versus -3240.8) indicates that the model of Table 5 fits the data considerably better than the model presented in Table 7.11 of Schmidt and Witte, 1989.

As in our previous work, we therefore proceed to remove variables from one or both parts of the model, attempting to find a specification in which (i) the coefficients remaining in the model are significantly different from zero, and (ii) the coefficients that have been deleted from the model are jointly insignificantly different from zero. This was not a trivial undertaking, because the level of significance of some coefficients was very sensitive to the inclusion or exclusion of other coefficients. Our final specification is given in Table 6. It may be noted that the coefficient of FELON is only marginally significant in the logit equation (t-ratio equal to 1.93), but it was included because the likelihood ratio test statistic for its exclusion, 6.70, is significant at the 5% level. The specification of Table 6 contains 11 variables in the logit portion of the model and six variables in the lognormal portion. It deletes seven coefficients that were present in the specification of Table 5, and the likelihood ratio tests statistic for these seven deletions, 4.58, is very insignificant.

It is interesting to compare the results in Table 6 to the results given in Table 7.12 (p. 109) of Schmidt and Witte, 1989. They found seven variables to be significantly related to the probability of eventual failure: TSERVD, AGE, PRIORS, WHITE, ALCHY, JUNKY and MALE. All of these are also significant in the logit portion of the model in Table 5, except that AGE has been replaced by LN(AGE). However, we now find four more variables to have significant effects on the probability of eventual failure: FELON, YOUNG, TSERVD*LN(AGE), and SHORT*NOPRIOR. Similarly, they found six variables to be significantly related to the timing of failure: TSERVD, AGE, PRIORS, FELON, ALCHY and PROPTY. Our results are similar in that we find all of these variables to be significant, except that LN(AGE) replaces AGE and now has an <u>insignificant</u> coefficient, and the added variable TSERVD*LN(AGE) has a significant coefficient.

Our final specification in Table 6 therefore contains four more parameters than the specification in Table 7.12 of Schmidt and Witte, 1989. Its log likelihood value is also much larger (-3205.5 versus -3240.8), and this is a very considerable increase in likelihood from the addition of only four parameters. We therefore conclude that we have succeeded in significantly improving the fit of the logit/individual lognormal model.

## 3. Results for the 1980 Estimation Sample

This section contains estimates of lognormal models applied to the 1980 estimation sample, which consists of 1435 observations. The basic structure of our anlysis is much the same as for the 1978 sample, as described in section 2, so we will present the results of this section more concisely than we did in the last section.

### Lognormal Model

We begin with the same specification as the final specification achieved for the proportional hazards model. This is the specification of Table 30 of REPORT 1, and includes the following 14 variables: LN(AGE), PRIORS, MARRIED, JUNKY, PROPTY, YOUNG, SHORT, NOPRIOR, TSERVD*ALCHY, TSERVD*NOPRIOR, LN(AGE)*SHORT, PRIORS*PROPTY, WHITE*NOPRIOR, and MALE*NOPRIOR.

Table 7 gives our results for this specification of the lognormal model. The results are qualitatively quite similar to the results for the proportional hazards model, as would be expected. However, two variables in this specification, YOUNG and PRIORS*PROPTY, have coefficients that are insignificantly different from zero, as judged by their t-ratios of -1.53 and 1.71, respectively. The variables are however jointly significant at the 5% level; dropping them both decreases the log likelihood from -2838.8 to -2842.4, generating a likelihood ratio test statistic of 7.2, which is significant at the 5% level. The fact that the two variables are individually insignificant but jointly significant indicates that we should probably keep one but not both in the the specification. A higher log likelihood value is achieved by keeping PRIORS*PROPTY in the specification and dropping YOUNG than by doing vice-versa, and the coefficient of YOUNG is still insignificant (as judged by its t-ratio or a likelihood ratio test statistic) if it is kept in the specification while PRIORS*PROPTY is dropped. For both reasons we decided to drop YOUNG from the model. This leads to our final specification for the lognormal mdoel, the results for which are given in Table 8. Note that PRIORS*PROPTY is still only marginally significant, as judged by its t-ratio of 1.81, but the likelihood ratio test statistic (4.3) generated by dropping it from the specification is significant at the 5% level.

The results in Table 8 are quite similar to the results in Table 30 of REPORT 1 for the proportional hazards model. The signs of all coefficients are reversed, as they should be, and the t-ratios are generally of comparable magnitude.

A more interesting comparison is the one between the results of Table 8 and the results for the lognormal model given in Table 7.7 (p. 105) of Schmidt and Witte, 1989. Our current model contains 13 explanatory variables, whereas the model of Schmidt and Witte contained nine. There is some overlap between the sets of variables: both specifications contain PRIORS, MARRIED, JUNKY and PROPTY, and LN(AGE) in the current specification replaces AGE in the previous one. The current specification contains a large number of interactions not considered in Schmidt and Witte. However, as described in section 3 of REPORT 1, a verbal summary of the results of the current specification is not really very different from a summary of the results of the previous specification: the type of individual most likely to return to prison is a young, unmarried drug and alcohol abuser with many previous incarcerations and a long previous time served, and whose previous sentence was for a crime against property. The biggest difference is that in the current specification race and sex matter only for individuals with no prior incarcerations.

The model in Table 8 fits considerably better than the model of Table 7.7 of Schmidt and Witte, 1989. The log likelihood value has increased from -2868.7 to -2840.2, a large increase given the addition of four parameters. Thus we conclude that we have succeeded in significantly improving the fit of the lognormal model.

## Split Lognormal Model

We were unable to fit the split lognormal model, with any reasonable expanded specification, to the 1980 estimation sample. For each specification that we tried, the parameter ($\delta$) representing the probability of eventual failure converged to one, thus reducing the split lognormal model to the lognormal model of the previous section. This model (e.g., the model as specified in Table 8) does fit the data better than the split lognormal model reported in Table 7.8 (p. 106) of Schmidt and

Witte, 1989, so in a sense we have improved on that model, but the resulting model is not really a split model.

## Logit Lognormal Model

We begin with the same (14 variable) specification as in Table 30 of REPORT 1 or Table 7 of this report. The results for the logit lognormal model with this specification are given in Table 9. We immediately encounter the same problem that we encountered in fitting the lognormal model with this specification: the variables YOUNG and PRIORS*PROPTY have t-ratios that are insignificant at the 5% level. Furthermore, dropping both of these variables yields a likelihood ratio test statistic of 7.9, which is significant at the 5% level. Thus we should keep one but not both of the variables in the specification. Unlike the lognormal case, however, in this case it turns out to be better to drop PRIORS*PROPTY from the specification and leave in YOUNG. This yields the specification for which results are given in Table 10. Note that the coefficient of YOUNG is now significantly different from zero (t-ratio equal to -1.98).

The specification in Table 10 contains 13 variables. The results are qualitatively quite similar to the results for the 13 variable specification of the lognormal model given in Table 8. The logit lognormal model fits the data better, though not strikingly better, than the lognormal model. Its log likelihood value is higher by 1.7 (-2838.5 versus -2840.2), a moderate increase given that it includes one more parameter than the lognormal model.

The logit lognormal model with the specification of Table 10 fits the data considerably better than the logit lognormal model presented in Table 7.9 (p. 109) of Schmidt and Witte, 1989. That specification included nine explanatory variables, and the 13 variable specification of Table 10 results in a considerable increase in log likelihood (from -2853.1 to -2838.5). Thus we conclude that we have succeeded in significantly improving the fit of the logit lognormal model.

Logit/Individual Lognormal Model

We begin with a specification that allows all fourteen variables found in the specification of Table 7 (or Table 30 of REPORT 1) to affect both the probability of eventual failure and the mean failure time. The results for this specification are given in Table 11. These results are rather confusing. Virtually no coefficients in the lognormal portion of the model are significantly different from zero, as judged by their t-ratios. Schmidt and Witte, 1989 found more or less the same thing; see Table 7.11, p. 108. The parameters of the logit portion of the model also suffer from a lack of significance, though not to the same extent, and there is no clear relationship between what is significant here and what was found to be significant in Schmidt and Witte, 1989.

As in our previous work, we therefore proceed to remove variables from one or both parts of the model, attempting to find a specification in which (i) the coefficients remaining in the model are significantly different from zero, and (ii) the coefficients that have been deleted from the model are jointly insignificantly different from zero. Once again this was not a trivial undertaking, because the level of significance of some coefficients was very sensitive to the inclusion or exclusion of other coefficients. Our final specification is given in Table 12. It contains eleven variables in the logit portion of the model and six variables in the lognormal portion, and therefore differs from the specification of Table 11 by setting eleven parameters to zero. The likelihood ratio test statistic for these eleven deletions, 15.4, is quite insignificant, indicating that the coefficients that were removed from the model were jointly (as well as individually) insignificant. Several of the variables still left in the model have coefficients that are not very significant (YOUNG in the logit model has a t-ratio of -1.58, and SHORT and NOPRIOR in the lognormal model have t-ratios of 1.71 and 1.86). They were left in the model because they are jointly significant, and dropping them individually did not improve the level of significance of the other two marginally significant coefficients.

Compared to the final specification for the logit/individual lognormal model given by Schmidt and Witte, 1989 (Table 7.13, p. 110), the specification of Table 12 is a clear improvement. There are eleven variables in the logit portion of the model here, instead of seven, and six variables in the lognormal portion of the model, instead of two. Furthermore, the log likelihood is improved from -2850.7 to -2826.9, a large increase. Thus we conclude that we have succeeded in significantly improving the fit of the logit/individual lognormal model.

## 4. Concluding Remarks and Further Research

In the work performed under our previous grant, and reported in Schmidt and Witte, 1989, an enormous amount of effort was put into finding appropriate distributions to use for a failure time model of time until return to prison. We estimated parametric models based on the exponential, Weibull, lognormal, log-logistic and LaGuerre distributions. We considered models in which everyone is assumed to fail eventually as well as split models in which some individuals are assumed never to fail, and we allowed explanatory variables to affect both the probability of eventual failure and the distribution of failure times for those individuals who would eventually fail. We established convincingly the superiority (for our data sets) of a model in which not everyone is an eventual failure, the probability of eventual failure varies across individuals according to a logit model, and the distribution of failure times for those who will eventually fail is lognormal with a mean that also varies across individuals. This model is the logit/individual lognormal model.

However, in our earlier work we did not experiment very much with the way that explanatory variables entered the model. They were just entered linearly. The point of our current work is to see how much our survival models can be improved by a more thorough and careful consideration of ways to enter explanatory variables into the models. We have considered transformations of variables, such as replacing AGE by LN(AGE) and YOUNG, and we have also considered numerous interactions between explanatory variables.

The work described in REPORT 1 used the proportional hazards model to consider a very large number of different combinations of explanatory variables and their transformations and interactions. We found that the data would support more complex specifications than those used by Schmidt and Witte, 1989. For the 1978 estimation sample, we found 12 variables significantly related to time until failure, instead of nine; for the 1980 estimation sample, we found 14 such variables instead of nine. These additions to the specification resulted in reasonable improvements in fit, as measured by the maximized value of the log likelihood function.

The work described in this report extends this work to parametric models based on the lognormal distribution; that is, to the logit/individual lognormal model and its variants. The results differ slightly across models and across samples, but speaking generally we found that the data would support about the same degree of added complexity for the specification of explanatory variables in these models as in the proportional hazards model. For example, for the 1978 logit/individual lognormal model, we now have eleven variables in the logit portion of the model, whereas Schmidt and Witte had seven; and we now have six variables in the lognormal portion of the model, the same number as before. For the 1980 logit/individual lognormal model, we now have eleven variables in the logit portion of the model, whereas Schmidt and Witte had seven; and we now have six variables in the lognormal portion of the model, whereas they had only two.

The expanded models fit the data better than the less complex models of Schmidt and Witte, 1989, and this is clear by comparing log likelihood values. However, since log likelihood values are not very easy to interpret, it may be useful to illustrate the improvement in fit in a more intuitive way. In a linear regression model, the most commonly used measure of goodness of fit is $R^2$, which indicates the proportion of the variance of the dependent variable explained by the independent variables. There is no $R^2$ for nonlinear models like the models of this report. However, Schmidt and Witte, 1989 (p. 117, footnote 9) define an "$R^2$ equivalent" by calculating the level of joint significance of the variables in the model, and asking what $R^2$ would be necessary in a linear regression model to yield the same level of significance. For example, they report an $R^2$ equivalent of .10 for their 1978 logit lognormal model and of .12 for their 1980 logit lognormal model.

Calculating the $R^2$ equivalents for the models of this report reveals a moderate increase in fit relative to the results of Schmidt and Witte, 1989. For 1978, the changes (from the specifications of Schmidt and Witte, 1989 to the final specifications of this report) in $R^2$ equivalent are as follows: lognormal model, .112 to .147; split lognormal model, .092 to .122; logit lognormal model, .101 to .128; logit/individual lognormal model, .117 to .152. Similarly, for 1980 we have the following changes in $R^2$ equivalent: lognormal model, .123 to .151; logit lognormal model, .108 to .141; logit/individual lognormal model, .128 to .153. As a broad statement, we have increased the proportion of the variation in time until recidivism that is explained by our models by about 30%. This is a sizeable increase, even though it is still clearly the case that we explain only a small fraction of the variation in time until recidivism.

In any case, our primary interest is not in explaining outcomes in our estimation samples, but rather in predicting outcomes in other independent samples, such as our validation samples. The extent to which our newly specified models lead to better predictions of recidivism, for individuals and for groups of individuals, will be considered in our next report.

## TABLE 1

### LOGNORMAL MODEL - 1978 DATA

### 12 VARIABLES: FINAL SPECIFICATION

### LOG-LIKELIHOOD: -3238.45

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| TSERVD | -21.0208 | 4.6348 | -4.54 |
| LN(AGE) | 0.7505 | 0.2745 | 2.73 |
| PRIORS | -0.9346 | 0.1880 | -4.97 |
| WHITE | 0.5311 | 0.1166 | 4.56 |
| FELON | 0.8263 | 0.1980 | 4.17 |
| ALCHY | -0.5447 | 0.1365 | -3.99 |
| JUNKY | -0.3835 | 0.1343 | -2.85 |
| PROPTY | -0.4580 | 0.1909 | -2.40 |
| MALE | -0.7121 | 0.2517 | -2.83 |
| YOUNG | -0.7920 | 0.1951 | -4.06 |
| TSERVD*LN(AGE) | 3.2599 | 0.7536 | 4.33 |
| SHORT*NOPRIOR | 0.6951 | 0.1402 | 4.96 |
| CNST | 1.2621 | 1.6289 | 0.77 |
| SIGMA | 1.7567 | 0.0663 | 26.48 |

## TABLE 2

### SPLIT LOGNORMAL MODEL - 1978 DATA

### 12 VARIABLES: FINAL SPECIFICATION

### LOG-LIKELIHOOD: -3235.97

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| TSERVD | -19.0517 | 4.6622 | -4.09 |
| LN(AGE) | 0.7743 | 0.2764 | 2.80 |
| PRIORS | -1.1685 | 0.2661 | -4.39 |
| WHITE | 0.5204 | 0.1136 | 4.58 |
| FELON | 0.8062 | 0.1863 | 4.33 |
| ALCHY | -0.5802 | 0.1405 | -4.13 |
| JUNKY | -0.3683 | 0.1286 | -2.86 |
| PROPTY | -0.4810 | 0.1783 | -2.70 |
| MALE | -0.6717 | 0.2580 | -2.60 |
| YOUNG | -0.7479 | 0.1841 | -4.06 |
| TSERVD*LN(AGE) | 2.9525 | 0.7798 | 3.79 |
| SHORT*NOPRIOR | 0.6335 | 0.1458 | 4.35 |
| CNST | 0.6773 | 1.6578 | 0.41 |
| DELTA | 0.8082 | 0.0684 | 11.81 |
| SIGMA | 1.5572 | 0.0920 | 16.92 |

TABLE 3

LOGIT LOGNORMAL MODEL - 1978 DATA

12 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -3228.47

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| TSERVD | -31.7528 | 7.1001 | -4.47 |
| LN(AGE) | 0.6948 | 0.3418 | 2.03 |
| PRIORS | -1.2764 | 0.3856 | -3.31 |
| WHITE | 0.6968 | 0.1426 | 4.89 |
| FELON | 0.8957 | 0.2537 | 3.53 |
| ALCHY | -0.6058 | 0.1779 | -3.40 |
| JUNKY | -0.5050 | 0.1635 | -3.09 |
| PROPTY | -0.4270 | 0.2431 | -1.76 |
| MALE | -0.8089 | 0.2938 | -2.75 |
| YOUNG | -0.9426 | 0.2467 | -3.82 |
| TSERVD*LN(AGE) | 4.9622 | 1.1317 | 4.38 |
| SHORT*NOPRIOR | 0.6929 | 0.1788 | 3.87 |
| CNST | -2.8007 | 2.0470 | -1.37 |
| MU | 3.1589 | 0.0842 | 37.51 |
| SIGMA | 1.1641 | 0.0511 | 22.79 |

TABLE 4

LOGIT LOGNORMAL MODEL - 1978 DATA

11 VARIABLES: FINAL SPECIFICATION

LOG-LIKELIHOOD: -3230.24

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| TSERVD | -33.4600 | 7.1130 | -4.70 |
| LN(AGE) | 0.7101 | 0.3413 | 2.08 |
| PRIORS | -1.3453 | 0.3878 | -3.47 |
| WHITE | 0.6942 | 0.1425 | 4.87 |
| FELON | 0.6207 | 0.1859 | 3.34 |
| ALCHY | -0.5904 | 0.1774 | -3.33 |
| JUNKY | -0.4912 | 0.1631 | -3.01 |
| MALE | -0.8153 | 0.2939 | -2.77 |
| YOUNG | -0.9321 | 0.2471 | -3.77 |
| TSERVD*LN(AGE) | 5.2470 | 1.1322 | 4.63 |
| SHORT*NOPRIOR | 0.6906 | 0.1790 | 3.86 |
| CNST | -2.8991 | 2.0449 | -1.42 |
| MU | 3.1624 | 0.0845 | 37.43 |
| SIGMA | 1.1662 | 0.0513 | 22.72 |

TABLE 5

LOGIT/INDIVIDUAL LOGNORMAL MODEL - 1978 DATA

12+12 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -3203.19

| | LOGIT MODEL | | | LOGNORMAL MODEL | | |
|---|---|---|---|---|---|---|
| VARIABLE | ESTIMATE | S.E. | T-RATIO | ESTIMATE | S.E. | T-RATIO |
| TSERVD | -22.6560 | 7.1545 | -3.17 | -10.9227 | 6.3621 | -1.72 |
| LN(AGE) | 0.6843 | 0.3794 | 1.80 | 0.1109 | 0.3286 | 0.34 |
| PRIORS | -0.5768 | 0.2497 | -2.31 | -0.6969 | 0.1755 | -3.97 |
| WHITE | 0.6400 | 0.1534 | 4.17 | 0.0551 | 0.1307 | 0.42 |
| FELON | 0.5463 | 0.2874 | 1.90 | 0.5773 | 0.2337 | 2.47 |
| ALCHY | -0.3955 | 0.1799 | -2.20 | -0.3179 | 0.1489 | -2.14 |
| JUNKY | -0.4802 | 0.1761 | -2.73 | -0.0253 | 0.1436 | -0.18 |
| PROPTY | -0.1726 | 0.2696 | -0.64 | -0.4373 | 0.2077 | -2.11 |
| MALE | -0.8248 | 0.3833 | -2.15 | -0.0346 | 0.5749 | -0.06 |
| YOUNG | -0.7985 | 0.2679 | -2.98 | -0.2924 | 0.2096 | -1.40 |
| TSERVD*LN(AGE) | 3.5491 | 1.1742 | 3.02 | 1.6852 | 1.0855 | 1.55 |
| SHORT*NOPRIOR | 0.7476 | 0.1919 | 3.90 | 0.0928 | 0.1583 | 0.59 |
| CNST | -3.0261 | 2.2833 | -1.33 | 2.9588 | 1.9967 | 1.48 |
| SIGMA | | | | 1.1031 | 0.0477 | 23.11 |

## TABLE 6

### LOGIT/INDIVIDUAL LOGNORMAL MODEL - 1978 DATA

### 11+6 VARIABLES: FINAL SPECIFICATION

### LOG-LIKELIHOOD: -3205.48

| | LOGIT MODEL | | | LOGNORMAL MODEL | | |
|---|---|---|---|---|---|---|
| VARIABLE | ESTIMATE | S.E. | T-RATIO | ESTIMATE | S.E. | T-RATIO |
| TSERVD | -22.1331 | 7.2371 | -3.06 | -12.5721 | 4.9487 | -2.54 |
| LN(AGE) | 0.7517 | 0.3385 | 2.22 | | | |
| PRIORS | -0.6107 | 0.2466 | -2.48 | -0.6552 | 0.1574 | -4.16 |
| WHITE | 0.6700 | 0.1413 | 4.74 | | | |
| FELON | 0.3829 | 0.1985 | 1.93 | 0.6855 | 0.2101 | 3.26 |
| ALCHY | -0.4022 | 0.1771 | -2.27 | -0.2920 | 0.1406 | -2.08 |
| JUNKY | -0.4914 | 0.1610 | -3.05 | | | |
| PROPTY | | | | -0.5004 | 0.1823 | -2.75 |
| MALE | -0.8337 | 0.2976 | -2.80 | | | |
| YOUNG | -0.9755 | 0.2595 | -3.76 | | | |
| TSERVD*LN(AGE) | 3.4745 | 1.1920 | 2.91 | 1.9503 | 0.8428 | 2.31 |
| SHORT*NOPRIOR | 0.8054 | 0.1715 | 4.70 | | | |
| CNST | -3.4364 | 2.0262 | -1.70 | 3.5712 | 0.1280 | 27.90 |
| SIGMA | | | | 1.1104 | 0.0471 | 23.59 |

TABLE 7

LOGNORMAL MODEL - 1980 DATA

14 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -2838.79

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| LN(AGE) | 2.4863 | 0.4163 | 5.97 |
| PRIORS | -1.5252 | 0.3126 | -4.88 |
| MARRIED | 0.3743 | 0.1356 | 2.76 |
| JUNKY | -0.3324 | 0.1275 | -2.61 |
| PROPTY | -0.4449 | 0.1205 | -3.69 |
| YOUNG | -0.3907 | 0.2558 | -1.53 |
| SHORT | 7.0257 | 2.5593 | 2.75 |
| NOPRIOR | 1.5173 | 0.3777 | 4.02 |
| TSERVD*ALCHY | -1.5674 | 0.5124 | -3.06 |
| TSERVD*NOPRIOR | -1.0465 | 0.5325 | -1.97 |
| LN(AGE)*SHORT | -1.1640 | 0.4390 | -2.65 |
| PRIORS*PROPTY | 0.5963 | 0.3480 | 1.71 |
| WHITE*NOPRIOR | 0.5054 | 0.1391 | 3.63 |
| MALE*NOPRIOR | -1.1598 | 0.3442 | -3.37 |
| CNST | -9.9292 | 2.4083 | -4.12 |
| SIGMA | 1.5872 | 0.0609 | 26.07 |

TABLE 8

LOGNORMAL MODEL - 1980 DATA

13 VARIABLES: FINAL SPECIFICATION

LOG-LIKELIHOOD: -2840.24

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| LN(AGE) | 2.5239 | 0.4171 | 6.05 |
| PRIORS | -1.5708 | 0.3122 | -5.03 |
| MARRIED | 0.3805 | 0.1359 | 2.80 |
| JUNKY | -0.3194 | 0.1270 | -2.51 |
| PROPTY | -0.4638 | 0.1206 | -3.85 |
| SHORT | 6.4752 | 2.5235 | 2.57 |
| NOPRIOR | 1.5348 | 0.3793 | 4.05 |
| TSERVD*ALCHY | -1.5669 | 0.5123 | -3.06 |
| TSERVD*NOPRIOR | -1.0197 | 0.5312 | -1.92 |
| LN(AGE)*SHORT | -1.0709 | 0.4329 | -2.47 |
| PRIORS*PROPTY | 0.6307 | 0.3479 | 1.81 |
| WHITE*NOPRIOR | 0.4968 | 0.1386 | 3.58 |
| MALE*NOPRIOR | -1.1965 | 0.3457 | -3.46 |
| CNST | -10.1369 | 2.4128 | -4.20 |
| SIGMA | 1.5888 | 0.0611 | 26.00 |

TABLE 9

LOGIT LOGNORMAL MODEL - 1980 DATA

14 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -2837.06

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| LN(AGE) | 3.1542 | 0.6335 | 4.98 |
| PRIORS | -2.4263 | 0.6027 | -4.03 |
| MARRIED | 0.4633 | 0.1852 | 2.50 |
| JUNKY | -0.4531 | 0.1908 | -2.38 |
| PROPTY | -0.5956 | 0.1778 | -3.35 |
| YOUNG | -0.7545 | 0.4012 | -1.88 |
| SHORT | 8.4742 | 3.9052 | 2.17 |
| NOPRIOR | 1.8312 | 0.5370 | 3.41 |
| TSERVD*ALCHY | -2.3537 | 0.8871 | -2.65 |
| TSERVD*NOPRIOR | -1.6699 | 0.7082 | -2.36 |
| LN(AGE)*SHORT | -1.3960 | 0.6675 | -2.09 |
| PRIORS*PROPTY | 1.1238 | 0.6635 | 1.69 |
| WHITE*NOPRIOR | 0.6926 | 0.1964 | 3.53 |
| MALE*NOPRIOR | -1.4864 | 0.4939 | -3.01 |
| CNST | -17.9161 | 3.7098 | -4.83 |
| MU | 2.9855 | 0.0871 | 34.29 |
| SIGMA | 1.1117 | 0.0543 | 20.48 |

TABLE 10

LOGIT LOGNORMAL MODEL - 1980 DATA

13 VARIABLES: FINAL SPECIFICATION

LOG-LIKELIHOOD: -2838.51

| VARIABLE | ESTIMATE | S.E. | T-RATIO |
|---|---|---|---|
| LN(AGE) | 3.2814 | 0.6384 | 5.14 |
| PRIORS | -2.0588 | 0.4712 | -4.37 |
| MARRIED | 0.4522 | 0.1860 | 2.43 |
| JUNKY | -0.4292 | 0.1924 | -2.23 |
| PROPTY | -0.4693 | 0.1618 | -2.90 |
| YOUNG | -0.7989 | 0.4040 | -1.98 |
| SHORT | 9.1857 | 3.9356 | 2.33 |
| NOPRIOR | 1.8057 | 0.5373 | 3.36 |
| TSERVD*ALCHY | -2.3463 | 0.8961 | -2.62 |
| TSERVD*NOPRIOR | -1.7047 | 0.7146 | -2.39 |
| LN(AGE)*SHORT | -1.5233 | 0.6724 | -2.27 |
| WHITE*NOPRIOR | 0.7058 | 0.1966 | 3.59 |
| MALE*NOPRIOR | -1.4764 | 0.4933 | -2.99 |
| CNST | -18.6873 | 3.7373 | -5.00 |
| MU | 2.9973 | 0.0878 | 34.13 |
| SIGMA | 1.1186 | 0.0548 | 20.39 |

TABLE 11

LOGIT/INDIVIDUAL LOGNORMAL MODEL - 1980 DATA

14+14 VARIABLES: ORIGINAL SPECIFICATION

LOG-LIKELIHOOD: -2819.18

| | LOGIT MODEL | | | LOGNORMAL MODEL | | |
|---|---|---|---|---|---|---|
| VARIABLE | ESTIMATE | S.E. | T-RATIO | ESTIMATE | S.E. | T-RATIO |
| LN(AGE) | 2.5515 | 0.8038 | 3.17 | 1.1205 | 0.5388 | 2.08 |
| PRIORS | -2.1568 | 0.6030 | -3.58 | -0.3994 | 0.2642 | -1.51 |
| MARRIED | 0.3004 | 0.2504 | 1.20 | 0.2280 | 0.2028 | 1.12 |
| JUNKY | -0.3466 | 0.2217 | -1.56 | -0.1414 | 0.1467 | -0.96 |
| PROPTY | -0.5156 | 0.2265 | -2.28 | -0.1481 | 0.1551 | -0.95 |
| YOUNG | -1.0771 | 0.7315 | -1.47 | 0.1064 | 0.2891 | 0.37 |
| SHORT | 5.9537 | 4.9625 | 1.20 | 4.6099 | 3.2741 | 1.41 |
| NOPRIOR | 1.4658 | 0.7189 | 2.04 | 0.6364 | 0.5999 | 1.06 |
| TSERVD*ALCHY | -1.7821 | 0.9200 | -1.94 | -0.8875 | 0.4592 | -1.93 |
| TSERVD*NOPRIOR | -1.4531 | 0.9270 | -1.57 | -0.4066 | 0.5697 | -0.71 |
| LN(AGE)*SHORT | -0.9499 | 0.8525 | -1.11 | -0.8012 | 0.5675 | -1.41 |
| PRIORS*PROPTY | 1.1611 | 0.6600 | 1.76 | -0.0307 | 0.3259 | -0.09 |
| WHITE*NOPRIOR | 0.6435 | 0.2749 | 2.34 | 0.1346 | 0.1929 | 0.70 |
| MALE*NOPRIOR | -1.3465 | 0.6820 | -1.97 | -0.3038 | 0.5734 | -0.53 |
| CNST | -14.5458 | 4.6742 | -3.11 | -3.3348 | 3.0681 | -1.09 |
| SIGMA | | | | 1.1136 | 0.0599 | 18.60 |

TABLE 12

LOGIT/INDIVIDUAL LOGNORMAL MODEL - 1980 DATA

11+6 VARIABLES: FINAL SPECIFICATION

LOG-LIKELIHOOD: -2826.87

| | LOGIT MODEL | | | LOGNORMAL MODEL | | |
|---|---|---|---|---|---|---|
| VARIABLE | ESTIMATE | S.E. | T-RATIO | ESTIMATE | S.E. | T-RATIO |
| LN(AGE) | 1.8657 | 0.4130 | 4.52 | 0.5753 | 0.2893 | 1.99 |
| PRIORS | -2.1346 | 0.5996 | -3.56 | -0.4182 | 0.2297 | -1.82 |
| MARRIED | | | | 0.4061 | 0.1482 | 2.74 |
| JUNKY | -0.4710 | 0.1977 | -2.38 | | | |
| PROPTY | -0.6514 | 0.1902 | -3.42 | | | |
| YOUNG | -0.7232 | 0.4578 | -1.58 | | | |
| SHORT | | | | 0.2275 | 0.1330 | 1.71 |
| NOPRIOR | 1.8954 | 0.5573 | 3.40 | 0.3017 | 0.1622 | 1.86 |
| TSERVD*ALCHY | -2.0173 | 0.8640 | -2.33 | -0.7499 | 0.3167 | -2.37 |
| TSERVD*NOPRIOR | -2.4461 | 0.8414 | -2.91 | | | |
| PRIORS*PROPTY | 1.2370 | 0.6145 | 2.01 | | | |
| WHITE*NOPRIOR | 0.7278 | 0.2160 | 3.37 | | | |
| MALE*NOPRIOR | -1.5893 | 0.5136 | -3.09 | | | |
| CNST | -10.1640 | 2.4172 | -4.20 | -0.4523 | 1.6348 | -0.28 |
| SIGMA | | | | 1.1316 | 0.0591 | 19.15 |

# PREDICTIONS FROM PROPORTIONAL HAZARDS AND PARAMETRIC MODELS, NORTH CAROLINA DATA

Ching-Fan Chung, Peter Schmidt and Ann Witte

Report 3 of results under NIJ grant 89-IJ-CX-0010

## 1. Introduction

The work described in this report is a continuation of the work described in our previous reports, "Additional Results on Functional Form, Proportional Hazards Model, North Carolina Data," September, 1989, hereafter called REPORT 1; and "Results on Functional Form, Parametric (Lognormal) Models, North Carolina Data," March, 1990, hereafter called REPORT 2. This work is an extension of the analyses performed under our previous grant (84-IJ-CX-0021) and reported in P. Schmidt and A.D. Witte, Predicting Recidivism Using Survival Models, Springer-Verlag, 1989, hereafter called Schmidt and Witte, 1989.

REPORT 1 dealt with the proportional hazards model. It investigated ways to improve the models of Schmidt and Witte, 1989, by introducing explanatory variables into the proportional hazards model in nonlinear ways. Specifically, logarithmic transformations of certain variables and interactions between variables were found to improve the fit of the model significantly, for both the 1978 and 1980 estimation samples.

REPORT 2 dealt with parametric models based on the lognormal distribution, which were found in Schmidt and Witte, 1989 to predict better than the proportional hazards model and other parametric models. Specifically, it considered four types of models based on the lognormal distribution. (1) The lognormal model assumes that the probability of eventual recidivism is one, and it allows the mean time until recidivism to depend on explantory variables. (2) The split lognormal model assumes that the probability of eventual recidivism is the same for all individuals, though not necessarily equal to one, and it allows the mean time until recidivism to depend on explantory variables. (3) The logit lognormal model allows the probability of eventual recidivism to depend on explanatory variables, but it assumes that the mean time until recidivism is the same for all individuals. (4) The logit/individual lognormal model allows both the probability of eventual recidivism and the mean time until recidivism to depend on explanatory variables. For each type of model, the fit of the model was improved significantly using the same types of transformations of variables and interactions as were used to improve the fit of the proportional hazards model. While the details varied across models and across the 1978 and 1980 data sets, as a rough statement we were able to improve the $R^2$ equivalent (which we used to measure goodness of fit) from about 0.12 to about 0.15. This corresponds to an improvement in explanatory power of about 25 to 30%.

In this report we investigate the way in which this improvement in explanatory power translates into improvements in out-of-sample predictions. As in Schmidt and Witte, 1989, we consider four types of predictions. First, we consider predictions for our entire validation sample; that is, for essentially a random sample of releasees. Second, we consider predictions of the recidivism behavior of the 1980 validation sample using models fit to the 1978 estimation sample; that is, predictions across different years of release. Third, we consider predictions for various interesting subsamples of releasees, such as youthful offenders or felons. Fourth, we consider predictions for individuals. These four types of predictions will be reported in sections two through five of this report, while section six will give our concluding remarks.

## 2. Predictions for the 1978 and 1980 Validation Samples

For both our 1978 and 1980 data sets, the sample of individuals with complete records was split randomly into an estimation sample, which was used to fit models, and a validation sample, which was used to evaluate predictive performance. In this section, we consider using models fit to the 1978 and 1980 estimation samples to make predictions for the corresponding entire validation samples. These are very large samples: 3078 individuals for 1978, and 4304 individuals for 1980.

The predictions we consider are <u>out of sample</u> predictions, since the individuals in the validation samples, for whom we are making predictions, were not in the estimation samples. Out of sample prediction is generally considered to be a more severe test of a model's adequacy than is its within-sample goodness of fit. Since we have engaged in a concerted effort to find models that maximize within-sample goodness of fit, it is reasonable to worry whether we have <u>overfit</u> the model; that is, whether we have improved the goodness of fit by modeling peculiar random features of the estimation samples that we would not expect to find in other similar samples, such as the validation samples. We therefore wish to see whether our expanded specifications do indeed lead to better predictions than the original specifications considered in Schmidt and Witte, 1989.

Table 1 gives summary statistics for our predictions for the 1978 and 1980 validation samples. We will ignore for the moment the results in the last two columns, labeled "Use 1978 to predict 1980," which will be discussed in the next section, and focus on the results for 1978 (i.e., models fit to the 1978 estimation sample and used to make predictions for the 1978 validation sample) and 1980. The variable that is being predicted is the number of individuals in the sample that return to prison in each month from release until the end of the followup period. The method of making these predictions is described in some detail in Schmidt and Witte, 1989, Chapter 3, pp. 38-47, in the section "Predictions Using Survival Time Models." We will use two statistics to summarize the quality of our predictions. The first is a chi-squared statistic, labeled $\chi^2$, which is the standard statistic based on the expected versus observed numbers of observations in various "cells." In our case the cells correspond to months after release, and the number of cells is approximately the number of months after release for which we make predictions (71 for 1978, and 47 for 1980); the word approximately is used because we combine some months after release to ensure at least five observations per cell. The expected numbers of observations in a given cell are those that are predicted by the estimated model. For more details, see Schmidt and Witte, 1989, pp. 43-44. The second statistic, labeled KS, is the maximum difference between the actual and the predicted cumulative distribution function (cumulative recidivism rate). In an uncensored sample, this would be the Kolmogorov-Smirnov statistic for goodness of fit. For more details, see Schmidt and Witte, 1989, p. 46.

The first eight lines of Table 1 are taken from Schmidt and Witte, 1989, Table 3.2, p. 45. The first three lines give their results for predictions made without using any explanatory variables; the nonparametric "model" just predicts the same failure rate in the validation sample as in the estimation sample, while the lognormal and split lognormal models are the same as the models described in REPORT2 except that they do not contain any explanatory variables (the distribution of time until recidivism is taken to be the same for each individual). As described in Schmidt and Witte, 1989, Chapter 5, the split lognormal model without explanatory variables gives quite good predictions for the validation samples; see especially Figure 5.1 and Tables 5.1 and 5.2, pp. 71-77.

The next five lines of Table 1 give the results of predictions using models with the set of explanatory variables of Schmidt and Witte, 1989. These models do not predict much better (or differently) than the corresponding models without explanatory variables. The logit lognormal model seems to give the best predictions, though the logit/individual lognormal model also predicts well, and either model really yields more or less the same predictions, month by month, as the split lognormal model without explanatory variables.

Finally, the last five lines give the summary statistics for the predictions from the proportional hazards model and the models based on the lognormal distribution, using the expanded specifications described in REPORT1 and REPORT2. For both 1978 and 1980, the logit lognormal model and the logit/individual lognormal model give very similar predictions. The use of an expanded specification instead of the specification of Schmidt and Witte, 1989, has resulted in very minor improvements in predictive ability, but these models still do not give better predictions than the split lognormal model without explantory variables.

These results are not surprising. Since the estimation and validation samples were generated by a random split of the sample of complete observations, there should be no differences between the estimation and validation samples (for a given year of release) except those due solely to the randomness of the sample division. The main point of explanatory variables is to correct for differences across groups, and there should be no differences to correct for between the estimation sample and the validation sample for a given year of release. The real test of the usefulness of our expanded specifications will be predicting outcomes for <u>nonrandom</u> samples of releasees, or for individuals.

Table 2 gives a detailed (month by month) listing of the predicted and actual outcomes for the 1978 validation sample, with the predictions coming from the logit/individual lognormal model. The format of this table is identical to the format of the tables in Schmidt and Witte, 1989, and a detailed description of the meaning of the entries of the table is given on their pp. 39-43. Briefly, the first column (with the heading "N") lists months after release, with N=0 corresponding to the first 15 days after release, N=1 corresponding to more than 15 days but less than one month and 15 days, etc. The next two columns, with the headings "Predicted cdf" and "cdf," give the predicted and actual values of the cumulative distribution function, which represents the proportion of the original cohort that fails no later than (i.e., before or during) the month in question (month N). The next column, with heading "<u>t</u>", gives the t-statistic used to test the significance of the difference between the actual and the predicted cdf. Under the null hypothesis that the model is correct, the t-statistic should be distributed as standard normal, for a large sample (which we believe we have). The next three columns give the predicted density, the actual density, and the t-statistic used to test the significance of the difference between the actual and the predicted density, where the density corresponds to the proportion of the original cohort that fails during month N. The next three columns are similar, except that they deal with the hazard instead of the density; the hazard corresponds to the proportion of the surviving cohort that fails during month N. (That is, the hazard for month N is the ratio of the number of failures during month N to the number of individuals who survived at least through month N-1.) Finally, the last column gives the $\chi^2$ statistic for goodness of fit. For each month, the change in the $\chi^2$ statistic is the square of the t-statistic for the difference between the actual and predicted hazards, and the last value of the $\chi^2$ statistic is the summary statistic reported in Table 1.

A glance at Table 2 shows that the logit/indidividual lognormal model predicts the recidivism behavior of the 1978 validation sample quite well. There are very few significant differences between actual and predicted outcomes. Although we do not display them here, the predictions given by the logit lognormal model would be very similar to the predictions in Table 2. For that matter, the predictions given in Schmidt and Witte, 1989, Table 7.14, p. 112, which correspond to the logit lognormal model with the Schmidt-Witte set of explanatory variables, are also very similar to those in Table 2; for example, the predicted cdf's differ by at most 0.003.

Table 3 is essentially the same as Table 2 except that it gives results for 1980 instead of 1978. The predictions for 1980 are less satisfactory than for 1978. We tend to underpredict recidivism in the 1980 validation sample, no matter what model is used, because the recidivism rate in the 1980 validation sample is higher than the recidivism rate in the 1980 estimation sample, due simply to bad luck in the random split of the 1980 data into its two subsamples. Nevertheless, the predictions in Table 3 are good enough to be satisfactory for most purposes. Furthermore, they are virtually indistinguishable from the predictions from other models, such as the logit lognormal model or even the logit lognormal model with the Schmidt-Witte set of explanatory variables.

These comparisons can be considered as an elaboration of the basic point, made three paragraphs above on the basis of summary statistics, that even fairly considerable expansions of our models do not result in very different predictions, for samples (like the entire validation sample) that do not differ systematically from the samples used to estimate the model. Different models do indeed lead to very different predictions, but different sets of explanatory variables for the same model do not.

## 3. Predictions Across Release Cohorts

In this section, we use our models of time until recidivism, estimated using data from the 1978 estimation sample, to make predictions of the rate of recidivism in the 1980 validation sample. This is an interesting exercise because practical uses of models like ours (for example, for program evaluations or population projections) will rely on the accuracy of predictions for cohorts of releasees different than the cohort used to estimate the models. This is true because our models are too difficult and expensive to estimate for it to be reasonable to reestimate them every year, and also because we may wish to make predictions for a cohort without waiting for a long enough followup period to allow precise estimation of our models.

The recidivism rate in the 1980 release cohort is considerably higher than the recidivism rate in the 1978 cohort, so we can expect at the outset that models fit to the 1978 data will tend to underpredict the rate of recidivism in the 1980 data. However, we can hope that some or all of the difference in recidivism rates can be explained by differences in the values of explanatory variables (individual characteristics and correctional histories) across the cohorts.

Schmidt and Witte, 1989 (Table 3.2, p. 45) report summary statistics for the predictions for the 1980 validation sample generated by various models fit to the 1978 validation sample. Their results varied considerably across models. For almost all models, the predictions for 1980 were much better when the model was estimated using 1980 data than when the models were estimated using 1978 data, an unsurprising result. However, for the logit/individual lognormal model, they found the interesting result that the predictions for the 1980 validation sample were almost equally good whether the model was fit to the 1980 estimation sample or to the 1978 estimation sample. In other words, with the best of their models, the observed differences in recidivism rates between 1978 and 1980 are essentially all explained by differences in values of the explanatory variables, and the model predicts well at least across a two year difference in year of release.

The corresponding results for the proportional hazards model and for our models based on the lognormal distribution, with the expanded specifications developed in REPORT1 and REPORT2, are given in the last two columns of Table 1. The results are similar to the results of Schmidt and Witte, 1989, in two regards. First, for all models except the logit/individual lognormal model, estimates based on the 1978 data do not predict outcomes for 1980 nearly as well as estimates based on the 1980 data. Second, for the logit/individual lognormal model, the predictions based on the 1978 estimates are almost (though not quite) as good as the predictions based on the 1980 estimates, so that once again the observed differences in recidivism rates between 1978 and 1980 are essentially all explained by differences in values of the explanatory variables.

Finally, we can compare the results for the logit/individual lognormal model with its expanded specification, as given in Table 1 ($\chi^2$ = 63.3, KS = .031) to the corresponding results for the smaller specification of Schmidt and Witte, 1989, as given in their Table 3.2 ($\chi^2$ = 66.3, KS = .029). A more detailed listing of our predictions for 1980, using the 1978 logit/individual lognormal model, can be found in Table 4, and these results can be compared to the results in Schmidt and Witte, 1989, Table 8.1, pp. 126-127. The results are clearly not very different, so that once again we find that our expanded specifications do not lead to improved predictions for samples (like an entire validation sample) that are essentially random samples of releasees.

## 4. Predictions for Non-Random Samples

In this section we consider predictions for nonrandom subsamples of release cohorts. Prediction for nonrandom samples is important because many practical uses of our models, such as program evaluation, require that the models be used to make predictions for groups that are not random samples of releasees. The function of the model is to correct for differences between the group in question and the larger population of releasees, and so it is important that it do so adequately. For example, in the context of program evaluation, if differences between the treated group and the rest of the population that are due to differences in individual characteristics are not adequately captured by the model, they will be attributed to the program, and this may result in a seriously biased estimate of its effect.

Our analysis is similar to the analysis of Schmidt and Witte, 1989, Chapter 8, "Subsample Predictions," pp. 131-138. They made predictions for eleven different groups of releasees, such as youthful offenders, nonwhites, and participants in the North Carolina work release program. Their predictions are summarized in their Table 8.3, p. 133, most of which we reproduce as Table 10 below. For most groups the predictions are not satisfactory, and this led Schmidt and Witte to conclude that their models needed to be improved before they could be useful in making predictions for nonrandom groups of releasees. In particular, since their models contained as explanatory variables the individual characteristics that defined their subsamples, the failure of the models to predict accurately for subsamples was taken as evidence that the models did not adequately capture the effects of these variables on time until recidivism. The work in REPORT1 and REPORT2 was intended to improve the earlier models of Schmidt and Witte, 1989, by allowing more flexibility in the ways that explanatory variables were allowed to affect time until recidivism, and we now see whether these improved models yield more accurate predictions for selected groups than the original models of Schmidt and Witte did.

We consider ten groups of individuals for whom we make predictions. Eight of these are the same as groups that were considered by Schmidt and Witte: OLD AGE, defined as AGE ≥ 40 years (480 months); PRIORS > 0; WHITE = 0; ALCHY = 1; JUNKY = 1; FELON = 1; PROPTY = 1; and MALE = 0. In addition, we consider two groups they did not consider: LONG TSERVD, defined as SHORT = 0, that is, as TSERVD > 30 months for 1978 and TSERVD > 22 months for 1980; and YOUNG AGE, defined as YOUNG = 1, that is, as AGE ≤ 240 months for 1978 and AGE ≤ 216 months for 1980. These represent slightly different definitions of long time served and young age than were used by Schmidt and Witte. In addition, they considered a group consisting of individuals who had participated in the North Carolina Prisoner Work Release Program, but we do not consider this group because membership in it was not a significant explanatory variable in our models.

Tables 5 - 9 give summary statistics for our predictions for the ten groups listed above, based on the logit/individual lognormal, lognormal, split lognormal, logit lognormal, and proportional hazards models, respectively. For each group we give the number (n) of individuals in the group and the $\chi^2$ and Kolmogorov-Smirnov (KS) statistics, as discussed in section 2 above. We give results for predictions for 1978 and 1980, and also for predictions for 1980 using models fit to the 1978 data.

We first note that the logit/individual lognormal model (Table 5) clearly yields the best predictions for the groups we consider. For almost all groups and years its predictions are much better than those of the lognormal, split lognormal and proportional hazards models. Its predictions are also clearly superior to those of the logit lognormal model for 1978; for 1980, they are of about the same average quality as those of the logit lognormal model. Recall that in terms of predictions for the entire validation sample (Table 1), the logit/individual lognormal model did not predict better (or differently) than the logit lognormal model. However, in making predictions for nonrandom samples, the additional flexibility of the logit/individual lognormal model is useful. This is a reasonable result.

A more difficult question to answer is the extent to which our expanded models improve on the predictions of the simpler models of Schmidt and Witte, 1989. Their Table 8.3, p. 133, gives summary statistics for predictions for the eleven groups they considered, and is essentially identical in format to Table 5. Thus we need to compare the results in Table 5 with the results in their Table 8.3. To make this comparison easier, we have reproduced their Table 8.3 as Table 10 (omitting the groups that were defined differently there than in the present report). In making these comparisons, two points should be kept in mind. First, our Table 5 gives results for the logit/individual lognormal model, while their Table 8.3 gives results for their logit lognormal model. They did not generate predictions for subsamples with their logit/individual lognormal model. However, because their specifications differed very little across models, compared to ours, the choice of model would probably not have made much difference in their predictions. Second, some of their groups differ from ours, and so it is legitimate to make comparisons only for the eight groups that are defined identically here and in their analyses. These are the last eight groups listed in Table 5.

Comparing the results in Table 5 to the results in Table 10 is not easy because the model that predicts best depends too much on the year, the group and the choice of summary statistic to allow easy generalization. It is accurate to say that our expanded models (Table 5) usually predict better than the original models of Schmidt and Witte (Table 10). They generate smaller $\chi^2$ statistics for seven groups out of eight for 1978, and also for seven groups out of eight for 1980; and they generate smaller KS statistics for four groups out of eight for 1978, and for five groups out of eight for 1980. When the 1978 models are used to make predictions for 1980, the comparison is clearer. The expanded models generate smaller $\chi^2$ statistics than the original models for all eight groups, and they generate smaller KS statistics for seven groups out of eight. Thus it is fair to say that our expansion of the Schmidt and Witte specifications has resulted in improvement in the predictions for nonrandom samples of releasees. The extent of the improvement probably does not match the 25 - 30% improvement in within-sample explanatory power that was reported in REPORT2, but it is still an improvement.

In fact, for most groups the size of the difference in the quality of predictions using the Schmidt and Witte specification (Table 10) or our expanded specification (Table 5) is not large. In no case does the expanded specification result in much worse predictions. In a few cases, the expanded specification does result in much better predictions. This is so for the group defined by ALCHY = 1 for the 1978 sample, and for the group defined by PRIORS > 0 for both 1978 and 1980.

To give a more detailed idea of the quality of predictions that our models give for nonrandom subsamples, we also provide a more detailed display of predicted and actual recidivism for three of our eight subsamples. These results are for the 1978 validation sample and use the logit/individual lognormal model. They are given in Tables 11, 12 and 13, which have the same format as Tables 2, 3 and 4 above.

Table 11 gives our results for the group defined by WHITE = 0, a group for which our predictions were quite good. We can see in Table 11 that the model overpredicts the rate of recidivism for about the first six months after release, and predicts quite accurately thereafter.

Table 12 gives our results for the group defined by YOUNG = 1 (AGE ≤ 240 months). This is a group for which the quality of predictions is about average among the groups we have considered. There is a fairly serious tendency for the model to underpredict the rate of recidivism during the first year (except for the first two months) after release. However, because the group is relatively small, most of the differences between actual and predicted recidivism rates (density, hazard or cdf) are not statistically significantly different from zero.

Table 13 gives our results for the group defined by JUNKY = 1, which is a group for which we predict quite poorly. Interestingly, the model overpredicts the rate of recidivism of this group over virtually the entire range of times after release. The differences between predicted and actual outcomes are substantial and are often statistically significantly different from zero.

Overall, our results for prediction for nonrandom subsamples of releasees are somewhat disappointing. Schmidt and Witte, 1989, regarded their results as "not very satisfactory," to the point that they cast doubt on the model's usefulness in program evaluation. The present results are better than theirs, but they still really do not change this pessimistic conclusion.

## 5. Individual Predictions

In this section we consider the use of our models to make predictions for individuals, instead of for groups. The event which we will attempt to predict is recidivism before the end of the follow-up period, a discrete (yes/no) outcome. Each of our models yields a probability of this event, for each individual in the sample, and the basic question is how well these (predicted) probabilities of recidivism agree with the observed outcomes.

We begin with the following standard calculation. We take as given the number of individuals in the sample who fail, and we see how well we can predict <u>which</u> individuals these will be. For example, in the 1978 estimation sample, the failure rate is .366. Since there are 3078 individuals in the 1978 validation sample, we predict failure for (.366)(3078) = 1127 individuals. We do this, for a given model, by predicting failure for the 1127 individuals in the validation sample with the highest predicted probabilities of failure, and by predicting no failure for the remaining 3078 - 1127 = 1951 individuals. The calculations for the 1980 validation sample follow the same pattern.

For the 1978 validation sample and their proportional hazards model, Schmidt and Witte, 1989 (p. 142) report a false positive rate of 0.472 (532 of the 1127 predicted failures do not fail) and a false negative rate of 0.277 (540 of the 1951 predicted successes fail). With our expanded specifications, we now find lower false positive and false negative rates, though the differences are not large. For 1978, our false positive rate is now 0.455 for the proportional hazards model, 0.452 for the logit lognormal model, and 0.456 for the logit/individual lognormal model; each of these is less than the Schmidt and Witte false positive rate of 0.472. Similarly, our false negative rates for the same three models are 0.268, 0.266, and 0.269, and each of these is less than the Schmidt and Witte false negative rate of 0.277. The results for 1980 are similar, though both the false positive rates and the false negative rates are slightly higher. (For the models in the same order as above, the false positive rates are 0.475, 0.474 and 0.480, while the false negative rates are 0.282, 0.282 and 0.285.) As noted by Schmidt and Witte, 1989, p. 142, these error rates are less than those commonly found in the literature. However, it is clearly the case that our augmentations of the models of Schmidt and Witte have resulted in only a modest improvement in these error rates.

For practical purposes, such as a policy of selective incapacitation, a false positive rate of over 40% is clearly unacceptable. However, we might be satisfied if we could predict recidivism with considerable assurance even for a very limited proportion of the sample. We therefore follow Schmidt and Witte, 1989, by making predictions for many different proportions of the sample, arranged in order of predicted probability of failure; the question of interest is how small the proportion of the sample for which we predict failure must be in order that the false positive rate becomes acceptably small. These results are given in Tables 14-19, which are of essentially the same form as Tables 8.7 - 8.10 of Schmidt and Witte, 1989. Tables 14, 15 and 16 correspond to the 1978 validation sample and the proportional hazards, logit lognormal, and logit/individual lognormal models, respectively, while Tables 17 - 19 give the same results for the 1980 validation sample.

Consider, for example, the results in Table 16, which are predictions made for the 1978 validation sample using the logit lognormal model. The upper .5% percentile corresponds to the 15 (.5% of 3078 = 15) individuals with the highest predicted probabilities of failure, according to the logit lognormal model. Of these 15 individuals, the actual failure rate is 93.3%, because 14 of these 15 individuals actually failed. Thus the model is successful in identifying a (very small) group of individuals who are very likely to fail. For these 15 individuals, the average probability of failure as indicated by the model is 93.8%, so that the model is also successful in predicting how many individuals in the group will fail.

For the 1978 validation sample, the logit/individual lognormal model gives better predictions than the logit lognormal or proportional hazard models. It is better both at identifying groups of individuals who are very likely to fail and also at predicting accurately what percentage of the group will fail. We can indeed use the model to identify a group for which we can predict recidivism with considerable assurance, if we restrict our predictions to a small enough fraction of the sample. For example, as noted above, the actual failure rate is 93.3% in the worst (highest predicted probability of failure) .5% of the sample, and it is 87.1% in the worst 1% of the sample, 79.2% in the worst 5% of the sample, and 72.2% in the worst 10% of the sample. These may be reasonable fractions of the sample to consider, in the sense that a policy of selective incapacitation would presumably be applied only to a small fraction of potential releasees. Furthermore, the logit/individual lognormal model predicts the failure rate in these groups quite accurately.

The results in Table 16 are considerably better than those reported by Schmidt and Witte, 1989. Their best predictions for the 1978 validation sample came from the logit lognormal model, and they had failure rates of 80.0% in the worst .5% of the sample, 83.9% in the worst 1% of the sample, and 70.1% in the worst 5% of the sample. When we use our models to identify these small groups of very likely failures, our expansion of the Schmidt and Witte models has reduced the false positive rate very considerably: by over one half for the worst .5% of the sample, and by about one third for the worst 5%.

Our results for 1980 are a little less optimistic, because we have higher false positive rates and because we are able to make less of an improvement on the results of the earlier models. Schmidt and Witte, 1989, Table 8.8, report failure rates of 81.8% in the worst .5% of the sample, 81.4% in the worst 1% of the sample, and 69.8% in the worst 5% of the sample. The corresponding failure rates for the groups identified by our expanded logit lognormal model (Table 18) are 81.8%, 86.0% and 70.7%, which are not strikingly different from the earlier results. The results for the proportional hazards model (Table 17) are similar. For the 1980 sample, unlike the 1978 sample, the logit/individual lognormal model does not predict as well as the proportional hazards and logit lognormal models.

Because most individuals in the sample do not fail, we are better able to identify groups that will <u>not</u> fail than we are to identify groups that will fail. Looking at the results for the lower percentiles of the predicted probability of failure, we see that for the 1978 sample and logit/individual lognormal model (Table 16), false negative rates are fairly low. For example, the best (lowest predicted probability of failure) 10% of the sample has a failure rate of only 9.1%, and this false negative rate is much lower than the false positive rate in the worst 10% of the sample. It is also lower than the false negative rate of 13.0% reported in Schmidt and Witte, 1989, Table 8.10. The same sorts of comparisons hold for 1980, though the false negative rates are higher, and the 1980 models do not predict the false negative rate as accurately as the 1978 models do.

## 6. Concluding Remarks

The point of this research project was to investigate the extent to which the previous models of Schmidt and Witte, 1989, could be improved by more careful consideration of the ways in which explanatory variables were entered into the models. These models included proportional hazards models and also parametric models based on the lognormal distribution.

We found that the data did support more elaborate specifications than those previously used. Our expanded models give a more complete picture of the way in which explanatory variables affect recidivism. In particular, age and number of prior incarcerations were found to have strong nonlinear effects that the previous models did not reveal. Our expanded models also have a higher degree of explanatory power than our previous models; we were able to increase one measure of $R^2$ (variance explained) from about 0.12 to about 0.15, an increase of about 25 to 30%.

Perhaps unsurprisingly, this increase in explanatory power was not matched by a commensurate increase in the quality of out-of-sample predictions. Our predictions for the entire 1978 and 1980 validation samples were improved only very slightly by using the expanded models. Similarly, when we use the estimates based on the 1978 estimation sample to make predictions for the 1980 validation sample, the predictions from our expanded models were not very different than those from the original (Schmidt and Witte) specifications of the models. These results are not surprising or discouraging, since the models of Schmidt and Witte already predicted quite well for random samples of releasees.

Predictions for nonrandom subsamples of releasees are more challenging than predictions for random samples, and the ability to predict accurately for nonrandom samples is important in practical uses of the models, such as program evaluation. We considered eight different groups of releasees, selected on the basis of individual characteristics and correctional histories. Our expanded models did lead to improvement in the ability to predict recidivism for most groups. For a few groups, such as the group of individuals with more than one prior incarceration, this improvement was substantial. However, for most groups any improvement in predictive ability was rather small, and the accuracy with which we can predict for nonrandom samples is still not encouraging.

Our expanded specifications result in more considerable improvements in predictive ability when we make predictions for individuals instead of for groups. In particular, when we use our models to identify small groups of individuals with very high probabilities of failure, these groups have very low false positive rates if the groups are small enough. For example, when we use the logit/individual lognormal model to identify the "worst" one percent of the sample of releasees (that is, the one percent of the sample that is judged most likely to fail), the actual failure rate is almost 90%. This is a very considerable increase in predictive accuracy over that provided by the models of Schmidt and Witte, 1989; the false positive rate has been cut about in half. These individuals have unusual values of the explanatory variables (individual characterstics and correctional histories), and so it is not surprising that it takes a more elaborate model to predict accurately for them than it does to predict accurately for individuals or groups of individuals with more common characteristics.

No research project can ever provide the last word on any topic, and we do not feel that we have found all possible improvements in our models. However, for our data set (and presumably for others that are similar), we doubt that substantial improvements in explanatory power or in predictive ability would be achieved by more careful use of the same types of individual characteristics and correctional histories as we have considered. In fact, although Schmidt and Witte, 1989 used only linear specifications, they sacrificed relatively little explanatory power or predictive accuracy by doing so. We were able to find relevant nonlinearities easily enough, but they simply didn't make as much difference in prediction as might have been suspected beforehand. This is not really an argument for the careless use of linear specifications, but it does argue that substantial improvements in prediction of recidivism are likely to depend on new and better data, or at the least on improved measures of important individual characteristics (e.g., accurate identification of drug abusers). Improvements in criminological theory that would indicate what measures ought to be collected and used would be especially useful.

TABLE 1

PREDICTIONS FOR VALIDATION SAMPLES: SUMMARY STATISTICS

| Model | Explanatory Variables | 1978 | | 1980 | | Use 1978 to predict 1980 | |
|---|---|---|---|---|---|---|---|
| | | $X^2$ | KS | $X^2$ | KS | $X^2$ | KS |
| Nonparametric | None | 130.3 | .012 | 222.2 | .028 | 273.1 | .054 |
| Lognormal | None | 142.2 | .037 | 172.4 | .045 | 350.3 | .085 |
| Split Lognormal | None | 50.6 | .005 | 53.9 | .023 | 126.3 | .053 |
| Proportional Hazards | SW1989 | 127.1 | .012 | 232.4 | .034 | 197.6 | .037 |
| Lognormal | SW1989 | 142.5 | .034 | 181.9 | .051 | 245.0 | .063 |
| Split Lognormal | SW1989 | 110.7 | .030 | 173.9 | .050 | 197.0 | .057 |
| Logit Lognormal | SW1989 | 50.8 | .006 | 57.1 | .024 | 97.6 | .042 |
| Logit/individual Log. | SW1989 | 51.6 | .013 | 60.0 | .027 | 66.3 | .029 |
| Proportional Hazards | Expanded | 111.1 | .011 | 226.8 | .032 | 205.0 | .037 |
| Lognormal | Expanded | 143.2 | .035 | 177.6 | .050 | 232.3 | .060 |
| Split Lognormal | Expanded | 124.2 | .034 | Not available | | 199.4 | .056 |
| Logit Lognormal | Expanded | 52.5 | .005 | 53.2 | .022 | 97.3 | .042 |
| Logit/individual Log. | Expanded | 50.0 | .006 | 55.2 | .026 | 63.3 | .031 |

TABLE 2

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL
1978 VALIDATION SAMPLE

| N | Predicted cdf | cdf | $t$ | Predicted density | density | $t$ | Predicted hazard | hazard | $t$ | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .000 | -1.00 | .000 | .000 | -1.00 | .000 | .000 | -1.00 | 0.0 |
| 1 | .005 | .005 | 0.44 | .004 | .005 | 0.73 | .004 | .005 | 0.71 | 0.2 |
| 2 | .013 | .010 | -1.63 | .008 | .005 | -2.33 | .008 | .005 | -2.38 | 5.8 |
| 3 | .024 | .019 | -1.67 | .011 | .009 | -0.64 | .011 | .010 | -0.79 | 6.5 |
| 4 | .035 | .031 | -1.50 | .012 | .011 | -0.22 | .012 | .012 | -0.43 | 6.7 |
| 5 | .048 | .042 | -1.57 | .012 | .011 | -0.48 | .013 | .012 | -0.72 | 7.2 |
| 6 | .060 | .055 | -1.28 | .012 | .013 | 0.27 | .013 | .014 | 0.03 | 7.2 |
| 7 | .073 | .069 | -0.81 | .012 | .014 | 0.82 | .014 | .015 | 0.57 | 7.5 |
| 8 | .085 | .082 | -0.50 | .012 | .013 | 0.62 | .013 | .014 | 0.45 | 7.7 |
| 9 | .096 | .094 | -0.49 | .012 | .012 | -0.03 | .013 | .013 | -0.17 | 7.7 |
| 10 | .108 | .107 | -0.22 | .011 | .013 | 0.68 | .013 | .014 | 0.55 | 8.0 |
| 11 | .119 | .119 | -0.02 | .011 | .012 | 0.57 | .013 | .013 | 0.45 | 8.2 |
| 12 | .129 | .131 | 0.30 | .011 | .012 | 0.99 | .012 | .014 | 0.92 | 9.1 |
| 13 | .139 | .144 | 0.72 | .010 | .013 | 1.43 | .012 | .015 | 1.38 | 11.0 |
| 14 | .149 | .151 | 0.39 | .010 | .008 | -1.07 | .011 | .009 | -1.08 | 12.2 |
| 15 | .158 | .160 | 0.20 | .009 | .008 | -0.67 | .011 | .010 | -0.68 | 12.6 |
| 16 | .167 | .171 | 0.64 | .009 | .012 | 1.66 | .011 | .014 | 1.67 | 15.4 |
| 17 | .176 | .178 | 0.41 | .009 | .007 | -0.83 | .010 | .009 | -0.79 | 16.0 |
| 18 | .184 | .186 | 0.35 | .008 | .008 | -0.22 | .010 | .009 | -0.18 | 16.1 |
| 19 | .192 | .193 | 0.25 | .008 | .007 | -0.42 | .010 | .009 | -0.39 | 16.2 |
| 20 | .199 | .200 | 0.15 | .007 | .007 | -0.43 | .009 | .008 | -0.39 | 16.4 |
| 21 | .206 | .206 | -0.05 | .007 | .006 | -0.87 | .009 | .007 | -0.83 | 17.1 |
| 22 | .213 | .212 | -0.15 | .007 | .006 | -0.47 | .009 | .008 | -0.44 | 17.3 |
| 23 | .220 | .218 | -0.30 | .007 | .006 | -0.74 | .008 | .007 | -0.71 | 17.8 |
| 24 | .226 | .224 | -0.23 | .006 | .007 | 0.34 | .008 | .009 | 0.36 | 17.9 |
| 25 | .232 | .232 | -0.08 | .006 | .007 | 0.76 | .008 | .009 | 0.78 | 18.5 |
| 26 | .238 | .239 | 0.10 | .006 | .007 | 0.95 | .008 | .009 | 0.98 | 19.5 |
| 27 | .244 | .245 | 0.18 | .006 | .006 | 0.42 | .007 | .008 | 0.46 | 19.7 |
| 28 | .249 | .251 | 0.24 | .005 | .006 | 0.34 | .007 | .008 | 0.39 | 19.8 |
| 29 | .254 | .258 | 0.46 | .005 | .007 | 1.26 | .007 | .009 | 1.30 | 21.5 |
| 30 | .259 | .262 | 0.36 | .005 | .004 | -0.61 | .007 | .006 | -0.57 | 21.9 |
| 31 | .264 | .264 | 0.01 | .005 | .002 | -2.03 | .006 | .003 | -2.02 | 25.9 |
| 32 | .269 | .267 | -0.26 | .005 | .003 | -1.66 | .006 | .004 | -1.66 | 28.7 |
| 33 | .273 | .271 | -0.29 | .004 | .004 | -0.20 | .006 | .006 | -0.20 | 28.7 |
| 34 | .277 | .276 | -0.25 | .004 | .005 | 0.21 | .006 | .006 | 0.21 | 28.8 |
| 35 | .282 | .280 | -0.20 | .004 | .005 | 0.35 | .006 | .006 | 0.36 | 28.9 |
| 36 | .286 | .284 | -0.21 | .004 | .004 | -0.09 | .006 | .005 | -0.08 | 28.9 |
| 37 | .289 | .288 | -0.12 | .004 | .005 | 0.61 | .005 | .006 | 0.62 | 29.3 |
| 38 | .293 | .291 | -0.23 | .004 | .003 | -0.73 | .005 | .004 | -0.73 | 29.8 |
| 39 | .297 | .295 | -0.27 | .004 | .003 | -0.33 | .005 | .005 | -0.33 | 29.9 |
| 40 | .300 | .298 | -0.34 | .003 | .003 | -0.53 | .005 | .004 | -0.53 | 30.2 |
| 41 | .304 | .301 | -0.32 | .003 | .004 | 0.20 | .005 | .005 | 0.19 | 30.2 |
| 42 | .307 | .303 | -0.48 | .003 | .002 | -1.27 | .005 | .003 | -1.28 | 31.9 |
| 43 | .310 | .307 | -0.34 | .003 | .004 | 1.06 | .005 | .006 | 1.05 | 33.0 |
| 44 | .313 | .310 | -0.40 | .003 | .003 | -0.45 | .004 | .004 | -0.47 | 33.2 |
| 45 | .316 | .314 | -0.32 | .003 | .004 | 0.64 | .004 | .005 | 0.62 | 33.6 |

TABLE 2, CONTINUED

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 46 | .319 | .317 | -0.23 | .003 | .004 | 0.74 | .004 | .005 | 0.73 | 34.1 |
| 47 | .322 | .319 | -0.33 | .003 | .002 | -0.87 | .004 | .003 | -0.88 | 34.9 |
| 48 | .324 | .323 | -0.18 | .003 | .004 | 1.30 | .004 | .006 | 1.29 | 36.5 |
| 49 | .327 | .325 | -0.22 | .003 | .002 | -0.36 | .004 | .003 | -0.36 | 36.7 |
| 50 | .329 | .328 | -0.13 | .003 | .003 | 0.80 | .004 | .005 | 0.79 | 37.3 |
| 51 | .332 | .329 | -0.31 | .002 | .001 | -1.66 | .004 | .001 | -1.66 | 40.0 |
| 52 | .334 | .332 | -0.29 | .002 | .003 | 0.25 | .004 | .004 | 0.24 | 40.1 |
| 53 | .337 | .334 | -0.37 | .002 | .002 | -0.79 | .003 | .002 | -0.80 | 40.8 |
| 54 | .339 | .336 | -0.41 | .002 | .002 | -0.34 | .003 | .003 | -0.36 | 40.9 |
| 55 | .341 | .337 | -0.52 | .002 | .001 | -1.04 | .003 | .002 | -1.07 | 42.0 |
| 56 | .343 | .340 | -0.42 | .002 | .003 | 0.98 | .003 | .004 | 0.94 | 42.9 |
| 57 | .345 | .342 | -0.35 | .002 | .003 | 0.67 | .003 | .004 | 0.64 | 43.3 |
| 58 | .347 | .344 | -0.35 | .002 | .002 | -0.06 | .003 | .003 | -0.08 | 43.3 |
| 59 | .349 | .346 | -0.39 | .002 | .002 | -0.40 | .003 | .002 | -0.43 | 43.5 |
| 60 | .351 | .348 | -0.34 | .002 | .002 | 0.49 | .003 | .003 | 0.46 | 43.7 |
| 61 | .353 | .352 | -0.17 | .002 | .003 | 1.83 | .003 | .005 | 1.79 | 46.9 |
| 62 | .355 | .354 | -0.11 | .002 | .002 | 0.64 | .003 | .004 | 0.61 | 47.3 |
| 63 | .356 | .356 | -0.04 | .002 | .002 | 0.71 | .003 | .004 | 0.68 | 47.8 |
| 64 | .358 | .358 | -0.01 | .002 | .002 | 0.34 | .003 | .003 | 0.32 | 47.9 |
| 65 | .360 | .361 | 0.11 | .002 | .003 | 1.30 | .003 | .004 | 1.27 | 49.5 |
| 66 | .361 | .363 | 0.19 | .002 | .002 | 0.92 | .003 | .004 | 0.91 | 49.5 |
| 67 | .363 | .365 | 0.20 | .002 | .002 | 0.08 | .002 | .003 | 0.07 | 50.0 |
| 68 | .364 | .366 | 0.17 | .002 | .001 | -0.32 | .002 | .002 | -0.34 | 50.0 |
| 69 | .366 | .367 | 0.19 | .001 | .002 | 0.20 | .002 | .003 | 0.18 | 50.0 |
| 70 | .367 | .369 | 0.17 | .001 | .001 | -0.22 | .002 | .002 | -0.24 | 50.0 |

TABLE 3

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL
1980 VALIDATION SAMPLE

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .001 | 3.82 | .000 | .001 | 3.82 | .000 | .001 | 3.82 | 0.0 |
| 1 | .005 | .005 | -0.63 | .005 | .003 | -1.64 | .005 | .003 | -1.65 | 0.4 |
| 2 | .015 | .014 | -0.55 | .010 | .010 | -0.21 | .010 | .010 | -0.27 | 0.5 |
| 3 | .028 | .030 | 0.83 | .013 | .016 | 1.81 | .013 | .016 | 1.68 | 3.3 |
| 4 | .042 | .046 | 1.58 | .014 | .016 | 1.50 | .014 | .017 | 1.40 | 5.3 |
| 5 | .056 | .064 | 2.35 | .014 | .018 | 1.84 | .015 | .019 | 1.79 | 8.5 |
| 6 | .070 | .079 | 2.37 | .014 | .015 | 0.52 | .015 | .016 | 0.49 | 8.7 |
| 7 | .084 | .094 | 2.24 | .014 | .014 | 0.14 | .015 | .016 | 0.11 | 8.7 |
| 8 | .098 | .111 | 2.93 | .014 | .017 | 2.07 | .015 | .019 | 2.04 | 12.9 |
| 9 | .111 | .126 | 3.17 | .013 | .015 | 1.04 | .015 | .017 | 1.09 | 14.1 |
| 10 | .124 | .141 | 3.56 | .013 | .015 | 1.50 | .015 | .018 | 1.57 | 16.5 |
| 11 | .136 | .157 | 4.05 | .012 | .015 | 1.86 | .014 | .018 | 1.95 | 20.3 |
| 12 | .148 | .171 | 4.44 | .012 | .014 | 1.66 | .014 | .017 | 1.79 | 23.5 |
| 13 | .159 | .183 | 4.36 | .011 | .011 | 0.14 | .013 | .014 | 0.31 | 23.6 |
| 14 | .170 | .194 | 4.35 | .011 | .011 | 0.32 | .013 | .014 | 0.48 | 23.8 |
| 15 | .180 | .205 | 4.48 | .010 | .011 | 0.80 | .012 | .014 | 0.99 | 24.8 |
| 16 | .190 | .214 | 4.28 | .010 | .009 | -0.42 | .012 | .011 | -0.25 | 24.9 |
| 17 | .199 | .224 | 4.30 | .009 | .010 | 0.35 | .011 | .012 | 0.51 | 25.1 |
| 18 | .208 | .234 | 4.40 | .009 | .010 | 0.66 | .011 | .013 | 0.81 | 25.8 |
| 19 | .216 | .240 | 3.98 | .008 | .006 | -1.54 | .011 | .008 | -1.41 | 27.8 |
| 20 | .224 | .249 | 4.14 | .008 | .009 | 0.92 | .010 | .012 | 1.05 | 28.9 |
| 21 | .232 | .257 | 4.17 | .008 | .008 | 0.34 | .010 | .011 | 0.49 | 29.1 |
| 22 | .239 | .262 | 3.66 | .007 | .004 | -2.25 | .010 | .006 | -2.13 | 33.7 |
| 23 | .246 | .269 | 3.62 | .007 | .007 | -0.04 | .009 | .009 | 0.07 | 33.7 |
| 24 | .253 | .274 | 3.37 | .007 | .005 | -1.11 | .009 | .007 | -1.02 | 34.7 |
| 25 | .260 | .282 | 3.50 | .006 | .007 | 0.82 | .009 | .010 | 0.90 | 35.5 |
| 26 | .266 | .287 | 3.42 | .006 | .006 | -0.30 | .008 | .008 | -0.23 | 35.6 |
| 27 | .272 | .294 | 3.50 | .006 | .007 | 0.51 | .008 | .009 | 0.58 | 35.9 |
| 28 | .277 | .298 | 3.25 | .006 | .004 | -1.30 | .008 | .006 | -1.25 | 37.5 |
| 29 | .283 | .303 | 3.11 | .005 | .005 | -0.71 | .008 | .007 | -0.66 | 37.9 |
| 30 | .288 | .307 | 2.89 | .005 | .004 | -1.16 | .007 | .006 | -1.13 | 39.2 |
| 31 | .293 | .311 | 2.79 | .005 | .004 | -0.56 | .007 | .006 | -0.54 | 39.5 |
| 32 | .298 | .318 | 3.07 | .005 | .007 | 1.81 | .007 | .010 | 1.83 | 42.8 |
| 33 | .302 | .322 | 2.99 | .005 | .004 | -0.44 | .007 | .006 | -0.42 | 43.0 |
| 34 | .307 | .327 | 3.08 | .004 | .005 | 0.64 | .007 | .008 | 0.65 | 43.4 |
| 35 | .311 | .334 | 3.41 | .004 | .007 | 2.22 | .006 | .010 | 2.24 | 48.5 |
| 36 | .315 | .338 | 3.44 | .004 | .004 | 0.28 | .006 | .007 | 0.31 | 48.6 |
| 37 | .319 | .340 | 3.18 | .004 | .002 | -1.73 | .006 | .004 | -1.72 | 51.5 |
| 38 | .323 | .344 | 3.08 | .004 | .003 | -0.63 | .006 | .005 | -0.62 | 51.9 |
| 39 | .327 | .348 | 3.11 | .004 | .004 | 0.26 | .006 | .006 | 0.26 | 52.0 |
| 40 | .330 | .351 | 3.09 | .004 | .003 | -0.10 | .005 | .005 | -0.10 | 52.0 |
| 41 | .334 | .354 | 3.05 | .003 | .003 | -0.22 | .005 | .005 | -0.23 | 52.0 |
| 42 | .337 | .357 | 2.93 | .003 | .003 | -0.89 | .005 | .004 | -0.90 | 52.8 |
| 43 | .340 | .360 | 2.96 | .003 | .003 | 0.30 | .005 | .005 | 0.28 | 52.9 |
| 44 | .344 | .362 | 2.81 | .003 | .002 | -1.21 | .005 | .003 | -1.23 | 54.4 |
| 45 | .347 | .365 | 2.70 | .003 | .002 | -0.82 | .005 | .004 | -0.86 | 55.2 |
| 46 | .350 | .368 | 2.68 | .003 | .003 | -0.15 | .005 | .004 | -0.19 | 55.2 |

TABLE 4

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL FIT TO 1978 ESTIMATION SAMPLE
USED TO MAKE PREDICTIONS FOR 1980 VALIDATION SAMPLE

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $X^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .001 | 3.48 | .000 | .001 | 3.48 | .000 | .001 | 3.48 | 0.0 |
| 1 | .006 | .005 | -0.77 | .005 | .003 | -1.74 | .005 | .003 | -1.75 | 0.6 |
| 2 | .015 | .014 | -0.62 | .010 | .010 | -0.19 | .010 | .010 | -0.28 | 0.7 |
| 3 | .028 | .030 | 0.85 | .012 | .016 | 1.93 | .013 | .016 | 1.75 | 3.7 |
| 4 | .041 | .046 | 1.71 | .014 | .016 | 1.68 | .014 | .017 | 1.51 | 6.0 |
| 5 | .055 | .064 | 2.58 | .014 | .018 | 2.05 | .015 | .019 | 1.92 | 9.7 |
| 6 | .069 | .079 | 2.68 | .014 | .015 | 0.75 | .015 | .016 | 0.66 | 10.1 |
| 7 | .083 | .094 | 2.64 | .014 | .014 | 0.37 | .015 | .016 | 0.30 | 10.2 |
| 8 | .096 | .111 | 3.42 | .013 | .017 | 2.33 | .015 | .019 | 2.28 | 15.4 |
| 9 | .109 | .126 | 3.73 | .013 | .015 | 1.30 | .014 | .017 | 1.31 | 17.1 |
| 10 | .122 | .141 | 4.19 | .012 | .015 | 1.77 | .014 | .018 | 1.80 | 20.4 |
| 11 | .133 | .157 | 4.74 | .012 | .015 | 2.13 | .014 | .018 | 2.20 | 25.2 |
| 12 | .145 | .171 | 5.20 | .011 | .014 | 1.92 | .013 | .017 | 2.04 | 29.2 |
| 13 | .155 | .183 | 5.17 | .011 | .011 | 0.37 | .013 | .014 | 0.54 | 29.7 |
| 14 | .166 | .194 | 5.20 | .010 | .011 | 0.56 | .012 | .014 | 0.73 | 30.2 |
| 15 | .176 | .205 | 5.38 | .010 | .011 | 1.04 | .012 | .014 | 1.25 | 31.7 |
| 16 | .185 | .214 | 5.22 | .009 | .009 | -0.21 | .011 | .011 | -0.01 | 31.7 |
| 17 | .194 | .224 | 5.28 | .009 | .010 | 0.58 | .011 | .012 | 0.78 | 32.4 |
| 18 | .202 | .234 | 5.42 | .009 | .010 | 0.88 | .011 | .013 | 1.09 | 33.5 |
| 19 | .211 | .240 | 5.03 | .008 | .006 | -1.36 | .010 | .008 | -1.18 | 34.9 |
| 20 | .218 | .249 | 5.22 | .008 | .009 | 1.14 | .010 | .012 | 1.34 | 36.7 |
| 21 | .226 | .257 | 5.29 | .007 | .008 | 0.54 | .010 | .011 | 0.75 | 37.3 |
| 22 | .233 | .262 | 4.79 | .007 | .004 | -2.10 | .009 | .006 | -1.93 | 41.0 |
| 23 | .240 | .269 | 4.78 | .007 | .007 | 0.15 | .009 | .009 | 0.31 | 41.1 |
| 24 | .246 | .274 | 4.55 | .006 | .005 | -0.94 | .009 | .007 | -0.80 | 41.8 |
| 25 | .252 | .282 | 4.71 | .006 | .007 | 1.01 | .008 | .010 | 1.16 | 43.1 |
| 26 | .258 | .287 | 4.66 | .006 | .006 | -0.13 | .008 | .008 | 0.02 | 43.1 |
| 27 | .264 | .294 | 4.75 | .006 | .007 | 0.69 | .008 | .009 | 0.84 | 43.8 |
| 28 | .270 | .298 | 4.52 | .005 | .004 | -1.15 | .008 | .006 | -1.03 | 44.9 |
| 29 | .275 | .303 | 4.40 | .005 | .005 | -0.56 | .007 | .007 | -0.43 | 45.1 |
| 30 | .280 | .307 | 4.20 | .005 | .004 | -1.02 | .007 | .006 | -0.91 | 45.9 |
| 31 | .285 | .311 | 4.12 | .005 | .004 | -0.42 | .007 | .006 | -0.31 | 46.0 |
| 32 | .289 | .318 | 4.42 | .005 | .007 | 1.99 | .007 | .010 | 2.13 | 50.5 |
| 33 | .294 | .322 | 4.35 | .004 | .004 | -0.30 | .006 | .006 | -0.18 | 50.5 |
| 34 | .298 | .327 | 4.46 | .004 | .005 | 0.79 | .006 | .008 | 0.92 | 51.4 |
| 35 | .302 | .334 | 4.80 | .004 | .007 | 2.39 | .006 | .010 | 2.56 | 57.0 |
| 36 | .306 | .338 | 4.85 | .004 | .004 | 0.42 | .006 | .007 | 0.58 | 58.3 |
| 37 | .310 | .340 | 4.60 | .004 | .002 | -1.63 | .006 | .004 | -1.52 | 60.6 |
| 38 | .314 | .344 | 4.52 | .004 | .003 | -0.51 | .005 | .005 | -0.39 | 60.7 |
| 39 | .317 | .348 | 4.56 | .004 | .004 | 0.39 | .005 | .006 | 0.52 | 61.0 |
| 40 | .321 | .351 | 4.55 | .003 | .003 | 0.02 | .005 | .005 | 0.14 | 61.0 |
| 41 | .324 | .354 | 4.53 | .003 | .003 | -0.11 | .005 | .005 | 0.00 | 61.0 |
| 42 | .328 | .357 | 4.41 | .003 | .003 | -0.78 | .005 | .004 | -0.68 | 61.5 |
| 43 | .331 | .360 | 4.46 | .003 | .003 | 0.42 | .005 | .005 | 0.53 | 61.8 |
| 44 | .334 | .362 | 4.31 | .003 | .002 | -1.11 | .005 | .003 | -1.03 | 62.8 |
| 45 | .337 | .365 | 4.21 | .003 | .002 | -0.73 | .004 | .004 | -0.65 | 63.3 |
| 46 | .339 | .368 | 4.20 | .003 | .003 | -0.05 | .004 | .004 | 0.04 | 63.3 |

TABLE 5

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
LOGIT/INDIVIDUAL LOGNORMAL MODEL

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| LONG TSERVD | 584 | 66.6 | .052 | 1248 | 61.3 | .037 | 852 | 66.1 | .028 |
| YOUNG AGE | 338 | 25.8 | .038 | 218 | 21.2 | .073 | 723 | 43.5 | .066 |
| OLD AGE | 387 | 16.9 | .017 | 566 | 51.4 | .071 | 566 | 37.3 | .057 |
| PRIORS > 0 | 1310 | 53.3 | .023 | 1823 | 68.5 | .036 | 1823 | 71.7 | .042 |
| WHITE = 0 | 1470 | 39.1 | .016 | 2106 | 69.4 | .029 | 2106 | 63.9 | .022 |
| ALCHY = 1 | 603 | 23.2 | .013 | 1551 | 51.0 | .035 | 1551 | 47.4 | .022 |
| JUNKY = 1 | 838 | 53.2 | .063 | 845 | 45.2 | .040 | 845 | 43.1 | .039 |
| FELON = 1 | 989 | 75.1 | .048 | 1778 | 40.1 | .016 | 1778 | 138.7 | .090 |
| PROPTY = 1 | 792 | 60.8 | .037 | 1895 | 67.1 | .030 | 1895 | 81.6 | .054 |
| MALE = 0 | 175 | 4.5 | .032 | 242 | 10.0 | .077 | 242 | 29.4 | .082 |

TABLE 6

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
LOGNORMAL MODEL

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| LONG TSERVD | 584 | 162.0 | .102 | 1248 | 119.5 | .058 | 852 | 126.2 | .069 |
| YOUNG AGE | 338 | 54.9 | .047 | 218 | 40.2 | .130 | 723 | 78.4 | .064 |
| OLD AGE | 387 | 34.6 | .029 | 566 | 104.3 | .093 | 566 | 85.1 | .082 |
| PRIORS > 0 | 1310 | 120.7 | .058 | 1823 | 129.6 | .062 | 1823 | 153.5 | .067 |
| WHITE = 0 | 1470 | 91.7 | .029 | 2106 | 110.2 | .045 | 2106 | 136.4 | .049 |
| ALCHY = 1 | 603 | 44.5 | .040 | 1551 | 108.4 | .057 | 1551 | 113.4 | .048 |
| JUNKY = 1 | 838 | 72.5 | .061 | 845 | 77.8 | .036 | 845 | 89.8 | .051 |
| FELON = 1 | 989 | 156.3 | .079 | 1778 | 91.5 | .039 | 1778 | 259.2 | .112 |
| PROPTY = 1 | 792 | 116.1 | .050 | 1895 | 132.9 | .051 | 1895 | 192.8 | .081 |
| MALE = 0 | 175 | 19.4 | .037 | 242 | 17.5 | .081 | 242 | 33.5 | .096 |

TABLE 7

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
SPLIT LOGNORMAL MODEL

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| LONG TSERVD | 584 | 160.6 | .113 | not available* | | | 852 | 123.1 | .076 |
| YOUNG AGE | 338 | 42.3 | .053 | not available | | | 723 | 72.6 | .067 |
| OLD AGE | 387 | 22.7 | .033 | not available | | | 566 | 56.4 | .072 |
| PRIORS > 0 | 1310 | 108.3 | .054 | not available | | | 1823 | 133.3 | .059 |
| WHITE = 0 | 1470 | 80.5 | .028 | not available | | | 2106 | 127.6 | .049 |
| ALCHY = 1 | 603 | 37.9 | .032 | not available | | | 1551 | 97.8 | .038 |
| JUNKY = 1 | 838 | 65.0 | .053 | not available | | | 845 | 82.3 | .050 |
| FELON = 1 | 989 | 151.6 | .082 | not available | | | 1778 | 241.8 | .110 |
| PROPTY = 1 | 792 | 111.4 | .052 | not available | | | 1895 | 172.4 | .077 |
| MALE = 0 | 175 | 17.5 | .038 | not available | | | 242 | 32.6 | .090 |

*split lognormal model was not fit successfully to 1980 estimation sample

TABLE 8

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
LOGIT LOGNORMAL MODEL

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| LONG TSERVD | 584 | 98.4 | .096 | 1248 | 65.1 | .047 | 852 | 79.9 | .063 |
| YOUNG AGE | 338 | 33.2 | .027 | 218 | 16.8 | .065 | 723 | 55.8 | .082 |
| OLD AGE | 387 | 24.3 | .025 | 566 | 56.4 | .073 | 566 | 56.1 | .067 |
| PRIORS > 0 | 1310 | 56.7 | .039 | 1823 | 85.0 | .050 | 1823 | 110.7 | .067 |
| WHITE = 0 | 1470 | 39.6 | .017 | 2106 | 72.6 | .027 | 2106 | 73.8 | .031 |
| ALCHY = 1 | 603 | 26.6 | .035 | 1551 | 61.0 | .043 | 1551 | 68.9 | .051 |
| JUNKY = 1 | 838 | 48.1 | .063 | 845 | 41.7 | .028 | 845 | 55.1 | .059 |
| FELON = 1 | 989 | 75.0 | .049 | 1778 | 37.5 | .016 | 1778 | 125.8 | .082 |
| PROPTY = 1 | 792 | 67.0 | .047 | 1895 | 62.4 | .030 | 1895 | 158.8 | .091 |
| MALE = 0 | 175 | 3.4 | .032 | 242 | 22.0 | .076 | 242 | 30.5 | .087 |

TABLE 9

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
PROPORTIONAL HAZARDS MODEL

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| LONG TSERVD | 584 | 76.8 | .063 | 1248 | 90.9 | .028 | 852 | 82.7 | .040 |
| YOUNG AGE | 338 | 43.7 | .038 | 218 | 19.1 | .054 | 723 | 59.8 | .042 |
| OLD AGE | 387 | 21.3 | .029 | 566 | 73.3 | .077 | 566 | 82.6 | .083 |
| PRIORS > 0 | 1310 | 93.5 | .036 | 1823 | 119.6 | .043 | 1823 | 163.8 | .062 |
| WHITE = 0 | 1470 | 80.3 | .023 | 2106 | 164.6 | .034 | 2106 | 109.5 | .020 |
| ALCHY = 1 | 603 | 25.1 | .038 | 1551 | 91.4 | .039 | 1551 | 111.2 | .049 |
| JUNKY = 1 | 838 | 81.8 | .075 | 845 | 73.1 | .051 | 845 | 69.1 | .035 |
| FELON = 1 | 989 | 64.6 | .039 | 1778 | 120.2 | .021 | 1778 | 162.0 | .078 |
| PROPTY = 1 | 792 | 58.6 | .021 | 1895 | 110.7 | .026 | 1895 | 142.5 | .054 |
| MALE = 0 | 175 | 6.1 | .031 | 242 | 16.8 | .075 | 242 | 25.6 | .081 |

TABLE 10

SUMMARY STATISTICS FOR SUBSAMPLE PREDICTIONS
LOGIT LOGNORMAL MODEL, SCHMIDT AND WITTE SPECIFICATION
(TABLE 8.3 OF SCHMIDT AND WITTE, 1989)

| | 1978 | | | 1980 | | | Use 1978 to predict 1980 | | |
|---|---|---|---|---|---|---|---|---|---|
| | n | $\chi^2$ | KS | n | $\chi^2$ | KS | n | $\chi^2$ | KS |
| OLD AGE | 387 | 33.2 | .016 | 566 | 58.8 | .079 | 566 | 63.1 | .068 |
| PRIORS > 0 | 1310 | 76.1 | .057 | 1823 | 92.4 | .057 | 1823 | 130.6 | .076 |
| WHITE = 0 | 1470 | 43.0 | .017 | 2106 | 67.0 | .021 | 2106 | 74.1 | .032 |
| ALCHY = 1 | 603 | 34.4 | .032 | 1551 | 57.8 | .042 | 1551 | 67.4 | .048 |
| JUNKY = 1 | 838 | 48.4 | .062 | 845 | 44.4 | .030 | 845 | 58.4 | .061 |
| FELON = 1 | 989 | 78.2 | .052 | 1778 | 40.4 | .014 | 1778 | 153.4 | .095 |
| PROPTY = 1 | 792 | 61.7 | .035 | 1895 | 68.0 | .030 | 1895 | 124.3 | .077 |
| MALE = 0 | 175 | 51.1 | .023 | 242 | 26.0 | .047 | 242 | 35.1 | .078 |

TABLE 11

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL
1978 VALIDATION SAMPLE, WHITE = 0

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .000 | -0.71 | .000 | .000 | -0.71 | .000 | .000 | -0.71 | 0.0 |
| 1 | .005 | .007 | 0.89 | .005 | .007 | 1.11 | .005 | .007 | 1.11 | 0.8 |
| 2 | .015 | .012 | -0.76 | .009 | .005 | -1.59 | .009 | .005 | -1.60 | 3.3 |
| 3 | .027 | .021 | -1.35 | .012 | .009 | -1.14 | .012 | .009 | -1.21 | 4.8 |
| 4 | .040 | .031 | -1.77 | .013 | .010 | -1.09 | .014 | .010 | -1.21 | 6.3 |
| 5 | .054 | .039 | -2.70 | .014 | .007 | -2.16 | .015 | .008 | -2.32 | 11.7 |
| 6 | .068 | .052 | -2.54 | .014 | .014 | -0.22 | .016 | .014 | -0.44 | 11.8 |
| 7 | .083 | .070 | -1.82 | .014 | .018 | 1.14 | .016 | .019 | 0.90 | 12.6 |
| 8 | .097 | .087 | -1.28 | .014 | .017 | 1.02 | .016 | .018 | 0.82 | 13.3 |
| 9 | .110 | .099 | -1.38 | .014 | .012 | -0.43 | .015 | .013 | -0.58 | 13.7 |
| 10 | .123 | .114 | -1.09 | .013 | .015 | 0.63 | .015 | .017 | 0.45 | 13.9 |
| 11 | .136 | .130 | -0.69 | .013 | .016 | 1.04 | .015 | .018 | 0.86 | 14.6 |
| 12 | .148 | .144 | -0.50 | .012 | .014 | 0.51 | .014 | .016 | 0.41 | 14.8 |
| 13 | .160 | .159 | -0.04 | .012 | .016 | 1.43 | .014 | .018 | 1.35 | 16.6 |
| 14 | .171 | .165 | -0.59 | .011 | .006 | -1.85 | .013 | .007 | -1.88 | 20.1 |
| 15 | .181 | .173 | -0.84 | .011 | .008 | -0.95 | .013 | .010 | -1.00 | 21.1 |
| 16 | .192 | .188 | -0.41 | .010 | .014 | 1.53 | .013 | .017 | 1.44 | 23.2 |
| 17 | .202 | .197 | -0.43 | .010 | .010 | -0.12 | .012 | .012 | -0.15 | 23.3 |
| 18 | .211 | .205 | -0.55 | .009 | .008 | -0.50 | .012 | .010 | -0.51 | 23.5 |
| 19 | .220 | .214 | -0.63 | .009 | .008 | -0.35 | .011 | .010 | -0.38 | 23.7 |
| 20 | .229 | .221 | -0.74 | .009 | .007 | -0.49 | .011 | .010 | -0.50 | 23.9 |
| 21 | .237 | .231 | -0.61 | .008 | .010 | 0.52 | .011 | .012 | 0.51 | 24.2 |
| 22 | .245 | .241 | -0.39 | .008 | .010 | 0.97 | .010 | .013 | 0.96 | 25.1 |
| 23 | .253 | .247 | -0.53 | .008 | .006 | -0.67 | .010 | .008 | -0.66 | 25.6 |
| 24 | .260 | .255 | -0.45 | .007 | .008 | 0.37 | .010 | .011 | 0.38 | 25.7 |
| 25 | .267 | .265 | -0.22 | .007 | .010 | 1.14 | .009 | .013 | 1.15 | 27.0 |
| 26 | .274 | .272 | -0.15 | .007 | .007 | 0.34 | .009 | .010 | 0.36 | 27.1 |
| 27 | .280 | .278 | -0.25 | .007 | .005 | -0.51 | .009 | .007 | -0.47 | 27.4 |
| 28 | .287 | .282 | -0.44 | .006 | .004 | -1.06 | .009 | .006 | -1.04 | 28.4 |
| 29 | .293 | .288 | -0.37 | .006 | .007 | 0.39 | .008 | .009 | 0.40 | 28.6 |
| 30 | .298 | .293 | -0.52 | .006 | .004 | -0.86 | .008 | .006 | -0.85 | 29.3 |
| 31 | .304 | .296 | -0.71 | .006 | .003 | -1.12 | .008 | .005 | -1.12 | 30.6 |
| 32 | .309 | .299 | -0.88 | .005 | .003 | -1.03 | .008 | .005 | -1.05 | 31.7 |
| 33 | .314 | .306 | -0.73 | .005 | .007 | 0.87 | .007 | .010 | 0.84 | 32.4 |
| 34 | .319 | .312 | -0.63 | .005 | .006 | 0.62 | .007 | .009 | 0.61 | 32.8 |
| 35 | .324 | .317 | -0.63 | .005 | .005 | -0.03 | .007 | .007 | -0.02 | 32.8 |
| 36 | .329 | .322 | -0.62 | .005 | .005 | 0.07 | .007 | .007 | 0.08 | 32.8 |
| 37 | .333 | .325 | -0.71 | .004 | .003 | -0.62 | .007 | .005 | -0.61 | 33.2 |
| 38 | .338 | .329 | -0.73 | .004 | .004 | -0.15 | .006 | .006 | -0.14 | 33.2 |
| 39 | .342 | .335 | -0.62 | .004 | .005 | 0.75 | .006 | .008 | 0.75 | 33.7 |
| 40 | .346 | .338 | -0.67 | .004 | .003 | -0.39 | .006 | .005 | -0.38 | 33.9 |
| 41 | .350 | .343 | -0.60 | .004 | .005 | 0.52 | .006 | .007 | 0.53 | 34.2 |
| 42 | .354 | .346 | -0.69 | .004 | .003 | -0.66 | .006 | .004 | -0.66 | 34.6 |
| 43 | .357 | .349 | -0.71 | .004 | .003 | -0.16 | .006 | .005 | -0.16 | 34.6 |
| 44 | .361 | .351 | -0.84 | .004 | .002 | -0.97 | .005 | .003 | -0.97 | 35.6 |
| 45 | .364 | .355 | -0.78 | .003 | .004 | 0.43 | .005 | .006 | 0.42 | 35.7 |

TABLE 11, CONTINUED

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 46 | .368 | .360 | -0.66 | .003 | .005 | 0.96 | .005 | .007 | 0.96 | 35.7 |
| 47 | .371 | .362 | -0.75 | .003 | .002 | -0.80 | .005 | .003 | -0.80 | 35.7 |
| 48 | .374 | .366 | -0.67 | .003 | .004 | 0.66 | .005 | .006 | 0.66 | 35.8 |
| 49 | .377 | .369 | -0.64 | .003 | .003 | 0.26 | .005 | .005 | 0.27 | 36.2 |
| 50 | .380 | .373 | -0.60 | .003 | .003 | 0.33 | .005 | .005 | 0.34 | 36.2 |
| 51 | .383 | .375 | -0.67 | .003 | .002 | -0.58 | .005 | .003 | -0.58 | 36.3 |
| 52 | .386 | .378 | -0.61 | .003 | .003 | 0.46 | .004 | .005 | 0.47 | 36.3 |
| 53 | .388 | .381 | -0.61 | .003 | .003 | 0.03 | .004 | .004 | 0.03 | 36.4 |
| 54 | .391 | .383 | -0.66 | .003 | .002 | -0.43 | .004 | .003 | -0.43 | 36.4 |
| 55 | .393 | .385 | -0.70 | .003 | .002 | -0.38 | .004 | .003 | -0.38 | 36.7 |
| 56 | .396 | .388 | -0.62 | .002 | .003 | 0.73 | .004 | .006 | 0.72 | 36.7 |
| 57 | .398 | .391 | -0.59 | .002 | .003 | 0.26 | .004 | .004 | 0.26 | 37.2 |
| 58 | .401 | .394 | -0.56 | .002 | .003 | 0.31 | .004 | .004 | 0.32 | 37.2 |
| 59 | .403 | .395 | -0.69 | .002 | .001 | -1.28 | .004 | .001 | -1.28 | 37.7 |
| 60 | .405 | .396 | -0.76 | .002 | .001 | -0.69 | .004 | .002 | -0.69 | 37.7 |
| 61 | .407 | .398 | -0.76 | .002 | .002 | -0.08 | .004 | .003 | -0.09 | 38.0 |
| 62 | .409 | .399 | -0.82 | .002 | .001 | -0.61 | .003 | .002 | -0.62 | 38.0 |
| 63 | .411 | .402 | -0.77 | .002 | .003 | 0.59 | .003 | .005 | 0.58 | 38.0 |
| 64 | .413 | .404 | -0.76 | .002 | .002 | 0.06 | .003 | .003 | 0.05 | 38.0 |
| 65 | .415 | .407 | -0.64 | .002 | .003 | 1.29 | .003 | .006 | 1.28 | 38.9 |
| 66 | .417 | .410 | -0.57 | .002 | .003 | 0.75 | .003 | .005 | 0.75 | 38.9 |
| 67 | .419 | .412 | -0.60 | .002 | .001 | -0.42 | .003 | .002 | -0.42 | 38.9 |
| 68 | .421 | .413 | -0.64 | .002 | .001 | -0.38 | .003 | .002 | -0.39 | 38.9 |
| 69 | .422 | .416 | -0.56 | .002 | .003 | 0.91 | .003 | .005 | 0.90 | 39.1 |
| 70 | .424 | .417 | -0.58 | .002 | .001 | -0.31 | .003 | .002 | -0.31 | 39.1 |

TABLE 12

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL
1978 VALIDATION SAMPLE, YOUNG AGE

| N | Predicted cdf | cdf | $t$ | Predicted density | density | $t$ | Predicted hazard | hazard | $t$ | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .000 | -0.33 | .000 | .000 | -0.33 | .000 | .000 | -0.33 | 0.0 |
| 1 | .005 | .003 | -0.51 | .005 | .003 | -0.44 | .005 | .003 | -0.44 | 0.0 |
| 2 | .014 | .003 | -1.77 | .009 | .000 | -1.78 | .009 | .000 | -1.81 | 0.0 |
| 3 | .026 | .024 | -0.29 | .012 | .021 | 1.45 | .013 | .021 | 1.33 | 0.1 |
| 4 | .040 | .044 | 0.42 | .014 | .021 | 1.07 | .014 | .021 | 1.03 | 0.1 |
| 5 | .055 | .056 | 0.11 | .015 | .012 | -0.47 | .016 | .012 | -0.48 | 0.3 |
| 6 | .070 | .080 | 0.71 | .015 | .024 | 1.24 | .017 | .025 | 1.20 | 1.8 |
| 7 | .086 | .089 | 0.20 | .016 | .009 | -0.99 | .017 | .010 | -0.98 | 2.7 |
| 8 | .101 | .121 | 1.27 | .015 | .033 | 2.56 | .017 | .036 | 2.55 | 9.2 |
| 9 | .116 | .139 | 1.35 | .015 | .018 | 0.39 | .017 | .020 | 0.45 | 9.4 |
| 10 | .131 | .151 | 1.11 | .015 | .012 | -0.45 | .017 | .014 | -0.38 | 9.4 |
| 11 | .146 | .169 | 1.25 | .014 | .018 | 0.52 | .016 | .021 | 0.59 | 9.5 |
| 12 | .160 | .189 | 1.55 | .014 | .021 | 1.06 | .016 | .025 | 1.17 | 9.5 |
| 13 | .173 | .210 | 1.87 | .013 | .021 | 1.15 | .016 | .026 | 1.28 | 12.9 |
| 14 | .186 | .216 | 1.47 | .013 | .006 | -1.15 | .016 | .007 | -1.08 | 12.9 |
| 15 | .199 | .228 | 1.40 | .013 | .012 | -0.12 | .015 | .015 | -0.01 | 13.4 |
| 16 | .211 | .246 | 1.63 | .012 | .018 | 0.95 | .015 | .023 | 1.11 | 13.4 |
| 17 | .222 | .254 | 1.47 | .012 | .009 | -0.48 | .014 | .012 | -0.36 | 13.8 |
| 18 | .234 | .260 | 1.21 | .011 | .006 | -0.93 | .014 | .008 | -0.84 | 13.8 |
| 19 | .244 | .275 | 1.37 | .011 | .015 | 0.71 | .014 | .020 | 0.83 | 13.8 |
| 20 | .255 | .293 | 1.67 | .010 | .018 | 1.34 | .014 | .024 | 1.48 | 13.8 |
| 21 | .265 | .299 | 1.47 | .010 | .006 | -0.75 | .013 | .008 | -0.63 | 14.3 |
| 22 | .274 | .305 | 1.30 | .010 | .006 | -0.70 | .013 | .008 | -0.58 | 14.3 |
| 23 | .284 | .311 | 1.14 | .009 | .006 | -0.64 | .012 | .009 | -0.54 | 14.9 |
| 24 | .293 | .314 | 0.88 | .009 | .003 | -1.16 | .012 | .004 | -1.09 | 14.9 |
| 25 | .301 | .331 | 1.26 | .009 | .018 | 1.83 | .012 | .026 | 1.98 | 15.3 |
| 26 | .309 | .340 | 1.27 | .008 | .009 | 0.12 | .011 | .013 | 0.26 | 15.3 |
| 27 | .317 | .355 | 1.55 | .008 | .015 | 1.41 | .011 | .022 | 1.59 | 17.1 |
| 28 | .325 | .358 | 1.34 | .008 | .003 | -0.99 | .011 | .005 | -0.90 | 17.1 |
| 29 | .332 | .364 | 1.28 | .007 | .006 | -0.32 | .011 | .009 | -0.22 | 17.1 |
| 30 | .340 | .367 | 1.10 | .007 | .003 | -0.91 | .010 | .005 | -0.84 | 18.3 |
| 31 | .346 | .373 | 1.06 | .007 | .006 | -0.22 | .010 | .009 | -0.13 | 18.3 |
| 32 | .353 | .379 | 1.02 | .007 | .006 | -0.17 | .010 | .009 | -0.08 | 18.3 |
| 33 | .360 | .391 | 1.23 | .006 | .012 | 1.25 | .010 | .019 | 1.36 | 18.9 |
| 34 | .366 | .399 | 1.34 | .006 | .009 | 0.63 | .009 | .015 | 0.76 | 18.9 |
| 35 | .372 | .399 | 1.10 | .006 | .000 | -1.43 | .009 | .000 | -1.38 | 18.9 |
| 36 | .377 | .405 | 1.10 | .006 | .006 | 0.03 | .009 | .010 | 0.13 | 18.9 |
| 37 | .383 | .408 | 0.99 | .006 | .003 | -0.65 | .009 | .005 | -0.58 | 18.9 |
| 38 | .388 | .408 | 0.78 | .005 | .000 | -1.36 | .009 | .000 | -1.32 | 18.9 |
| 39 | .394 | .411 | 0.68 | .005 | .003 | -0.58 | .008 | .005 | -0.52 | 20.9 |
| 40 | .399 | .414 | 0.60 | .005 | .003 | -0.55 | .008 | .005 | -0.50 | 20.9 |
| 41 | .404 | .417 | 0.52 | .005 | .003 | -0.51 | .008 | .005 | -0.47 | 20.9 |
| 42 | .408 | .417 | 0.34 | .005 | .000 | -1.27 | .008 | .000 | -1.25 | 20.9 |
| 43 | .413 | .420 | 0.27 | .005 | .003 | -0.45 | .008 | .005 | -0.41 | 22.5 |
| 44 | .418 | .423 | 0.21 | .004 | .003 | -0.42 | .007 | .005 | -0.39 | 22.5 |
| 45 | .422 | .429 | 0.28 | .004 | .006 | 0.45 | .007 | .010 | 0.49 | 22.5 |

TABLE 12, CONTINUED

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 46 | .426 | .432 | 0.23 | .004 | .003 | -0.35 | .007 | .005 | -0.32 | 22.5 |
| 47 | .430 | .441 | 0.41 | .004 | .009 | 1.39 | .007 | .016 | 1.45 | 23.0 |
| 48 | .434 | .444 | 0.37 | .004 | .003 | -0.29 | .007 | .005 | -0.25 | 23.0 |
| 49 | .438 | .444 | 0.23 | .004 | .000 | -1.14 | .007 | .000 | -1.12 | 23.0 |
| 50 | .442 | .447 | 0.20 | .004 | .003 | -0.23 | .006 | .005 | -0.20 | 23.0 |
| 51 | .445 | .450 | 0.17 | .004 | .003 | -0.20 | .006 | .005 | -0.17 | 23.0 |
| 52 | .449 | .450 | 0.04 | .004 | .000 | -1.09 | .006 | .000 | -1.08 | 24.5 |
| 53 | .452 | .450 | -0.09 | .003 | .000 | -1.08 | .006 | .000 | -1.06 | 24.5 |
| 54 | .455 | .450 | -0.22 | .003 | .000 | -1.06 | .006 | .000 | -1.05 | 24.5 |
| 55 | .459 | .453 | -0.23 | .003 | .003 | -0.09 | .006 | .005 | -0.08 | 24.5 |
| 56 | .462 | .459 | -0.12 | .003 | .006 | 0.92 | .006 | .011 | 0.93 | 24.5 |
| 57 | .465 | .462 | -0.13 | .003 | .003 | -0.03 | .006 | .005 | -0.02 | 24.8 |
| 58 | .468 | .462 | -0.24 | .003 | .000 | -1.00 | .005 | .000 | -1.00 | 24.8 |
| 59 | .471 | .464 | -0.24 | .003 | .003 | 0.03 | .005 | .005 | 0.03 | 24.8 |
| 60 | .474 | .464 | -0.35 | .003 | .000 | -0.98 | .005 | .000 | -0.97 | 24.8 |
| 61 | .476 | .467 | -0.34 | .003 | .003 | 0.08 | .005 | .006 | 0.08 | 24.8 |
| 62 | .479 | .467 | -0.44 | .003 | .000 | -0.95 | .005 | .000 | -0.95 | 24.8 |
| 63 | .482 | .473 | -0.31 | .003 | .006 | 1.20 | .005 | .011 | 1.20 | 25.2 |
| 64 | .484 | .476 | -0.29 | .003 | .003 | 0.16 | .005 | .006 | 0.17 | 25.2 |
| 65 | .487 | .488 | 0.06 | .002 | .012 | 3.48 | .005 | .023 | 3.51 | 25.2 |
| 66 | .489 | .488 | -0.03 | .002 | .000 | -0.90 | .005 | .000 | -0.89 | 25.2 |
| 67 | .491 | .488 | -0.12 | .002 | .000 | -0.89 | .004 | .000 | -0.88 | 25.2 |
| 68 | .494 | .494 | 0.02 | .002 | .006 | 1.40 | .004 | .012 | 1.43 | 25.2 |
| 69 | .496 | .494 | -0.06 | .002 | .000 | -0.87 | .004 | .000 | -0.86 | 25.2 |
| 70 | .498 | .494 | -0.15 | .002 | .000 | -0.86 | .004 | .000 | -0.85 | 25.8 |

TABLE 13

PREDICTED VERSUS ACTUAL RECIDIVISM RATES
LOGIT/INDIVIDUAL LOGNORMAL MODEL
1978 VALIDATION SAMPLE, JUNKY = 1

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .000 | .000 | -0.62 | .000 | .000 | -0.62 | .000 | .000 | -0.62 | 0.0 |
| 1 | .006 | .002 | -1.36 | .005 | .002 | -1.23 | .006 | .002 | -1.24 | 0.0 |
| 2 | .016 | .008 | -1.82 | .010 | .006 | -1.21 | .010 | .006 | -1.25 | 3.1 |
| 3 | .029 | .020 | -1.51 | .013 | .012 | -0.19 | .013 | .012 | -0.30 | 3.2 |
| 4 | .043 | .031 | -1.73 | .014 | .011 | -0.80 | .015 | .011 | -0.95 | 4.1 |
| 5 | .057 | .045 | -1.54 | .015 | .014 | -0.05 | .016 | .015 | -0.25 | 4.2 |
| 6 | .072 | .060 | -1.42 | .015 | .014 | -0.07 | .016 | .015 | -0.27 | 4.3 |
| 7 | .086 | .074 | -1.33 | .014 | .014 | -0.04 | .016 | .015 | -0.23 | 4.3 |
| 8 | .100 | .085 | -1.59 | .014 | .011 | -0.84 | .016 | .012 | -1.00 | 5.3 |
| 9 | .114 | .095 | -1.79 | .014 | .011 | -0.75 | .016 | .012 | -0.91 | 6.2 |
| 10 | .127 | .104 | -2.16 | .013 | .008 | -1.24 | .015 | .009 | -1.40 | 8.1 |
| 11 | .140 | .122 | -1.62 | .013 | .018 | 1.33 | .015 | .020 | 1.10 | 9.3 |
| 12 | .152 | .129 | -2.00 | .012 | .007 | -1.34 | .015 | .008 | -1.46 | 11.4 |
| 13 | .164 | .135 | -2.42 | .012 | .006 | -1.56 | .014 | .007 | -1.67 | 14.2 |
| 14 | .175 | .138 | -2.98 | .011 | .004 | -2.11 | .014 | .004 | -2.23 | 19.2 |
| 15 | .186 | .154 | -2.54 | .011 | .016 | 1.33 | .013 | .018 | 1.08 | 20.4 |
| 16 | .196 | .167 | -2.27 | .010 | .013 | 0.80 | .013 | .016 | 0.62 | 20.8 |
| 17 | .206 | .177 | -2.26 | .010 | .010 | -0.10 | .012 | .011 | -0.23 | 20.8 |
| 18 | .216 | .185 | -2.31 | .009 | .008 | -0.33 | .012 | .010 | -0.45 | 21.0 |
| 19 | .225 | .190 | -2.59 | .009 | .005 | -1.31 | .012 | .006 | -1.41 | 23.0 |
| 20 | .234 | .192 | -3.01 | .009 | .002 | -1.96 | .011 | .003 | -2.07 | 27.3 |
| 21 | .242 | .198 | -3.15 | .008 | .006 | -0.75 | .011 | .007 | -0.90 | 28.1 |
| 22 | .250 | .206 | -3.09 | .008 | .008 | 0.13 | .011 | .010 | -0.06 | 28.1 |
| 23 | .257 | .216 | -2.92 | .008 | .010 | 0.63 | .010 | .012 | 0.44 | 28.3 |
| 24 | .265 | .222 | -2.99 | .007 | .006 | -0.46 | .010 | .008 | -0.62 | 28.7 |
| 25 | .272 | .226 | -3.21 | .007 | .004 | -1.20 | .010 | .005 | -1.33 | 30.4 |
| 26 | .279 | .230 | -3.32 | .007 | .005 | -0.70 | .009 | .006 | -0.85 | 31.2 |
| 27 | .285 | .235 | -3.42 | .006 | .005 | -0.62 | .009 | .006 | -0.78 | 31.8 |
| 28 | .291 | .243 | -3.25 | .006 | .008 | 0.78 | .009 | .011 | 0.56 | 32.1 |
| 29 | .297 | .247 | -3.40 | .006 | .004 | -0.91 | .009 | .005 | -1.06 | 33.2 |
| 30 | .303 | .253 | -3.37 | .006 | .006 | 0.07 | .008 | .008 | -0.13 | 33.2 |
| 31 | .309 | .254 | -3.64 | .006 | .001 | -1.70 | .008 | .002 | -1.83 | 36.6 |
| 32 | .314 | .259 | -3.67 | .005 | .005 | -0.23 | .008 | .006 | -0.43 | 36.6 |
| 33 | .319 | .264 | -3.68 | .005 | .005 | -0.15 | .008 | .006 | -0.36 | 36.9 |
| 34 | .324 | .274 | -3.28 | .005 | .011 | 2.38 | .007 | .015 | 2.06 | 36.9 |
| 35 | .329 | .277 | -3.43 | .005 | .002 | -1.01 | .007 | .003 | -1.15 | 37.4 |
| 36 | .334 | .280 | -3.48 | .005 | .004 | -0.44 | .007 | .005 | -0.61 | 37.4 |
| 37 | .338 | .289 | -3.22 | .004 | .008 | 1.69 | .007 | .012 | 1.43 | 37.7 |
| 38 | .342 | .292 | -3.26 | .004 | .004 | -0.32 | .007 | .005 | -0.47 | 37.7 |
| 39 | .346 | .292 | -3.52 | .004 | .000 | -1.87 | .006 | .000 | -1.96 | 40.6 |
| 40 | .350 | .296 | -3.54 | .004 | .004 | -0.20 | .006 | .005 | -0.37 | 40.6 |
| 41 | .354 | .298 | -3.63 | .004 | .002 | -0.70 | .006 | .003 | -0.84 | 41.4 |
| 42 | .358 | .300 | -3.78 | .004 | .001 | -1.21 | .006 | .002 | -1.33 | 41.4 |
| 43 | .362 | .304 | -3.70 | .004 | .005 | 0.55 | .006 | .007 | 0.34 | 41.8 |
| 44 | .365 | .305 | -3.85 | .004 | .001 | -1.14 | .006 | .002 | -1.26 | 41.8 |
| 45 | .369 | .310 | -3.75 | .003 | .005 | 0.68 | .005 | .007 | 0.46 | 42.2 |

TABLE 13, CONTINUED

| N | Predicted cdf | cdf | t | Predicted density | density | t | Predicted hazard | hazard | t | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 46 | .372 | .313 | -3.80 | .003 | .002 | -0.46 | .005 | .003 | -0.62 | 42.2 |
| 47 | .375 | .314 | -3.93 | .003 | .001 | -1.03 | .005 | .002 | -1.16 | 43.7 |
| 48 | .378 | .322 | -3.58 | .003 | .008 | 2.74 | .005 | .012 | 2.40 | 43.7 |
| 49 | .381 | .326 | -3.54 | .003 | .004 | 0.31 | .005 | .005 | 0.12 | 47.1 |
| 50 | .384 | .327 | -3.65 | .003 | .001 | -0.92 | .005 | .002 | -1.04 | 47.1 |
| 51 | .387 | .328 | -3.75 | .003 | .001 | -0.89 | .005 | .002 | -1.01 | 49.2 |
| 52 | .390 | .332 | -3.69 | .003 | .004 | 0.46 | .005 | .005 | 0.27 | 49.2 |
| 53 | .392 | .333 | -3.78 | .003 | .001 | -0.82 | .004 | .002 | -0.95 | 49.4 |
| 54 | .395 | .333 | -3.94 | .003 | .000 | -1.47 | .004 | .000 | -1.56 | 49.4 |
| 55 | .398 | .334 | -4.02 | .003 | .001 | -0.76 | .004 | .002 | -0.89 | 49.4 |
| 56 | .400 | .338 | -3.94 | .002 | .004 | 0.67 | .004 | .005 | 0.46 | 50.8 |
| 57 | .402 | .339 | -4.01 | .002 | .001 | -0.70 | .004 | .002 | -0.83 | 50.8 |
| 58 | .405 | .341 | -4.00 | .002 | .002 | 0.05 | .004 | .004 | -0.12 | 50.8 |
| 59 | .407 | .344 | -3.99 | .002 | .002 | 0.09 | .004 | .004 | -0.09 | 51.1 |
| 60 | .409 | .348 | -3.82 | .002 | .005 | 1.61 | .004 | .007 | 1.35 | 51.1 |
| 61 | .411 | .352 | -3.73 | .002 | .004 | 0.92 | .004 | .005 | 0.71 | 51.1 |
| 62 | .413 | .353 | -3.78 | .002 | .001 | -0.55 | .004 | .002 | -0.68 | 51.9 |
| 63 | .415 | .357 | -3.68 | .002 | .004 | 1.02 | .003 | .006 | 0.81 | 51.9 |
| 64 | .417 | .359 | -3.65 | .002 | .002 | 0.28 | .003 | .004 | 0.12 | 51.9 |
| 65 | .419 | .362 | -3.62 | .002 | .002 | 0.32 | .003 | .004 | 0.16 | 52.3 |
| 66 | .421 | .363 | -3.66 | .002 | .001 | -0.44 | .003 | .002 | -0.57 | 52.3 |
| 67 | .423 | .364 | -3.69 | .002 | .001 | -0.42 | .003 | .002 | -0.54 | 52.3 |
| 68 | .424 | .365 | -3.73 | .002 | .001 | -0.39 | .003 | .002 | -0.52 | 53.2 |
| 69 | .426 | .366 | -3.76 | .002 | .001 | -0.36 | .003 | .002 | -0.49 | 53.2 |
| 70 | .428 | .368 | -3.79 | .002 | .001 | -0.34 | .003 | .002 | -0.47 | 53.2 |

TABLE 14

INDIVIDUAL PREDICTIONS, 1978 VALIDATION SAMPLE
PROPORTIONAL HAZARDS MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 3078 | 36.9 | 37.4 |
| .5% | 15 | 93.3 | 99.5 | 99.5% | 3063 | 36.7 | 37.1 |
| 1% | 31 | 87.1 | 98.0 | 99% | 3047 | 36.4 | 36.8 |
| 5% | 154 | 76.6 | 85.5 | 95% | 2924 | 34.8 | 34.9 |
| 10% | 308 | 68.8 | 76.7 | 90% | 2770 | 33.4 | 33.0 |
| 20% | 616 | 59.7 | 66.8 | 80% | 2462 | 31.2 | 30.0 |
| 30% | 923 | 56.0 | 60.8 | 70% | 2155 | 28.8 | 27.4 |
| 40% | 1231 | 53.6 | 56.2 | 60% | 1847 | 25.8 | 24.9 |
| 50% | 1539 | 51.5 | 52.3 | 50% | 1539 | 22.4 | 22.5 |
| 60% | 1847 | 48.1 | 48.9 | 40% | 1231 | 20.2 | 20.2 |
| 70% | 2155 | 45.0 | 45.8 | 30% | 923 | 18.2 | 17.9 |
| 80% | 2462 | 42.2 | 42.9 | 20% | 616 | 15.9 | 15.5 |
| 90% | 2770 | 40.0 | 40.2 | 10% | 308 | 9.7 | 12.4 |
| 95% | 2924 | 38.5 | 38.8 | 5% | 154 | 7.8 | 10.2 |
| 99% | 3047 | 37.2 | 37.7 | 1% | 31 | 9.7 | 7.2 |
| 99.5% | 3063 | 37.1 | 37.4 | .5% | 15 | 6.7 | 6.3 |
| 100% | 3078 | 36.9 | 36.8 | 0% | 0 | ---- | ---- |

TABLE 15

INDIVIDUAL PREDICTIONS, 1978 VALIDATION SAMPLE
LOGIT LOGNORMAL MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 3078 | 36.9 | 36.8 |
| .5% | 15 | 93.3 | 81.4 | 99.5% | 3063 | 36.7 | 36.6 |
| 1% | 31 | 83.9 | 80.7 | 99% | 3047 | 36.5 | 36.3 |
| 5% | 154 | 78.6 | 74.9 | 95% | 2924 | 34.7 | 34.8 |
| 10% | 308 | 74.0 | 70.0 | 90% | 2770 | 32.8 | 33.1 |
| 20% | 616 | 61.2 | 63.3 | 80% | 2462 | 30.9 | 30.1 |
| 30% | 923 | 57.1 | 58.7 | 70% | 2155 | 28.3 | 27.4 |
| 40% | 1231 | 54.8 | 54.9 | 60% | 1847 | 25.0 | 24.7 |
| 50% | 1539 | 51.3 | 51.5 | 50% | 1539 | 22.6 | 22.1 |
| 60% | 1847 | 48.0 | 48.3 | 40% | 1231 | 20.4 | 19.5 |
| 70% | 2155 | 44.9 | 45.3 | 30% | 923 | 18.4 | 16.8 |
| 80% | 2462 | 42.2 | 42.5 | 20% | 616 | 16.1 | 13.9 |
| 90% | 2770 | 39.9 | 39.7 | 10% | 308 | 10.4 | 10.6 |
| 95% | 2924 | 38.5 | 38.3 | 5% | 154 | 7.8 | 8.5 |
| 99% | 3047 | 37.2 | 37.1 | 1% | 31 | 6.5 | 5.4 |
| 99.5% | 3063 | 37.1 | 36.9 | .5% | 15 | 6.7 | 4.6 |
| 100% | 3078 | 36.9 | 36.8 | 0% | 0 | ---- | ---- |

TABLE 16

INDIVIDUAL PREDICTIONS, 1978 VALIDATION SAMPLE
LOGIT/INDIVIDUAL LOGNORMAL MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 3078 | 36.9 | 36.8 |
| .5% | 15 | 93.3 | 93.8 | 99.5% | 3063 | 36.7 | 36.5 |
| 1% | 31 | 87.1 | 91.4 | 99% | 3047 | 36.4 | 36.3 |
| 5% | 154 | 79.2 | 79.5 | 95% | 2924 | 34.7 | 34.6 |
| 10% | 308 | 72.2 | 72.1 | 90% | 2770 | 33.0 | 32.9 |
| 20% | 616 | 61.2 | 64.1 | 80% | 2462 | 30.9 | 30.0 |
| 30% | 923 | 56.9 | 59.0 | 70% | 2155 | 28.6 | 27.3 |
| 40% | 1231 | 53.9 | 55.0 | 60% | 1847 | 25.6 | 24.7 |
| 50% | 1539 | 51.3 | 51.5 | 50% | 1539 | 22.5 | 22.2 |
| 60% | 1847 | 48.3 | 47.9 | 40% | 1231 | 20.6 | 19.6 |
| 70% | 2155 | 44.7 | 45.3 | 30% | 923 | 18.9 | 17.0 |
| 80% | 2462 | 42.1 | 42.5 | 20% | 616 | 16.2 | 14.2 |
| 90% | 2770 | 40.0 | 39.7 | 10% | 308 | 9.1 | 10.9 |
| 95% | 2924 | 38.5 | 38.3 | 5% | 154 | 7.8 | 8.7 |
| 99% | 3047 | 37.2 | 37.1 | 1% | 31 | 6.5 | 5.6 |
| 99.5% | 3063 | 37.1 | 37.0 | .5% | 15 | 6.7 | 4.7 |
| 100% | 3078 | 36.9 | 36.8 | 0% | 0 | ---- | ---- |

TABLE 17

INDIVIDUAL PREDICTIONS, 1980 VALIDATION SAMPLE
PROPORTIONAL HAZARDS MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 4304 | 36.9 | 35.6 |
| .5% | 22 | 90.9 | 98.2 | 99.5% | 4284 | 36.7 | 35.3 |
| 1% | 43 | 88.4 | 96.0 | 99% | 4261 | 36.4 | 35.0 |
| 5% | 215 | 69.8 | 82.3 | 95% | 4089 | 35.2 | 33.2 |
| 10% | 430 | 66.3 | 74.5 | 90% | 3874 | 33.7 | 31.3 |
| 20% | 861 | 59.1 | 65.9 | 80% | 3443 | 31.4 | 28.0 |
| 30% | 1291 | 54.5 | 59.9 | 70% | 3013 | 29.4 | 25.2 |
| 40% | 1722 | 51.2 | 55.3 | 60% | 2582 | 27.4 | 22.5 |
| 50% | 2152 | 48.5 | 51.3 | 50% | 2152 | 25.4 | 19.9 |
| 60% | 2582 | 46.2 | 47.8 | 40% | 1722 | 23.1 | 17.3 |
| 70% | 3013 | 43.5 | 44.6 | 30% | 1291 | 21.5 | 14.6 |
| 80% | 3443 | 41.2 | 41.6 | 20% | 861 | 20.0 | 11.7 |
| 90% | 3874 | 39.1 | 38.7 | 10% | 430 | 17.2 | 8.2 |
| 95% | 4089 | 38.1 | 37.2 | 5% | 215 | 15.3 | 6.1 |
| 99% | 4261 | 37.2 | 35.9 | 1% | 43 | 9.3 | 3.5 |
| 99.5% | 4282 | 37.1 | 35.8 | .5% | 22 | 9.1 | 2.8 |
| 100% | 4304 | 36.9 | 35.6 | 0% | 0 | ---- | ---- |

TABLE 18

INDIVIDUAL PREDICTIONS, 1980 VALIDATION SAMPLE
LOGIT LOGNORMAL MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 4304 | 36.9 | 34.9 |
| .5% | 22 | 81.8 | 75.9 | 99.5% | 4284 | 36.7 | 34.7 |
| 1% | 43 | 86.0 | 74.8 | 99% | 4261 | 36.4 | 34.5 |
| 5% | 215 | 70.7 | 69.0 | 95% | 4089 | 35.2 | 33.1 |
| 10% | 430 | 65.8 | 65.0 | 90% | 3874 | 33.7 | 31.6 |
| 20% | 861 | 59.8 | 60.1 | 80% | 3443 | 31.2 | 28.6 |
| 30% | 1291 | 55.1 | 56.1 | 70% | 3013 | 29.2 | 25.8 |
| 40% | 1722 | 51.3 | 52.6 | 60% | 2582 | 27.4 | 23.1 |
| 50% | 2152 | 48.6 | 49.5 | 50% | 2152 | 25.3 | 20.3 |
| 60% | 2582 | 46.2 | 46.6 | 40% | 1722 | 23.1 | 17.4 |
| 70% | 3013 | 43.3 | 43.7 | 30% | 1291 | 22.1 | 14.4 |
| 80% | 3443 | 41.0 | 40.9 | 20% | 861 | 20.6 | 11.1 |
| 90% | 3874 | 39.2 | 38.0 | 10% | 430 | 17.0 | 7.2 |
| 95% | 4089 | 38.1 | 36.5 | 5% | 215 | 14.9 | 5.0 |
| 99% | 4261 | 37.2 | 35.2 | 1% | 43 | 11.6 | 2.4 |
| 99.5% | 4282 | 37.1 | 35.1 | .5% | 22 | 9.1 | 1.8 |
| 100% | 4304 | 36.9 | 34.9 | 0% | 0 | ---- | ---- |

TABLE 19

INDIVIDUAL PREDICTIONS, 1980 VALIDATION SAMPLE
LOGIT/INDIVIDUAL LOGNORMAL MODEL

| Upper percentile | n | Percent recidivists | Predicted | Lower percentile | n | Percent recidivists | Predicted |
|---|---|---|---|---|---|---|---|
| 0% | 0 | ---- | ---- | 100% | 4304 | 36.9 | 35.1 |
| .5% | 22 | 77.3 | 88.9 | 99.5% | 4284 | 36.7 | 34.8 |
| 1% | 43 | 79.1 | 85.1 | 99% | 4261 | 36.5 | 34.6 |
| 5% | 215 | 69.3 | 73.2 | 95% | 4089 | 35.2 | 33.1 |
| 10% | 430 | 64.7 | 67.1 | 90% | 3874 | 33.9 | 31.5 |
| 20% | 861 | 58.3 | 60.7 | 80% | 3443 | 31.6 | 28.7 |
| 30% | 1291 | 54.3 | 56.3 | 70% | 3013 | 29.5 | 26.0 |
| 40% | 1722 | 50.7 | 52.8 | 60% | 2582 | 27.8 | 23.3 |
| 50% | 2152 | 48.7 | 49.7 | 50% | 2152 | 25.2 | 20.5 |
| 60% | 2582 | 46.6 | 46.8 | 40% | 1722 | 22.5 | 17.5 |
| 70% | 3013 | 43.5 | 44.0 | 30% | 1291 | 21.5 | 14.3 |
| 80% | 3443 | 41.2 | 41.1 | 20% | 861 | 20.0 | 10.9 |
| 90% | 3874 | 39.1 | 38.2 | 10% | 430 | 17.2 | 7.0 |
| 95% | 4089 | 38.0 | 36.7 | 5% | 215 | 16.7 | 4.6 |
| 99% | 4261 | 37.2 | 35.4 | 1% | 43 | 14.0 | 2.0 |
| 99.5% | 4282 | 37.1 | 35.3 | .5% | 22 | 4.5 | 1.5 |
| 100% | 4304 | 36.9 | 35.1 | 0% | 0 | ---- | ---- |