NUMBER 29

TESTING THE PREDICTIVE POWER OF A PROPORTIONAL HAZARDS SEMI-MARKOV MODEL OF POSTENTITLEMENT HISTORIES OF DISABLED MALE BENEFICIARIES

John C. Hennessey
Division of Disability Studies

NOVEMBER 1982

Social Security Administration
Office of Policy
Office of Research and Statistics

Abstract

In the Disability Amendments of 1980 (P.L. 96-265), Congress mandated that certain experiments be carried out which are designed to encourage disabled beneficiaries to return to work and save trust fund monies. A research plan has been developed which would offer alternative program provisions, experimentally, to different samples of beneficiaries. An observation period of three to four years will be possible before a report to Congress must be written. However, a period of this length is not sufficient to observe, fully, the postentitlement experience of disabled beneficiaries. In order to estimate the long run effects of the experiments, a method is needed which can project postentitlement behavior beyond the observation period.

This paper tests the ability of a proportional hazards semi-Markov model to make accurate predictions in this type of setting. The data are divided into two segments: the first 14 calendar quarters and the last 16 quarters. Various types of rate functions including proportional hazards rate functions are estimated on the first segment, then projected over the entire 30 quarters and compared to the actual data. The proportional hazards rate functions are then used in a simulation to estimate monthly benefit cost to the social security disability trust fund over the last 16 quarters, using an age-dependent, absorbing, semi-Markov model. The model does a very good job of capturing the dynamics of the process and should prove quite useful as one of the major components in an analysis of the Work Incentive Experiments.

Acknowledgements

The author wishes to thank Barry V. Bye for his constant support and advice throughout the entire project, Katherine Merrick for her technical assistance, and Sharon Long for preparation of the manuscript.

Table of Contents

List of Tables

Page

Figures

Introduction

In the Social Security Disability Amendments of 1980 (P.L. 96-265) the Congress gave the Secretary of Health and Human Services the authority to waive the normal provisions of Title II of the Social Security Act and mandated that he use this authority to carry out experiments which are designed to encourage disabled beneficiaries to return to work and leave the benefit rolls. The primary objective of the experiments is to save trust fund monies and control the costs of the Social Security Disability Insurance program. The bill itself contains several examples of the kind of changes that Congress had in mind. These include changes in the entitlement provisions for Medicare benefits, lengthening of the trial work period and alternative ways of treating postentitlement earnings such as the application of a benefit offset based on earnings such as the application of a benefit offset based on earnings, as in the retirement benefit program.

In order to measure the effect of these types of changes, the Office of Research and Statistics of the Social Security Administration has developed a plan 1/ which offers various alternative program provisions to five groups of newly entitled beneficiaries. Four of the groups will be offered one of the following alternatives: (1) benefits reduced 50 cents for each dollar of earnings when earnings exceed substantial gainful activity (and trial work is completed); (2) benefits reduced one dollar for each dollar of earnings; (3) more trial work months (15 instead of 9) and earlier eligibility for Medicare (after the sixth month of trail work); and (4) five years of Medicare coverage after completion of a trial work period (provided there is no medical recovery). The fifth group will be offered both the first and fourth alternatives on the assumption that this particular combination is likely to provide the most powerful work incentive and is therefore the best test of the efficacy of any of these changes. A control group will also be established.

Because all of the experimental alternatives will be more liberal than the current law, they will exhibit short run costs to the trust funds which are greater than those that are obtained under the current law. The potential for disability trust fund savings comes from sustained increases in work on the part of larger numbers of beneficiaries and significant increases in the termination of cash benefits.

Although short run labor force response will be of much interest, there is no necessary relationship between an early increase in work attempts and long run trust fund effects. At the same time, the reporting requirements to Congress preclude observations much beyond 4 or 5 years. It will, therefore, not be possible to directly observe long run effects on trust funds. Instead, estimates of these effects will have to come from mathematical models defined so that long run effects can be predicted in terms of estimated parameters using the several years of data available.

Because the focus of the experiments is on work incentives and on trust fund costs, the models developed should have the ability to answer such questions as: How many disabled beneficiaries, under each experimental treatment, are expected to make a work attempt? What is the expected length of time from date of award before a beneficiary tries to work? What is the probability that this work attempt will lead to a secure job and allow the beneficiary to leave the SSDI program? What is the expected length of time in the program for beneficiaries in each experimental group? What is the expected cost in family benefits?

In order to answer these types of questions, an analysis plan is being proposed which is based on a continuous time stochastic model. This decision is based on several factors. First, these models have been used quite successfully in the life sciences. The work of Mode, Littman and Soyka 2/ provides a nice example of this approach. The work of Tuma, Hannan, and Groeneveld 3/ provides an example of the use of stochastic models in the social sciences. Their interest, however, has been in explanatory models. Flinn and Heckman 4/ have been developing stochastic models for the analysis of labor force dynamics, but, these also are explanatory in nature.

Even though our main interest will be predicting long run effects of the experiments, the short run labor force response will be of considerable interest. The works of Tuma, et. al. and Flinn and Heckman present convincing arguments that the rate at which events are occuring over time should be the central focus of this type of analysis. Consider two groups with different experimental provisions. Suppose, after 2 years, we observe that the number of beneficiaries who made a work attempt in each group is the same, i.e., not statistically different. It is conceivable that two different processes are at work. One group may exhibit a very high rate of return to work in the early months with the rate consistently dropping to zero in the two years. The other group could have a very low rate initially, with a consistent increase over the two years. Both processes yield the same number of work attempts over the first two years. But, clearly, they are different processes and they will generate different results in the long run. The work attempts in the first group are essentially over, but the second group has not yet reached its full potential. Therefore, any short run analysis should include an analysis of the rates at which events are occurring over time.

As will be shown in detail later on, estimates of these rates can also be used to build a model which can predict all the long run effects mentioned above. All the quantities of interest are generated from the same model. It is important, then, to see how successful this approach is when dealing with postentitlement histories of disabled beneficiaries.

A first analysis of postentitlement histories of disabled male beneficiaries using stochastic models was made by Hennessey 5/. In that effort, an age dependent, absorbing, semi-Markov (AASM) model was developed and shown to have the ability to account for the postentitlement work patterns for disabled males. However, seven years of data were used to estimate the model, and the rates described above were not estimated.

In this paper the initial work is extended in several ways. The AASM model is generated through the estimation of the rate functions for male beneficiaries mentioned above. Exogenous variables are introduced into the rate functions and the increase in predictive power assessed. The model is estimated using an observation period of three and one half years and then tested by comparing estimated and actual behavior over an additional four year period.

The State Space

The analytic approach in this paper treats the postentitlement history as a stochastic process defined over some state space. The state space is the set of possible values or categories that a person can attain in the process being studied. The focus is on a set of discrete categories representing program and work status, starting at the time of initial award of benefits and ending with the termination of benefits for this initial period of disability. Termination of benefits can occur for three reasons: (1) The beneficiary can be terminated because the disabling condition has improved, so that the definition of disability is no longer satisfied; this is called a medical recovery. (2) The beneficiary can be terminated because he has demonstrated the ability to engage in substantial gainful activity (SGA), but the disabling conditions remain severe; this is called a work recovery. (3) The beneficiary may die. It is worth noting that the two types of recoveries mentioned are based on administrative decisions and that the available data do not distinguish between them. Thus, for this analysis, two reasons for termination will be possible, recovery or death.

During the initial period in the disability program, the work pattern is important since this will affect the length of time in the program and since the Work Incentive Experiments are aimed at modifying work behavior. Generally, it is important to distinguish between various types of work situations, since this could affect the work episode or subsequent episodes. For example, if the monthly earnings are above SGA, then the beneficiary will most likely be terminated from the program after completion of the trial work period. If monthly earnings are below SGA, termination will not occur. The way a beneficiary enters and leaves a job may also be important. If he returns to work for a previous employer, experience with the type of job and support from previous friends on the job could provide a better chance for a work recovery. If the work stops because of health reasons, as opposed to being laid off, the subsequent nonwork episode would be lengthened. However, the data available at this time permit only the distinction between working and not working. Therefore, the state space for this analysis admits only four possibilities in the combined administrative and work process:

E_1: Recovered

E_2: Deceased

E_3: Nonworking Beneficiary

E_4: Working Beneficiary

A beneficiary at time of award is assumed to be in state E_3. As time progresses, he can switch between states E_3 and E_4 until he finally switches to state E_1 or E_2, in which case the first episode in the disability program is over. State E_2 is an absorbing state since, once a beneficiary enters this state, he cannot leave it. Since our focus is on the first period of disability, we shall designate state E_1 as an absorbing state. When a disability beneficiary reaches age 65, he is converted to the retirement program. Since we are only interested in cost to the disability trust fund, all calculations will be done from time of award to age 65 or entry to state E_1 or E_2, whichever is first. Figure 1 describes this process.

The Rate Functions

The proposed model attempts to make projections by accounting for the rates of movement from states E_3 and E_4 to the other states, as functions of time, age, and other relevant variables. More precisely, for i=3 or 4 and j=1,2,3,4, $j \neq i$, we let $P_{ij}(s,t)$ represent the probability that a person in state E_i at time s will be in state E_j at time t. For example, $P_{31}(3,5)$ represents the probability that a beneficiary who is not working at time 3 is recovered at time 5.

We define $r_{ij}(t)$ as the instantaneous rate of transition from state E_i, i=3,4 to state E_j, j=1,2,3,4,$j \neq i$, at time t. This rate, $r_{ij}(t)$, is the limit, as Δt approaches 0, of the probability of a change from state E_i to state E_j between times t and $t+\Delta t$, per unit of time:

$$r_{ij}(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t,t+\Delta t)}{\Delta t} \qquad (1)$$

Intuitively, this represents the instantaneous rate of flow from E_i to E_j at time t.

These rate functions can be considered the "atoms" of the process since no mathematical restrictions are placed on them other than the fact that they are nonnegative. All other quantities are derivable from them. First passage probabilities are of special interest. First passage probabilities are the probabilities that various events will take place within a certain amount of time. We define $A_{ij}(t)$, for i=3,4 and $j \neq i$, to be the probability that a person, who enters state E_i, will next switch to state E_j within t time units after entering E_i. For example, if i=3 and j=4, then $A_{34}(t)$ is the probability that a beneficiary will make a work attempt sometime within the next t time units.

The relationship between the first passage probabilities and the rates can be shown as follows. Let $A_i(t)$ represent the probability that a person who enters state E_i at some time will leave that state within the next t time units. That is,

$$A_i(t) = \sum_{\substack{j=1 \\ j\neq i}}^{4} A_{ij}(t)$$

It follows from the definition of $P_{ij}(t,t+\Delta t)$ that, for small Δt:

$$P_{ij}(t,t+\Delta t) = \{A_{ij}(t+\Delta t) - A_{ij}(t)\}/\{1-A_i(t)\} \qquad (2)$$

Thus, from formulae (1) and (2), we obtain:

$$r_{ij}(t) = \lim_{\Delta t\to 0} \frac{A_{ij}(t+\Delta t) - A_{ij}(t)}{\Delta t\{1-A_i(t)\}}$$

$$= \lim_{\Delta t\to 0} \left(\frac{A_{ij}(t+\Delta t)-A_{ij}(t)}{\Delta t} \right) \cdot \frac{1}{1-A_i(t)}$$

or

$$r_{ij}(t) = \frac{a_{ij}(t)}{1-A_i(t)} \qquad (3)$$

where $a_{ij}(t) = A'_{ij}(t)$, is the density function for $A_{ij}(t)$.

To solve for $A_{ij}(t)$, formula (3) is first summed over j, to obtain:

$$r_i(t) = \sum_{\substack{j=1 \\ j\neq i}}^{4} r_{ij}(t) = \frac{a_i(t)}{1-A_i(t)}$$

where

$$a_i(t) = A_i'(t)$$

Thus,

$$\int_0^t r_i(s)\, ds = \int_0^t \frac{a_i(s)ds}{1-A_i(s)} = -\log\left(1 - A_i(t)\right)$$

or

$$A_i(t) = 1 - \exp\left(-\int_0^t r_i(s)ds\right) \qquad (4)$$

Thus, from formulae (3) and (4),

$$a_{ij}(t) = r_{ij}(t) \exp\left(-\int_0^t r_i(s)\, ds\right) \qquad (5)$$

and so

$$A_{ij}(t) = \int_0^t r_{ij}(u) \exp\left(-\int_0^u r_i(s)ds\right) du \qquad (6)$$

Therefore, a specification for the rate functions completely determines the first passage probability functions, $A_{ij}(t)$.

The rate functions can be made as simple or as complex as desired, guided only by the underlying conceptual structure and the availability of data to estimate the structure. In their most general form, they will depend on different types of variables. We shall call them static, clock, stochastic, and linkage variables.

A static variable is one which doesn't change over time, such as sex, race, diagnosis of primary disabling medical condition at time of initial award.

A clock variable is augmented as a nonstochastic function of time. Age at entry to a given state is one example.

A stochastic variable is one which changes randomly over time according to some process exogenous to the process being modeled. In the current context, variables such as the severity of the impairment of the beneficiary and the beneficiary's local labor market conditions are examples of possible relevant stochastic variables.

The fourth type of variable is so named because it establishes the linkage between the past, present, and future of the process. A linkage variable is one whose value is determined by the past events of the process being modeled. For example, the rate functions describing a third work episode might depend on the length of the previous nonwork episode, the length of the previous work episode, or the reason why the last work episode was terminated.

In the Work Incentive Experiments, three or four years of data will be available. During that time, work episodes beyond the second will most likely not be observed. Even second work episodes will, quite likely, not be completed during the observation period Thus, it may not be possible to establish or estimate a complex linkage. Also, no stochastic variables useful to this analysis are available. It would be of interest to know how well a model with only static and clock variables can make the type of predictions desired with three or four years of data. In the literature, this is called an age dependent semi-Markov model.

The Data

In order to test this model, a data file was extracted from the Social Security Continuous Disability History (CDHS) Sample. The CDHS is a longitudinal data file. New applicants for disability insurance benefits meeting the selection criteria are added to the file each year. Through annual updates, the file is extended to accumulate new earnings and entitlement data for each applicant. 6/

The data file for this analysis contains information on a subsample of males from the CDHS who were entitled to benefits in 1969. There were 20,749 persons. This constitutes a 10 percent self weighting sample of the population. For each sample case, the following information was obtained:

1) Month and year of birth.

2) Diagnosis of Primary Disabling Condition at time of award 7/

3) Monthly Benefit Payable - the monthly benefit paid to the beneficiary after adjustments for supplemental medical insurance benefits , decided at time of award.

4) Quarters of Coverage patterns, for each year from 1967 through 1977, if available.

5) Self-employment quarters of coverage (0 or 4) for each year, 1967 through 1977.

6) Agricultural Wages for each year, 1967 through 1977.

7) Date of Award of Benefits

Using the data items (4) through (7) a postentitlement history is computed. Each history takes the following form:

$$\{ T_o, X_o, T_1, X_1, \ldots T_k, X_k \} \qquad (7)$$

Since the CDHS records quarters of coverage on a quarterly basis, it was decided to measure time in quarters, with t = 0 on January 1, 1969. T_o equals the time of award of benefits. X_o is the initial state in the program, which is state E_3. From then on, whenever the person switches to a new state, the time T_i, and the new state, X_i, are recorded. This continues until the person enters either an absorbing state, E_1 or E_2, or the end of the data file is reached on December, 1977.

Using the CDHS to ascertain actual labor force status of beneficiaries at a particular point in time is problematic, given the fact that the data contain only quarters of coverage, i.e., posted earnings of $50 or more for a given quarter. Investigation of folders of beneficiaries have indicated that a posted quarter of coverage does not always correspond to a quarter of work. For example, it may represent late pay or sick pay. Some-times it represents sheltered work or some situation where substantive services are not performed. The Social Security Administration does not consider these situations as work attempts. This difficulty is compounded by the fact that, for annual earnings from self employment, the person is given either 0 or 4 quarters of coverage each year. Also, for annual earnings from agricultural work, only the number of quarters for that year is recorded. Other earnings are reported annually, not quarterly. And so, the actual pattern of quarters of work for the year is not known. In the data gathering scheme for the Work Incentive Experiments, this problem will not occur. However, for the CDHS data, the following decisions were made to construct the postentitlement history:

1) To compute the quarterly pattern of work for each year, following date of award:

   (a) Compute the quarters of coverage from self-employment. This is either 0 or all 4 quarters.

   (b) If (a) gives all 4 quarters, then assume he has worked all year. If not, compute the quarters for agricultural wages as follows:

      If agricultural wages ($) are less than $100, then no quarters worked.

      If $100 \leq \$ < 200$, then third quarter worked.

      If $200 \leq \$ < 300$, then second and third quarter worked.

      If $300 \leq \$ < 400$, then second, third, and fourth quarter worked.

      If $\$ \geq 400$, then all 4 quarters worked.

(c) Superimpose any quarterly pattern from other wages over the above.

2) If the quarter at time of award now indicates that the person is working, it is assumed that this comes from artificial sources and is not real work. Therefore, this quarter and all consecutive quarters of work are changed to nonwork. For example, if the quarterly pattern from time of award looks like

W, W, W, N, N, W, W, N,....

where W means work and N means nonwork, this is converted to:

N, N, N, N, N, W, W, N,....

3) From this quarterly pattern, the postentitlement history in formula (7) is constructed.

Clearly, these are crude adjustments and therefore, the predictions here are not to be construed as accurate predictions about the program. However, it was decided to be sufficient for this test of the model.

The First Nonwork Episode

Since everyone starts in state E_3 at time of award, this episode is investigated first. We wish to construct a model of the rates of leaving state E_3 for states E_1, E_2, and E_4 which will capture, as much as possible, the actual rates. This model will consist of certain functional forms for $r_{3j}(t)$, $j=1,2,4$, with a number of parameters to be estimated in each form. There are two issues present when choosing these functions. First, we must understand the basic shape of the actual rate functions; and, second, we must understand how the actual rates differ by exit state, j=1,2,4 and by different subpopulations, who might exhibit different rates. In order to do this the actual rates for different subpopulations are computed and graphed. To compute the actual rate, $r_{3j}(t)$, of leaving state E_3 for state E_j, at time t, the number of beneficiaries who switch from state E_3 to E_j from time t-1 to t is divided by the number of people in state E_3 at time t-1.

At first, the population is divided into 3 groups, based on the monthly benefit payable at time of award. The groups are: $0 - 149 , 150 - 249 , and $250 or more. Each of these groups is further subdivided by age. The age groups are: 18 - 30, 31 - 45, 46 - 65. For each of the nine subpopulations, the graphs for the 3 rates of exit from state E_3 were drawn. Since there are many graphs to consider, we present only typical examples of the results. Figure 2 contains the graphs for the middle monthly benefit group, all age groups. The graph which describes the rates of return to work for the three age groups, suggests that the rates for each age group have the same basic shape, but are shifted vertically. The more erratic behavior of the graph for the young age group could be due to the smaller size of the group. In the middle benefit group, 363 beneficiaries in the young age group, 552 beneficiaries in the middle age group, and 1,515 beneficiaries in the old age group return to work. The death rates and recovery rates show the same pattern.

If the recovery rates, death rates, and return to work rates are compared to each other, there is also a similarity to be noticed. They all seem to possess the same basic shape--high at the beginning and then tapering off. However, they seem to differ by more than some vertical shift. The death rates, for example, exhibit very little decline after the eighth quarter. The recovery rates, on the other hand, show a severe drop off from the eighth to the sixteenth quarter. The decreasing trend could be due to the fact that we still have a heterogeneous population with regard to the rate functions. This means, for example, that the middle benefit, old age group actually consists of two or more subgroups, each of which have different rate functions. The subpopulations with the higher rates will exit from state E_3 in the earlier times, leaving those subpopulations with the lower rates dominant in the later times. Hence the overall rate will decline over time. It may be the case, however, that all subpopulations actually have declining rates over time. 8/

To further investigate this phenomenon, the middle benefit, old age group was divided into categories based on the diagnosis of primary disabling condition. These categories are the type that are often used in other studies. The categories are: Circulatory System, Musculoskeletal System, Mental Disorders, Accidents, Neoplasms, Respiratory System, and all others combined. The graphs for the categories of Accidents and Neoplasms are presented in Figure 3. Again, the rate functions for recovery and return to work suggest the same pattern as before. The death rate for neoplasms is also similar. But, the death rate for the accident group seems to have an increasing trend, which suggests that some of the decreasing trend in the overall population rates is due to population heterogeneity. On the other hand, the differences in rates among the subpopulations is not a point of the analysis. The main purpose is to accurately account for the total population. The differences in the shapes of the rate functions for the different exit states are enough to decide that separate estimates are needed for each exit state. But, separate estimates for each subpopulation may be unnecessary and even damaging to the predictive accuracy for the whole population, because of the decrease in sample size of each subpopulation each time we subdivide the total population into more categories.

To investigate this phenomenon and to test the predictive accuracy as well, we shall assume that we only have the first 14 quarters of observation, beginning in January 1, 1969. This 3½ year period also more accurately reflects the length of the observation period for the Work Incentive Experiments. All decisions shall be made based on the data from these 14 quarters. The estimates will be computed for each monthly benefit group since they will be needed later to estimate costs. The predictions of the first passage probabilities, $A_{3j}(t)$, will be computed, using formula (6), and compared to the actual data, since this is a quantity of interest. More specifically, the following strategy will be employed:

For each monthly benefit group,

1) Compute estimates of the rate functions from state E_3 to all other states for the total population, using the first 14 quarters of data; this will be called the Total Population Model.

2) Repeat part one separately for each age group, then average the results to form the Age Model predictions.

3) For the middle benefit, old age group, repeat part one. Then estimate separately for each diagnostic group. Average the results to form the Diagnostic Model predictions.

4) For all models, compute the first passage probabilities from state E_3 to all other states for 7½ years, by year, and compare to the actual data.

Estimating the Rate Functions

As mentioned above, a function for r(t) is needed that can capture the basic shapes in figures 2 and 3. The analysis so far suggests that the same functional form might serve well to model all of the rates for all subpopulations. On the other hand, it should not be overly complicated. The one chosen is a combination of a number of those found in the literature: 9/

$$r(t) = exp\left(\alpha+\beta t+\gamma \log(t+1)\right) \qquad (8)$$

The exponential function is used to guarantee positivity of the rate fucntion. The constant term α can allow for the vertical displacement observed in the actual rates. The two terms, βt and $\gamma \log(t+1)$, can combine together to create large values for the rates for small t. Also, when t is large, βt will dominate to create a linear type trend for large t. Since testing various functional forms is not the focus of this work, and since this form allows a decent amount of flexibility, formula (8) is used for all rate functions, with the condition that, if a coefficient is not statistically significant, that term would be eliminated and the simpler form would be estimated and used.

The rate functions are then estimated using the method of maximum liklihood. The likelihood, L, becomes the product of the probability densities of the duration times for the initial nonwork episode of each individual. For an episode whose end is not observed, i.e., a right censored episode, the probability that the episode lasts longer than the observation period is included.

Thus, the likelihood function is:

$$L = \left[\prod_{\substack{j=1 \\ j\neq 3}}^{4} \prod_{t=1}^{14} a_{3j}(t)^{N_{3j}(t)} \right] \cdot \left[\prod_{t=1}^{14} 1-A_3(t)^{N_3(t)} \right] \qquad (9)$$

where: $N_{ij}(t)$ is the number of persons having an episode in state E_i, ending in state E_j, of duration t time units.

$a_{ij}(t)$ is the probability density of such an episode, i.e., a first passage probability density, defined in formula (3).

$N_i(t)$ is the number of individuals with a right censored episode in state E_i with an observation period in this episode of t time units.

$A_i(t)$ is the probability that a person who enters state E_i at some time will leave that state within the next t quarters, defined in formula (2).

In our case, i = 3, so we will drop this subscript and write:

$$L = \left[\prod_{\substack{j=1 \\ j\neq 3}}^{4} \prod_{t=1}^{14} a_j(t)^{N_j(t)} \right] \cdot \left[\prod_{t=1}^{14} \left(1-A(t)\right)^{N(t)} \right] \tag{10}$$

As usual, to simplify calculations, the log of the likelihood is maximized. Taking logs, we obtain:

$$\log L = \left[\sum_{\substack{j=1 \\ j\neq 3}}^{4} \sum_{t=1}^{14} N_j(t) \log \left(a_j(t)\right) + \sum_{t=1}^{14} N(t) \log \left(1 - A(t)\right) \right] \tag{11}$$

We replace $a_j(t)$ with formula (5) and $A(t)$ with formula (4) to obtain:

$$\log L = \sum_{\substack{j=1 \\ j\neq 3}}^{4} \sum_{t=1}^{14} N_j(t) \left[\log r_j(t) - \int_0^t r(s)ds \right] - \sum_{t=1}^{14} N(t) \int_0^t r(s)ds$$

$$= \sum_{\substack{j=1 \\ j\neq 3}}^{4} \left\{ \sum_{t=1}^{14} \left[N_j(t) \log r_j(t) - N^*(t) \int_0^t r_j(s)ds \right] \right\}$$

where $N^*(t) = \sum_{\substack{j=1 \\ j\neq 3}}^{4} N_j(t) + N(t)$

Another advantage to using rate functions now appears. If it is assumed that the rate functions, $r_j(t)$, have no common parameters, then each one can be estimated separately from the others -- even with censored data. This, in fact, was done. For each exit state, and for each model described above in the strategy, the rate functions were estimated, using a standard Newton-Raphson technique. 10/

Comparison of Results

Table 1 presents the coefficient estimates for the rate functions for each benefit group. The total population model (ALL) estimates are presented, along with the separate estimates by age group. Table 2 presents the coefficient estimates for the middle benefit group, oldest age group by diagnosis of primary disabling condition. It is worth noting in table 2 that, for the accident group, the β coefficient for the death rate is positive and significant. This lends credence to our earlier suspicions about heterogeneity.

These rate functions are then used to compute the first passage probabilities from state E_3 to all other states by year. The graphs appear in figure 4. In general, even the simple Total Population Model does a decent job of predicting the first passage probabilities in most categories. It is not surprising that the model does well for the first 3½ years, since these are the years which are used to estimate the model. The final four years are projections beyond the observation period. The worst case is the probabilities of death for the high monthly benefit group, in figure 4C. However, the empirical data take a sharp turn upward after 3 years, which is just after the data use to estimate the model. In almost all cases, the Total Population Model slightly underpredicts the probabilities at the beginning and the end. These estimates improve with the Age Model, although the same pattern is evident. This difficulty is, perhaps, a problem with the functional form.

In similar fashion, separate estimates were made for each diagnostic group in the middle benefit, old age group. The coefficient estimates are presented in table 2. The Total Population model for this group is the same as the Age model in the previous discussion for this group.

Using these estimates, the Diagnostic Model is constructed for the middle benefit old age group. The first passage probabilities for the Total Population Model for this group is compared to the Diagnostic Model in figure 5. Again, some improvement is noted, especially in the probabilities to death.

Thus, it seems that estimating a mixture of rate functions, even though the smaller subsamples may cause the use of a simpler functional form, (even constant in some cases), is preferable to estimating one rate function on the larger sample, in cases where prediction beyond the observation period is important.

A Proportional Hazards Model

In general, estimating rate functions for each subpopulation of interest will not be a practical approach, especially if one considers that this also must be done for the work episodes. However, as noted before, figures 2 and 3 suggest that, in most cases, the same rate function for the different subpopulations are similar, except for a vertical type shift. The one notable exception is the death rates in figure 3. The death rates for these individuals with neoplasms in the first few years is unusually high, compared to other types of disabling conditions. All others, however, seem to follow the vertical shift pattern, and, therefore, suggest the possibility of forcing the coefficients $\bar{\beta}$ and $\bar{\gamma}$ of the rate function in formula (8) to be the same for every subpopulation and allowing only the constant α to vary across subpopulations. This can be accomplished by writing:

$$r(t) = \exp\left(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_m X_m + \beta t + \gamma \ln(t+1)\right) \qquad (13)$$

Where $X_1, X_2, \ldots, X_m$ are dummy variables for the different categories of age, primary diagnosis, and monthly benefit. This is called a proportional hazards model. 11/ Since all of the data are used to estimate each rate function, more categories can be handled and so the number of age groups and monthly benefit groups are increased. The new age groups are: 18-30, 31-45, 46-55, 56-62. The ages 63-65 are dropped since their future in the program is short. Thus, their behavior might be different and bias the results. They could have been modeled separately, but this is not done. The new monthly benefit groups are: $0-100, 100-150, 150-200, 200-250, 250-300, and larger than $300. All categories are entered additively, with no interaction. The same maximum likelihood estimation method is used, with the rate functions replaced with the new formula (13).

Table 3 contains the estimates of the coefficients. The coefficient estimates indicate the expected differences between age groups. For example, the younger age group has a higher recovery rate and return to work rate than the older group. The reverse is true for the death rate. It is interesting to note that all the coefficients on the primary diagnosis variables for the return to work from the nonwork episode are not statistically significant. This is possibly due to the fact that the diagnosis of the primary disabling conditions at time of award is not a good proxy for the severity of the disabling condition, and it is the severity of the disabling condition at time t that affects the rate of returning to work at time t.

Figure 6 compares the first passage probabilities from state E_3 estimated from the proportional hazards model to the actual data for the whole population. The model does an excellent job of projecting the first passage probabilities of recovery past the 3½ year observation period. It does tend to underpredict the first passage probability to state E_2. The actual proportion of beneficiaries who die before making a work attempt or recovering within 7½ years of their time of award of benefits is 33 percent. The model predicts 29 percent.

The model does a very good job of predicting the proportion of beneficiaries who make a work attempt in t years. During the first 3½ years - the observation period - it does very well. In the subsequent years, the model slightly underpredicts the actual proportion. For example, the actual proportion of beneficiaries who make a work attempt within 7½ years of award of benefits is 23 percent. The model predicts 21 percent.

Figure 7 presents similar graphs for the first work episode. In this case, all the predictions are very good. Most work episodes end by returning to nonworking beneficiary status, as opposed to recovering or dying. The actual proportion that return to state E_3 within 7½ years is 72 percent. The model predicts 72.3 percent.

Simulation of Postentitlement Histories

In order to understand the behavior of beneficiaries under the different types of program, quantities other than first passage probabilities need to be predicted by the model. As was mentioned in the introduction, the model should have the ability to predict the expected length of time in the program, the expected cost in family benefits, Medicare costs, etc. One method for making such predictions is to use the first passage probabilities in a Monte Carlo simulation of the postentitlement history of a beneficiary over his time in the program. By simulating a large number of complete histories, direct calculations will produce estimates of the desired quantities. This method is appropriate since, closed form estimates for certain important quantities, when complex model structures are entertained, are simply not available.

In order to test the present model in this environment, the rate functions for the proportional hazards model, which were estimated with the first 3½ years of data, were used to simulate a history for each person for 7½ years, even if this was not the end of their time in the program. This was done since the actual data ends after 7½ years, and so all comparisions need to be limited to this period.

For each individual, under the assumption that the process is semi-Markov, one simulated history is created as follows: The actual date of award of benefits is used as entry time to state E_3. The monthly benefit amount is noted; the age at entry to E_3 is computed. A random draw is made from a uniform, $[0,1]$ distribution and compared to the first passage probabilities for an individual in that age group, monthly benefit group, primary diagnosis group, in state E_3. This decides when the person would switch to a new state and what the new state is. If the length of time for this even goes past 30 quarters, the episode is truncated at 30 quarters so as to compare with the original data. If not, and the person is still younger than 65, and the new state was E_4, a new random draw is made and compared to the first passage probabilities for an individual with the given monthly benefit and primary diagnosis, at the new age upon entry to state E_4. This continues until the person enters state E_1 or E_2 or reaches age 65 or 30 quarters are used up.

A number of quantities are computed for both the real and simulated histories. Some of them are presented in table 4. The average length of time in states E_3 and E_4 and the average length of time in the program compare quite favorably. For example, the average number of quarters in the disability program during the 30 quarter period of observation is 17.32 quarters. The model, which uses only the first 14 quarters for parameter estimation, predicts an average of 17.58 quarters. For those beneficiaries who were awarded a monthly benefit between $200 and $249, the average length of time as a working beneficiary during this 30 quarter observation period is .78 quarters. This small number is due in large part to the fact that many beneficiaries never work. The model predicts .72 quarters. The model does equally well across all age and monthly benefit groups. The greater discrepencies seem to occur when the sample size is small. In order to compare the sequencing of events between the actual and simulated histories, the mean number of work and nonwork episodes are tabulated by age and monthly benefit groups. In all categories, the model slightly underestimates the actual means. This is not surprising, in light of figure 6. The model underestimates the first passage probabilities to work. This will create less work episodes and subsequently create less returns to nonwork. It is also possible that the linkage assumptions in the AASM model are at fault.

Monthly Benefit Costs

To compute realistic cost estimates of any type, more information would be needed than is available from the CDHS. Monthly benefit costs should include the entire family benefit. Projections for the number of children receiving benefits and the amount of benefit would be needed. The wife's benefit would need to be estimated. Medicare costs would most likely be of interest. This information is not available from the present data. The only useful item is the monthly benefit payable to the primary beneficiary at the time of award of benefits. This is used to estimate the amount of monthly benefit, in 1969 dollars, paid to the primary beneficiary up until the end of the observation period. The variation in this monthly amount is due to the cost of living. 12/ Since we are computing cost in 1969 dollars, the amount remains constant during the person's time in the program. Therefore, we need only multiply the monthly benefit payable by the number of months in the program during the observation period. This is done for each beneficiary for both the actual history and the simulated history. Table 6 summarizes the results. In all categories the model estimates are quite good. For the total population, the cost per person for the last four years, a pure projection, the model predicts $5,097.80 compared to the actual cost of $5,011.00.

Conclusions

The main result of this work is that this methodology should be an effective tool for the analysis of the Work Incentive Experiments. The model is simple and straightforward. Even with no interaction terms, it does an excellent job of making predictions for a reasonable time beyond the observation period. This model was estimated with data spanning a 3½ year observation period, which is similar to the length of time expected for the Work Incentive Experiments. Even though only a couple of long term predictions were made with the model, others are possible. For example, the answers to all the questions described in the introduction are derivable. Examples of this type appear in the previous paper. 13/ This model also provides a method for analyzing short term effects. An analysis of the rate functions during the observation period can detect if any differences between groups are occuring.

Thus, this model provides a unified approach to the analysis of the data, has the advantage of dealing with the interdependencies among the various outcomes, and proves to be quite accurate in making long term predictions.

The work in this paper also suggests certain guidelines for a data gathering strategy. As was mentioned earlier, it seems that mixtures of rate functions provide more accurate predictions than one overall rate function. Thus, every effort should be made not only to capture the actual sequencing and timing of events, but also to gather the underlying details which would possibly cause individuals to have different rates of occurrence of these events. For example, whether the person is working above or below the SGA level may affect the rate of leaving a work episode. Since the diagnosis of the primary disabling condition did not seem to work as a proxy for severity of the disabling condition, perhaps some attempt at monitoring this stochastic variable over time should be made. All of this information should help to improve the accuracy of the model and provide good information about the different treatments in the Work Incentive Experiments.

TABLE 1.—Estimates of coefficients of rate functions, $r(t)=\exp(\alpha+\beta t+\gamma\log(t+1))$, for first nonwork episodes; which describe rates of recovery, return to work, and death; by age and benefit group; standard errors are in parentheses

Monthly Benefit group	Age group	Sample size	Recovery Rates α	Recovery Rates β	Recovery Rates γ	Death Rates α	Death Rates β	Death Rates γ	Work Rates α	Work Rates β	Work Rates γ
$0-149	**All....**	7,168	**-7.23** (.31)	**-.830** (.08)	**3.74** (.38)	**-3.23** (.03)	**-.048** (.00)	**0** ---	**-3.81** (.10)	**-.17** (.03)	**.44** (.13)
	18-30..	750	**-6.27** (.52)	**-.80** (.12)	**3.99** (.62)	**-4.58** (.13)	**0** ---	**0** ---	**-2.72** (.09)	**-.05** (.00)	**0** ---
	31-45..	1,079	**-7.05** (.61)	**-.94** (.14)	**4.46** (.74)	**-3.75** (.06)	**0** ---	**0** ---	**-3.65** (.22)	**-.26** (.05)	**.87** (.29)
	46-65..	5,339	**-7.44** (.49)	**-.75** (.13)	**3.01** (.65)	**-3.10** (.04)	**-.05** (.00)	**0** ---	**-3.85** (.06)	**-.10** (.00)	**0** ---
$150-249	**All....**	10,577	**-6.87** (.26)	**-.62** (.06)	**2.71** (.31)	**-3.82** (.00)	**0** ---	**0** ---	**-3.50** (.07)	**-.29** (.00)	**.67** (.10)
	18-30..	758	**-5.74** (.52)	**-.713** (.13)	**3.14** (.66)	**-5.11** (.16)	**0** ---	**0** ---	**-2.59** (.09)	**-.09** (.00)	**0** ---
	31-45..	1,761	**-6.66** (.48)	**-.68** (.10)	**3.26** (.57)	**-4.26** (.06)	**0** ---	**0** ---	**-3.39** (.16)	**-.29** (.03)	**.84** (.21)
	46-65..	8,058	**-7.13** (.37)	**-.52** (.09)	**2.14** (.47)	**-3.70** (.00)	**0** ---	**0** ---	**-3.63** (.09)	**-.33** (.03)	**.72** (.13)
$250-	**All....**	3,004	**-11.05** (1.23)	**-1.48** (.27)	**7.30** (1.44)	**-7.39** (.22)	**0** ---	**0** ---	**-3.15** (.12)	**-.28** (.03)	**.51** (.17)
	18-30..	211	**-5.16** (.27)	**0** ---	**0** ---	**-6.95** (.18)	**0** ---	**0** ---	**-3.14** (.13)	**0** ---	**0** ---
	31-45..	591	**-6.06** (.26)	**0** ---	**0** ---	**-6.98** (.07)	**0** ---	**0** ---	**-2.82** (.12)	-.15 (.00)	**0** ---
	46-65..	2,202	**-12.36** (1.89)	**-1.91** (.45)	**9.13** (2.31)	**-7.58** (.05)	**0** ---	**0** ---	**-3.30** (.14)	**-.40** (.04)	**.85** (.22)

TABLE 2.--Estimates of coefficients of rate functions, $r(t)=\exp(\alpha+\beta t+\gamma \log (t+1))$, for first nonwork episode, which describe rates of recovery, return to work, and death; for those beneficiaries with a monthly benefit of $150 through $249, whose age is 46 through 65, by primary diagnosis at date of award; standard errors in parentheses

Primary Diagnosis	Sample size	Recovery			Death			Work		
		α	β	γ	α	β	γ	α	β	γ
Total............	8,058	**-7.13** (.37)	**-.52** (.09)	**2.14** (.47)	**-3.70** (.00)	**0** ---	**0** ---	**-3.63** (.09)	**-.33** (.03)	**.72** (.13)
Circulatory......	3,214	**-7.80** (.72)	**-.74** (.19)	**3.08** (.93)	**-4.50** (.17)	**-.07** (.06)	**.68** (.19)	**-3.49** (.13)	**-.31** (.04)	**.61** (.19)
Musculoskeletal..	1,067	**-7.77** (1.10)	**-.55** (.21)	**2.79** (1.21)	**-4.88** (.10)	**0** ---	**0** ---	**-3.49** (.12)	**-.15** (.00)	**0** ---
Mentally Ill	554	**-6.70** (.35)	**0** ---	**0** ---	**-4.34** (.11)	**0** ---	**0** ---	**-3.64** (.18)	**-.14** (.03)	**0** ---
Accident.........	411	**-5.63** (.85)	**-.34** (.19)	**1.42** (.69)	**-5.77** (.40)	**.139** (.04)	**0** ---	**-3.82** (.37)	**-.47** (.10)	**1.54** (.51)
Neoplasm.........	707	**-6.25** (.38)	**0** ---	**0** ---	**-2.03** (.13)	**-.23** (.04)	**.87** (.20)	**-3.81** (.33)	**-.64** (.16)	**1.74** (.62)
Respiratory......	847	**-8.45** (.71)	**0** ---	**0** ---	**-3.73** (.07)	**0** ---	**0** ---	**-3.43** (.15)	**-.196** (.03)	**0** ---
Other............	1,258	**-7.12** (.90)	**-.62** (.23)	**2.51** (1.15)	**-4.30** (.18)	**0** ---	**.25** (.09)	**-3.78** (.23)	**-.474** (.08)	**1.22** (.35)

TABLE 3.--Estimates of coefficients for proportional hazards rate functions in formula (13); standard errors are in parentheses

Variable	Nonwork Episodes (N=19,714) Recovered	Nonwork Episodes (N=19,714) Deceased	Nonwork Episodes (N=19,714) Working Beneficiary	Work Episodes (N=3,166) Recovered	Work Episodes (N=3,166) Deceased	Work Episodes (N=3,166) Non-working Beneficiary
Constant	-8.58	-11.29	-3.40	-4.78	-17.02	-2.79
	(.31)	(3.76)	(.16)	(.06)	(.67)	(.20)
Age: 18-30	2.37	- .70	1.07	.41	- .70	- .38
	(.13)	(.10)	(.25)	(.02)	(.89)	(.46)
31-45	1.99	- .18	.65	.54	- .38	- .26
	(.12)	(.05)	(.21)	(.01)	(.36)	(.22)
46-55	1.34	.09	.25	.20	- .08	- .07
	(.13)	(.04)	(.27)	(.01)	(.26)	(.16)
Monthly Benefit:						
$0-100	.17	7.48	- .48	.16	11.22	.37
	(.22)	(3.76)	(.19)	(.04)	(.85)	(.16)
100-150	.11	7.52	- .62	.05	11.60	.29
	(.21)	(3.76)	(.12)	(.03)	(.65)	(.18)
150-200	- .01	7.42	- .42	.19	12.05	.38
	(.21)	(3.76)	(.11)	(.03)	(.62)	(.15)
200-250	- .72	6.34	- .48	.07	10.62	.51
	(.23)	(3.76)	(.14)	(.04)	(.83)	(.15)
250-300	-1.34	3.72	- .39	- .78	- .51	.76
	(.26)	(3.75)	(.17)	(.05)	(1.00)	(.14)
Diagnosis:						
Circulatory	- .63	.23	- .01	- .58	.56	.07
	(.13)	(.05)	(.17)	(.02)	(.35)	(.15)
Musculoskeletal	.45	- .98	.06	.08	- 1.56	.03
	(.11)	(.08)	(.35)	(.02)	(.94)	(.24)
Mentally Ill	-1.00	- .74	- .23	- .37	- .56	.15
	(.15)	(.08)	(.31)	(.02)	(.78)	(.37)
Accident	.91	- .86	.37	.37	-10.50	- .19
	(.10)	(.11)	(.37)	(.01)	(1.03)	(.44)
Neoplasm	- .33	1.88	.01	1.53	2.10	- .01
	(.24)	(.05)	(.54)	(.06)	(.36)	(.53)
Respiratory	-2.21	.14	- .32	-1.25	.68	. 44
	(.45)	(.06)	(.21)	(.18)	(.86)	(.18)
T	- .75	- .05	- .25	- .55	- .72	-1.02
	(.05)	(.01)	(.05)	(.00)	(.26)	(.07)
LOG (T+1)	3.55	.28	.57	3.34	2.72	3.43
	(.24)	(.07)	(.25)	(.05)	(.83)	(.21)

TABLE 4.--Mean number of quarters spent as a nonworking beneficiary, working beneficiary, and mean number of quarters as a beneficiary, during the 30 quarter period, by monthly benefit group, by age group; actual and proportional hazards model estimate

Monthly benefit group	Age group	Number of cases	Meantime as nonworking beneficiary		Meantime as working beneficiary		Meantime as beneficiary	
			Actual	Model estimate	Actual	Model estimate	Actual	Model estimate
$0-99	18-30	266	14.34	15.02	2.39	1.31	16.73	16.33
	31-45	284	16.48	16.20	1.22	1.00	17.70	17.20
	46-65	1,311	14.52	14.75	.65	.41	15.16	15.16
	Total	1,861	14.79	14.75	.98	.63	15.77	15.64
$100-149	18-30	483	13.14	14.65	1.71	1.39	14.85	16.04
	31-45	795	15.23	16.29	1.07	.72	16.30	17.01
	46-65	4,029	14.19	14.52	.48	.40	14.67	14.92
	Total	5,307	14.25	14.80	.68	.54	14.93	15.33
$150-199	18-30	472	15.53	14.85	1.90	1.44	17.44	16.29
	31-45	1,224	15.52	16.06	.92	.99	16.44	17.05
	46-65	5,697	14.36	14.61	.53	.47	14.89	15.08
	Total	7,393	14.63	14.87	.68	.62	15.31	15.48
$200-249	18-30	285	16.35	17.61	1.57	1.88	17.93	19.49
	31-45	536	20.08	19.65	1.27	1.02	21.35	20.68
	46-65	2,363	19.74	20.33	.58	.51	20.32	20.84
	Total	3,184	19.49	19.97	.78	.72	20.28	20.69
$250-299	18-30	122	19.91	20.26	1.99	2.46	21.90	22.72
	31-45	518	23.35	23.21	1.08	1.29	24.43	24.51
	46-65	2,132	24.23	24.63	.74	.67	24.97	25.29
	Total	2,772	23.88	24.17	.86	.86	24.74	25.03
$300 or more	18-30	88	19.14	15.32	2.30	2.61	21.43	17.93
	31-45	73	18.18	17.66	2.52	1.86	20.70	19.52
	46-65	71	11.55	21.59	2.56	1.32	14.11	22.92
	Total	232	16.51	17.97	2.45	1.98	18.96	19.96
TOTAL		20,749	16.55	16.92	.77	.66	17.32	17.58

TABLE 5.--Mean number of nonwork and work; episodes during the 30 quarter period; by monthly benefit group, by age group; actual and proportional hazards model estimate

Monthly benefit group	Age group	Number of cases	Mean number of nonwork episodes: Actual	Mean number of nonwork episodes: Model estimate	Mean number of work episodes: Actual	Mean number of work episodes: Model estimate
$0-99	18-30	266	1.35	1.21	.85	.41
	31-45	284	1.36	1.19	.46	.34
	46-65	1,311	1.14	1.10	.19	.15
	Total	1,861	1.22	1.13	.33	.22
$100-149	18-30	483	1.40	1.24	.65	.44
	31-45	795	1.27	1.17	.40	.27
	46-65	4,029	1.12	1.09	.17	.13
	Total	5,307	1.17	1.12	.25	.18
$150-199	18-30	472	1.49	1.29	.75	.51
	31-45	1,224	1.24	1.22	.40	.36
	46-65	5,697	1.15	1.12	.21	.18
	Total	7,393	1.19	1.15	.27	.23
$200-249	18-30	285	1.41	1.36	.67	.61
	31-45	536	1.33	1.24	.50	.36
	46-65	2,363	1.18	1.14	.24	.18
	Total	3,184	1.23	1.18	.32	.25
$250-299	18-30	122	1.53	1.57	.77	.75
	31-45	518	1.38	1.38	.47	.44
	46-65	2,132	1.27	1.25	.32	.27
	Total	2,772	1.30	1.29	.37	.33
$300 or more	18-30	88	1.59	1.43	.86	.81
	31-45	73	1.51	1.21	.77	.49
	46-65	71	1.34	1.23	.66	.34
	Total	232	1.49	1.30	.77	.56
TOTAL		20,749	1.21	1.16	.30	.23

TABLE 6.--Average monthly benefit cost, by monthly benefit group, by age group, for first 3½ years, for last 4 years, for total 7½ years; actual and proportional hazards model estimate

Monthly benefit group	Age group	Number of cases	First 3½ years		Last 4 years		All 7½ years	
			Actual	Model estimate	Actual	Model estimate	Actual	Model estimate
$0-99	18-30	266	$1,826.80	$1,878.90	$2,111.40	$1,966.50	$3,938.20	$3,845.30
	31-45	284	1,836.20	1,903.10	2,266.80	2,051.30	4,103.00	3,954.40
	46-65	1,311	1,996.10	2,059.70	1,650.30	1,576.00	3,646.30	3,635.70
	Total	1,861	1,947.50	2,010.00	1,810.00	1,704.20	3,757.50	3,714.20
$100-149	18-30	483	2,673.50	2,898.10	2,838.80	3,068.70	5,512.30	5,966.70
	31-45	795	2,883.80	3,110.60	3,321.70	3,383.30	6,205.60	6,493.90
	46-65	4,029	3,115.00	3,218.40	2,501.70	2,465.50	5,616.70	5,683.90
	Total	5,307	3,040.20	3,173.10	2,655.20	2,657.90	5,695.40	5,831.00
$150-199	18-30	472	4,150.80	4,099.30	4,926.50	4,339.30	9,077.40	8,438.60
	31-45	1,224	3,873.30	4,173.00	4,547.90	4,588.00	8,421.30	8,761.00
	46-65	5,697	4,180.80	4,269.70	3,482.20	3,479.30	7,663.10	7,749.00
	Total	7,393	4,128.00	4,242.80	3,750.90	3,717.80	7,878.90	7,960.60
$200-249	18-30	285	5,459.90	5,904.90	6,469.90	7,079.80	11,929.80	12,984.70
	31-45	536	6,214.10	5,846.90	8,102.70	7,974.80	14,316.80	13,794.70
	46-65	2,363	6,609.10	6,347.00	7,107.00	7,697.40	13,716.10	14,044.40
	Total	3,184	6,439.70	6,223.20	7,217.60	7,684.20	13,657.30	13,907.50
$250-299	18-30	122	7,745.70	7,781.10	10,135.70	10,723.40	17,881.40	18,504.40
	31-45	518	8,125.70	8,090.50	11,886.20	11,977.40	20,011.90	20,067.90
	46-65	2,132	8,345.20	8,346.20	12,142.20	12,382.10	20,487.40	20,728.30
	Total	2,772	8,277.80	8,273.50	12,006.00	12,233.50	20,283.80	20,507.00
$300 or more	18-30	88	9,457.30	8,888.40	11,762.60	9,103.20	21,219.90	17,991.60
	31-45	73	8,217.90	8,668.30	12,014.90	10,979.70	20,232.80	19,648.00
	46-65	71	6,211.80	8,602.30	8,029.10	13,751.20	14,240.80	22,353.50
	Total	232	8,051.70	8,729.40	10,676.70	11,148.30	18,728.40	19,877.70
TOTAL		20,749	4,608.70	4,663.10	5,011.00	5,097.80	9,619.70	9,760.90

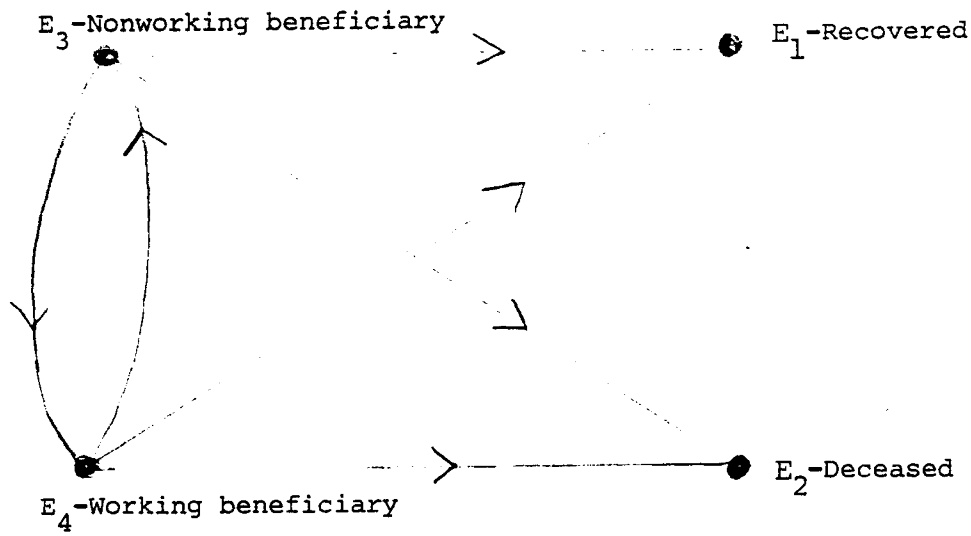

FIGURE 1.—Flow Between States During Initial Episode in Disability Program

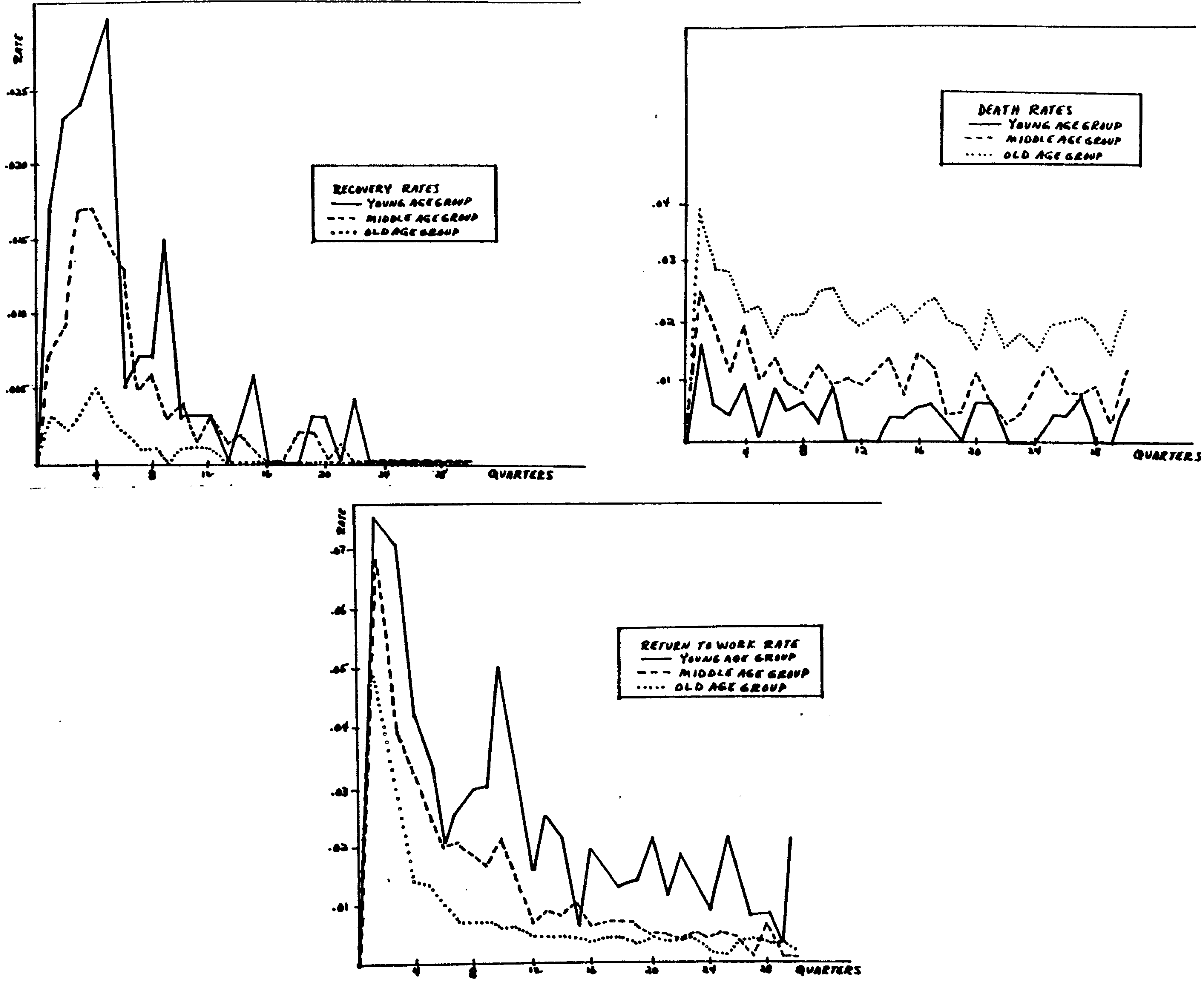

FIGURE 2.--**Rate functions for switches from nonwork to other states for middle monthly benefit group** (\$150-249), for young age group (18-30), middle age group (31-45), and old age group (46-65).

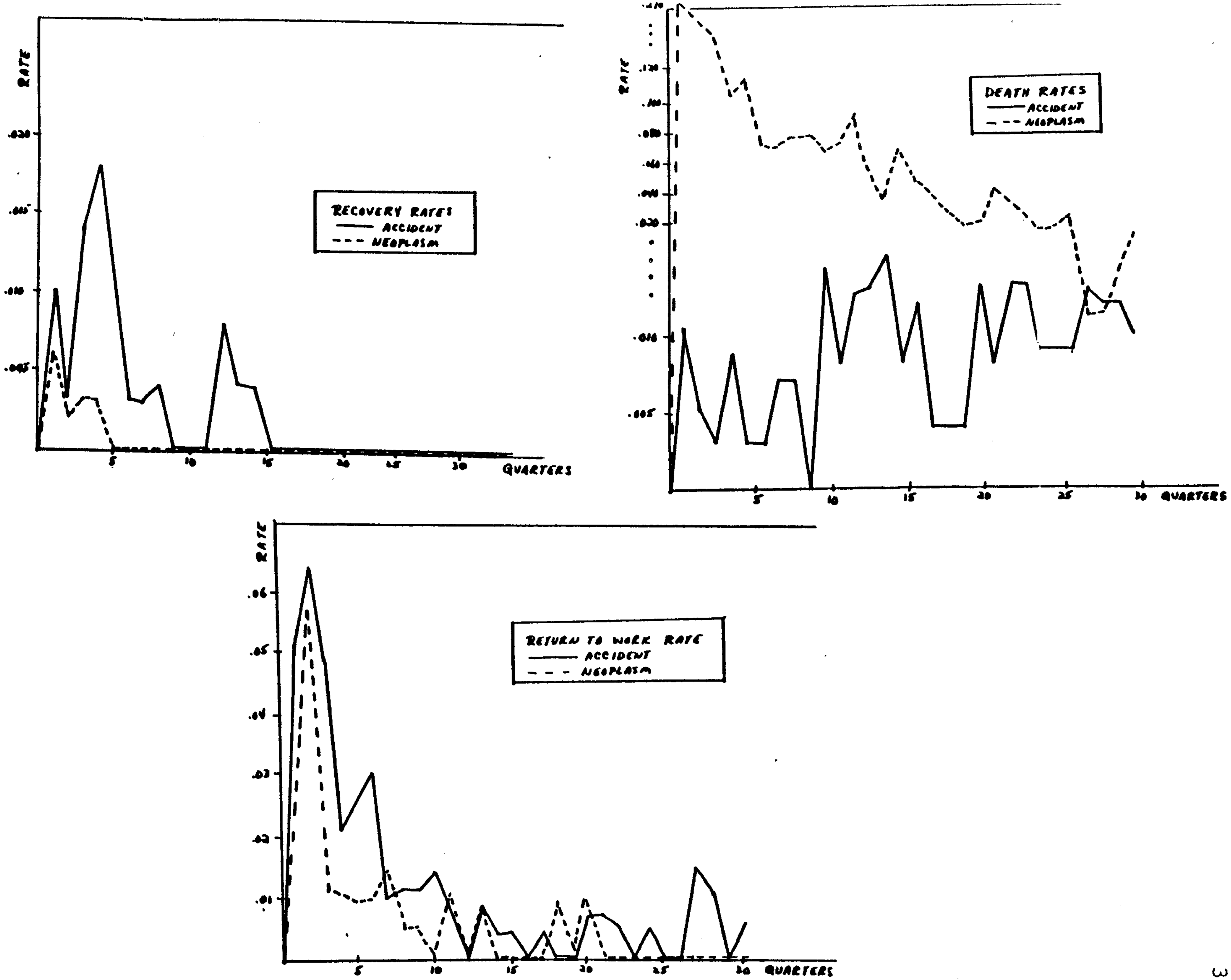

FIGURE 3.--Rate functions for switches from nonwork to other states for middle benefit group ($150-249), old age group (46-65), by diagnosis.

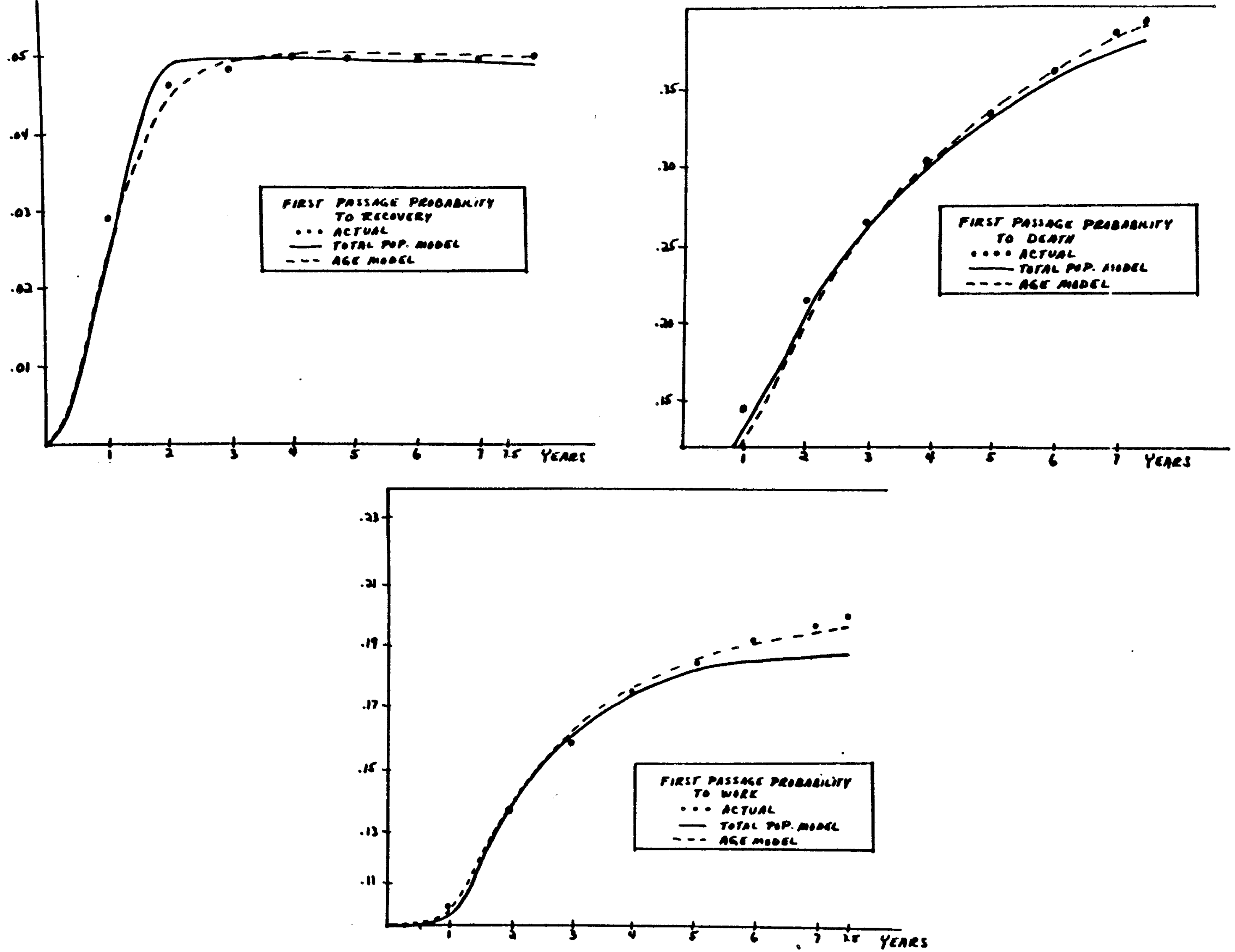

FIGURE 4A.--First passage probabilities from nonwork to other states; estimates from the total population and age models; low benefit group, ($0-149).

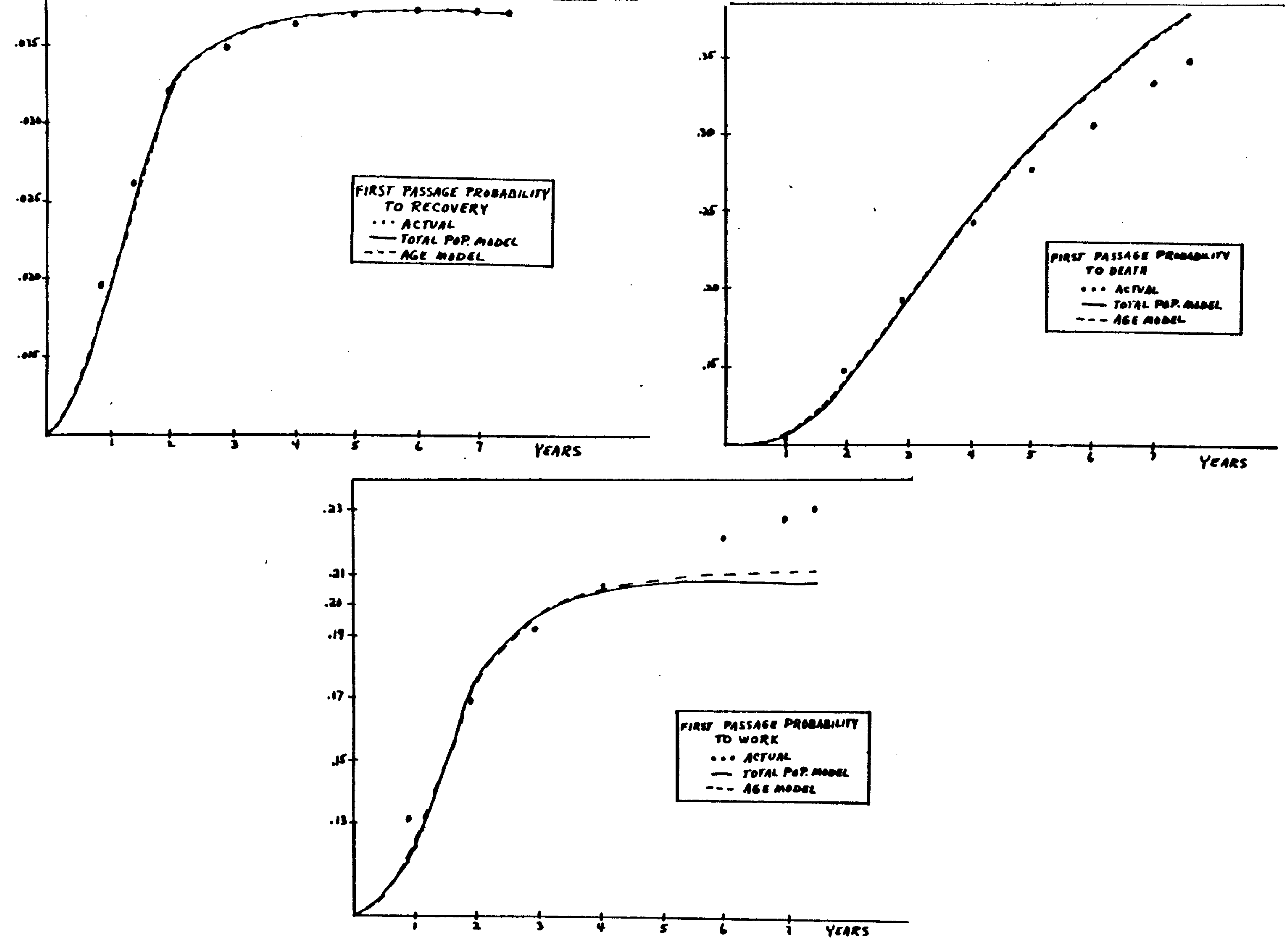

FIGURE 4.B.--First passage probabilities from nonwork to other states; estimates from the total population and age models; middle benefit group, ($150-249).

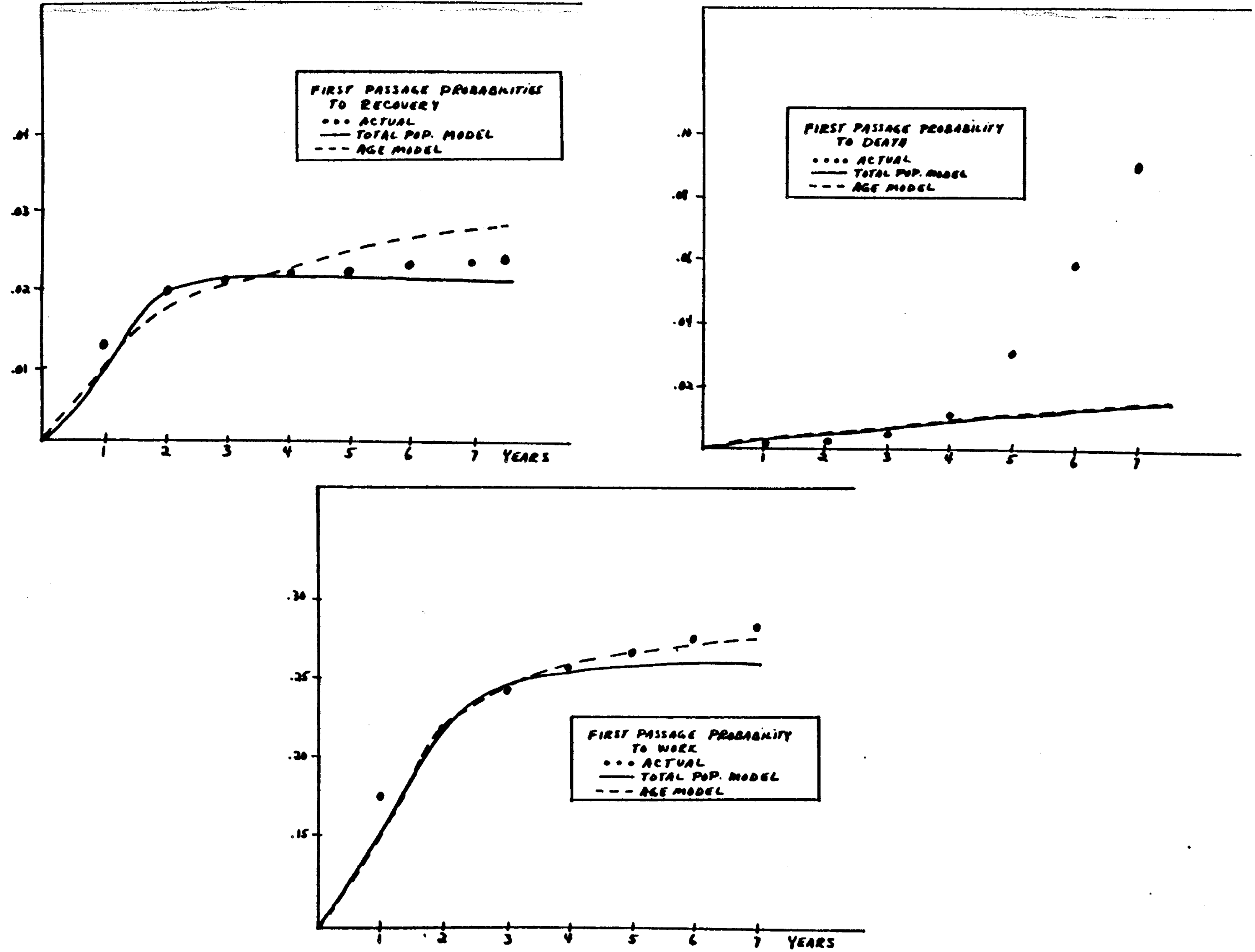

FIGURE 4C.--**First passage probabilities from nonwork to other states; estimates from the total population and age models;** high benefit group, ($250 or more).

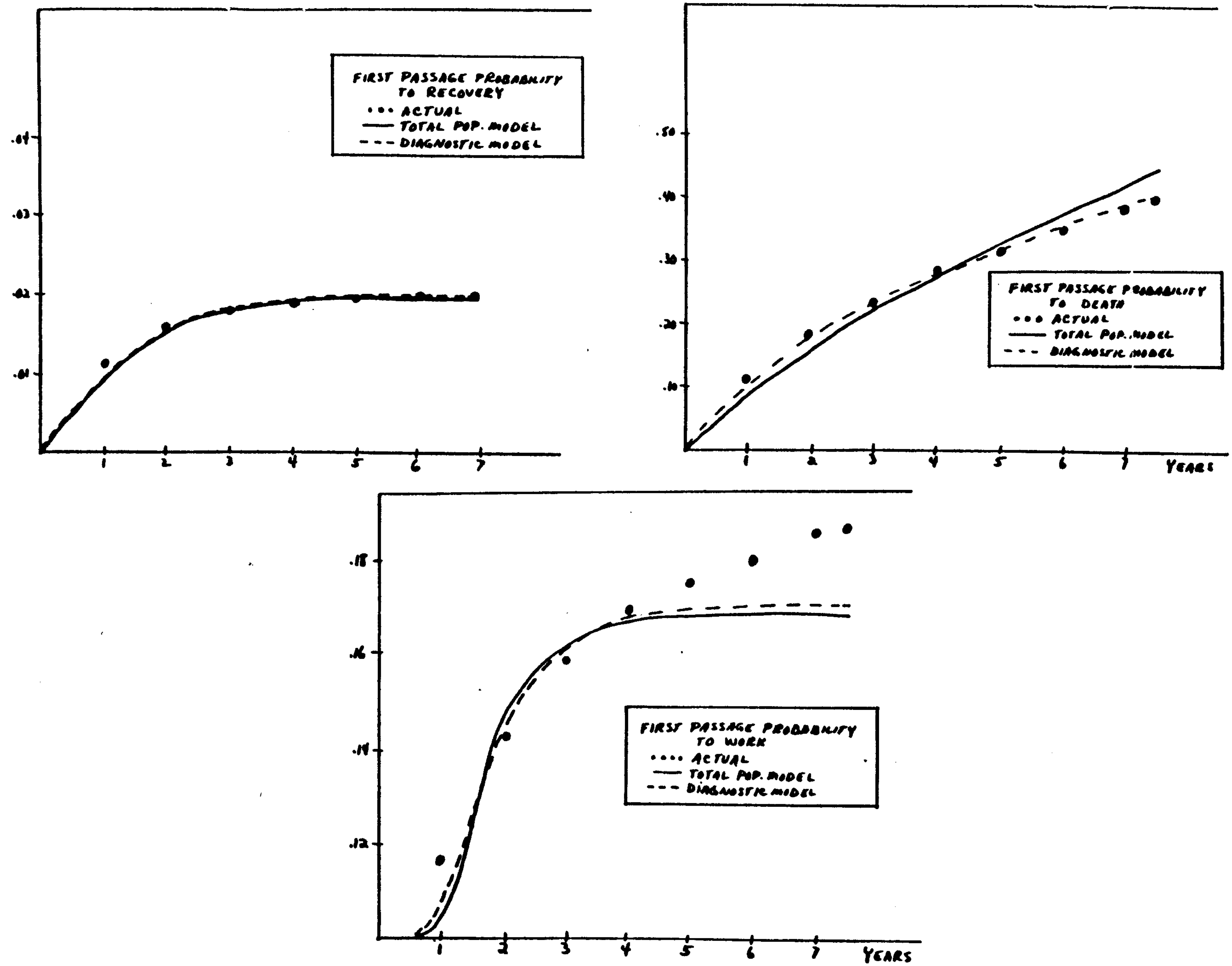

FIGURE 5 .--First passage probabilities from nonwork to other states; estimates from total population model and diagnostic model for the middle benefit, ($150-249), old age group, (46-65).

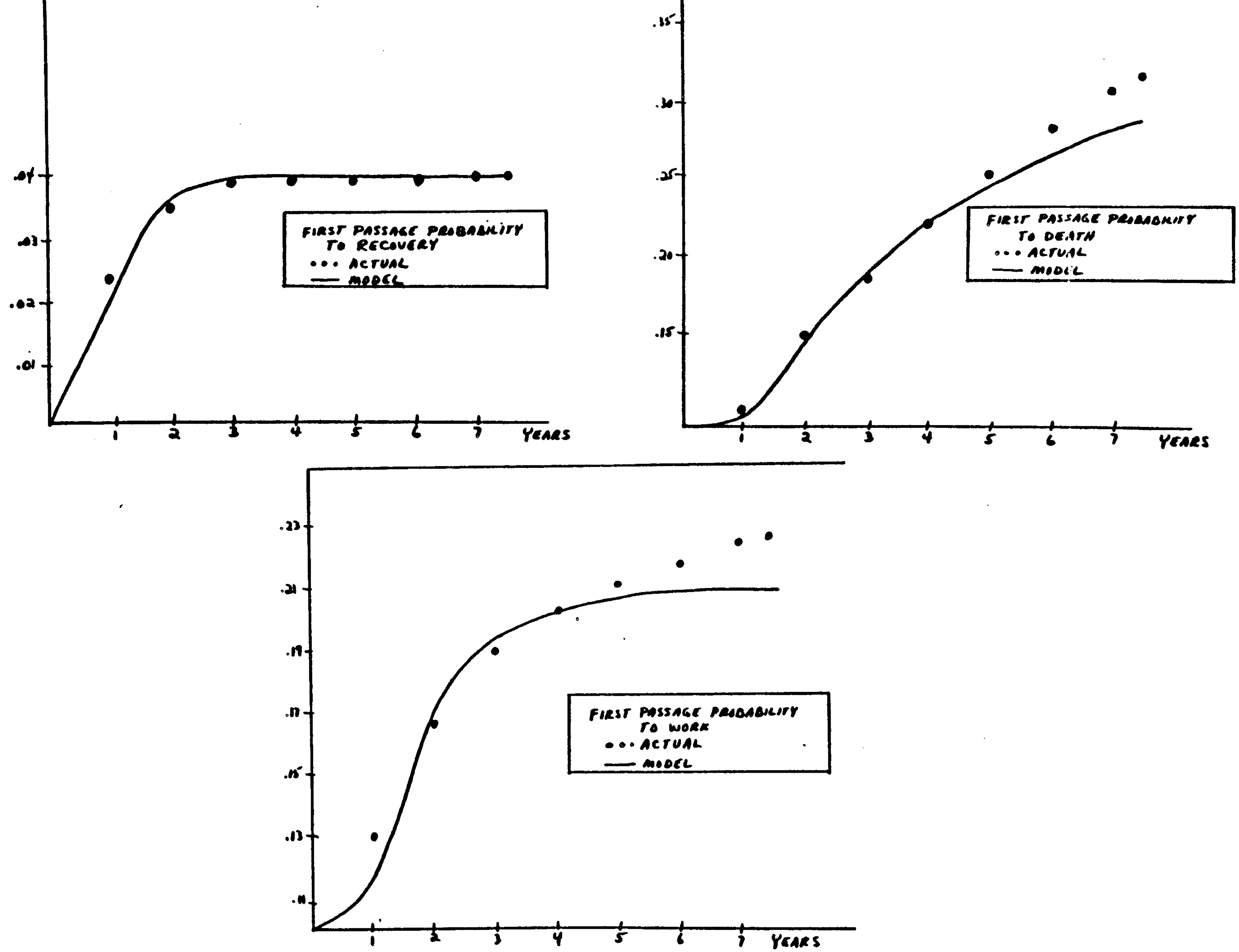

FIGURE 6.--First passage probabilities from nonwork to other states, estimate from the proportional hazards model.

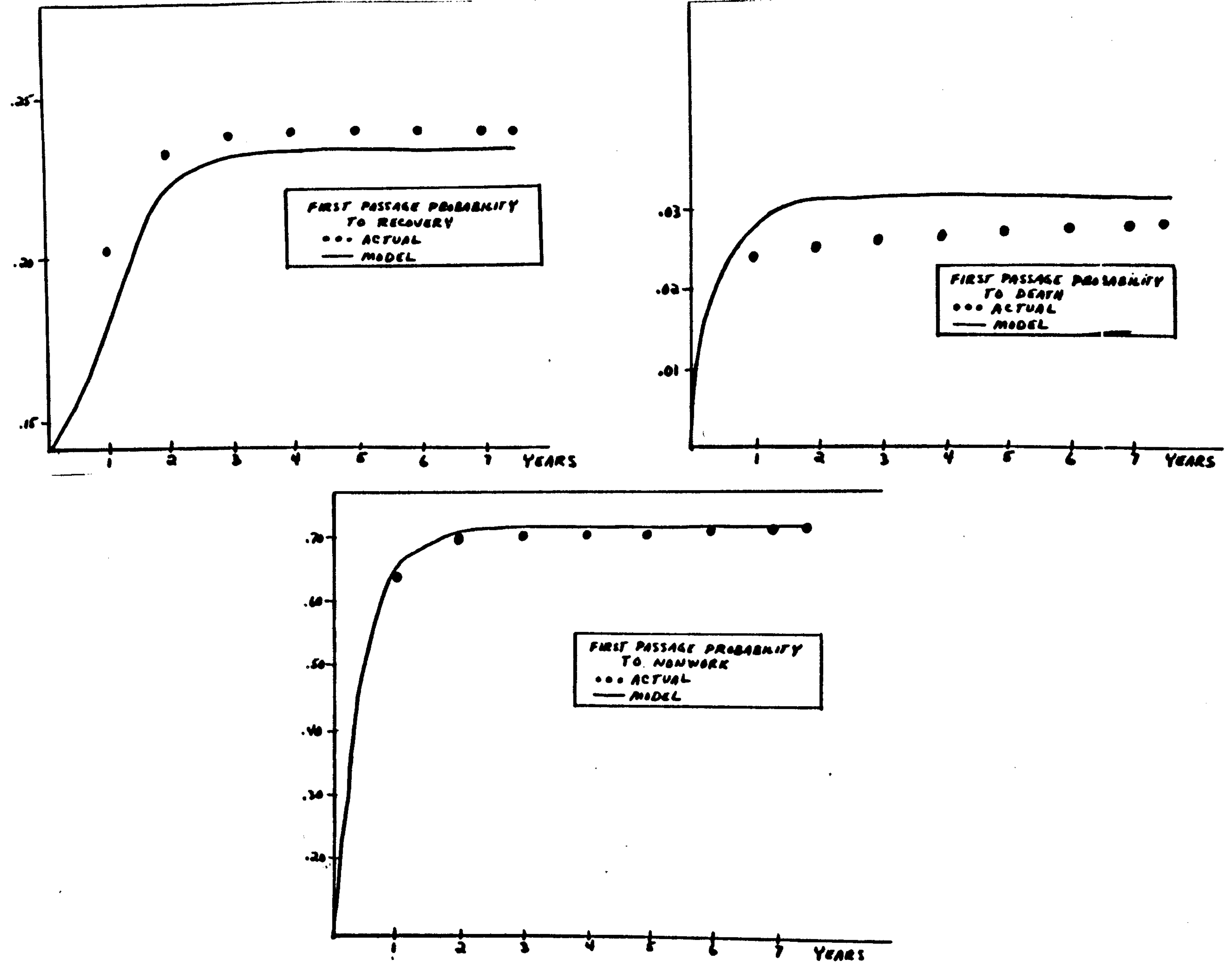

FIGURE 7.--First passage probabilities from work to other states, estimate from the proportional hazards model.

Footnotes

1/ For details, see "Disability Insurance Work Incentive Experiments: Project Statement", SSA/OP/ORS/DDS, March 1982

2/ See Charles Mode, "Algorithms for a Stochastic Model of Human Fertility under Contraception", Mathematical Biosciences, 38:217-246 (1978) and the references.

3/ See Nancy Tuma, Michael Hannan, and Lyle Groeneveld, "Dynamic Analysis of Event Histories", American Journal of Sociology, 84:820-854 (1979) and the references.

4/ Christopher Flinn and James Heckman, "Models for the Analysis of Labor Force Dynamics", unpublished manuscript, 1980, University of Chicago.

5/ John Hennessey, "An Age Dependent Absorbing Semi-Markov Model of Postentitlement Work Histories of the Disabled Beneficiaries", Social Security Administration, Office of Research and Statistics, Staff Paper No. 38, 1980.

6/ For more details, see Continuous Disability History Sample, Restricted Use Data File: Description and Documentation, Social Security Administration Office of Research and Statistics, Pub. No. 024 (1-78).

7/ This code is based on the International Classification of Diseases, Adapted, Eighth Revision.

8/ In James Heckman and Burton Singer, "The Identification Problem in Econometric Models for Duration Data" , in Advances in Econometrics, edited by W. Hildebrand, Cambridge University Press, 1982, this issue is discussed in more detail.

9/ See footnotes 3 and 8.

10/ This technique is described in G. R. Walsh, Methods of Optimization, J. Wiley, 1975.

11/ This model is used quite often in survival analysis where rate functions are called hazard functions.

12/ Actual benefits paid could be reduced if the person is receiving money from certain other federal programs. This information is not available and will not be taken into account.

13/ c. f. footnote 5