Hazard Rates from Competing/Repeated Risk Survival Analysis

Carl Formoso, Division of Child Support, Olympia, WA

Introduction

In survival analysis we look at duration of residence of individuals in particular states, and at rates of change from one state to another. Often we are interested in the effects of a particular set of covariates on these factors, while controlling for other covariates.

When more than one type of change is possible, individuals can exposed to multiple competing risks. In competing risks the occurrence of one event eliminates the risk of other events. A common example is that once an individual has died from a heart attack, there is no longer a risk of traffic-related death. It is usually easy to deal with competing risks in survival analysis, though the treatment generally will increase the uncertainty of the results.

The origins of survival analysis focused on death, a one-directional change. But in social programs, and other areas, we are interested in bi-directional changes: an individual can leave welfare and later return to welfare, or leave prison and later return to prison, etc. Thus each individual has the possibility of being exposed to a particular risk more than once, and of experiencing a particular change, or event, multiple times. Repeated events can be difficult to analyze, with possible short-comings whichever approach is taken.

Hazard rates, the number of events expected per individual per unit time, may be the most generally useful way to compare different survival functions. By determining the hazard at a set point in time, we have a single number which characterizes the survival function. The inverse of the hazard rate gives us the expected duration in the state under consideration, assuming the event under consideration is the only risk. For competing exit risks in the state under consideration, the hazard rates are directly additive.

The SAS® survival analysis procedure PHREG is generally useful because it does not require an assumption of the probability distribution of event times. The output from PHREG lists the "Risk Ratio" for each covariate. For a dichotomous variable this is the ratio of the hazard for "1" to the hazard for "0," while controlling for all other covariates. While the hazard function may be time dependent, the Risk Ratio from PHREG is not time dependent, because PHREG assumes proportional hazards. PHREG will also output survival curves for given sets of covariates. From the survival curve the magnitude of the hazard rate can be estimated. This combined with the Risk Ratio allows a complete set of comparable hazard rates to be determined.

Survival Analysis

Survival analysis is based on some fairly simple calculus. For individuals exiting a particular state,

$$-\frac{dN}{dt} = hN \qquad ,$$

where N is the number of individuals in the state at time t, h is the hazard rate , and dN/dt is the rate of change in N at time t. Assuming that h does not depend on t and integrating gives

$$S = \frac{N}{N_0} = e^{-ht} \qquad ,$$

an exponential survival function, S. The hazard, h, may be a function of explanatory variables:

$$h \;= k\,\exp\left\{ \underset{i}{\ } \beta_i x_i \right\} \qquad ,$$

where $\boldsymbol{\beta}_i \boldsymbol{x}_i$ are the relevant coefficients and covariates. Additionally, h may be time dependent, for example:

$$h = \kappa e^{pt} \qquad ,$$

then integration gives

$$\ln S = \ln \frac{N}{N_0} = \frac{\kappa}{p}(1 - e^{pt}) = a(1 - e^{bt})$$

in this case a Gompertz survival function.

The survival function gives the probability of survival until time t or longer. A related function, the probability density function ($f = h*S$), gives the probability of an event occurring in a small time increment at t. With observed events and right censored events both of these probabilities are used by PHREG in fitting observed survival times for a partial likelihood estimation of the $\boldsymbol{\beta}_i$ (Allison, 1995, p81-84 & chap 5).

Welfare and Work Four-State Model

Using a cohort of 116,377 adults who were on welfare in 4th Quarter 1993 (93Q4), welfare use history and work history were acquired from 91Q4 to 97Q1. Monthly state residence from 94Q1 to 97Q1 was established according to the four-state model indicated in Figure 1 (Formoso, 1999). Work status was only available on a quarterly basis, so work status is always constant for the three months of a quarter. History prior to 93Q4 is used in two explanatory variables. Other explanatory variables are based on information from, or prior to, 93Q4. Table 1 identifies the covariates.

Figure 1 also introduces a numerical notation for each state for ease of reference, and illustrates three competing risks for exit from state 2. Thus there are twelve unique events possible in the model, and there will be twelve unique hazards. Isolating the unique hazards can be done within PHREG simply by labeling the competing events as censored observations (Allison, 1995, Chap 6). For example, if we want to focus on the 2 $\to$ 1 event, the 2 $\to$ 0 and 2 $\to$ 3 events are treated as censored observations. However, as the extent of censoring increases the coefficient estimates are based on diminishing information on the event of interest, and bias may be introduced by the possibility of "informative" censoring (Allison, 1995, p9-14).

We treat repeated events in the most direct way, by creating a separate observation for each occurrence. While this can possibly attenuate coefficient estimates and amplify test statistics, a correction for statistics (Allison, 1997, p78-79) indicates statistical validity is barely affected in our work. Corrections for coefficients are more difficult and were not attempted, but are likely to also be small.

Survival for Competing/Repeated Risks

We are interested in the welfare/work patterns associated with different levels of child support enforcement (CSE) collections, controlling for other factors, including other social programs accessed by clients. To illustrate the results, we show comparisons of three program categories:

CPJN signifies poor CSE collections and clients who had not entered the JOBS program,

CGJN signifies good CSE collections and clients who were not JOBS entrants,

CPJY signifies poor CSE collections and clients who were JOBS entrants.

The differences between CGJN and CPJN then show the effects associated with increasing the level of CSE collections, without the influence of JOBS. The differences between CPJY and CPJN show the effects associated with JOBS, without the influence of the level of CSE collections (Formoso, 1999).

Figure 2 shows survival curves for average welfare clients for the 1 $\to$ 2 event, where the effect associated with good child support collections is a lengthening of the time off welfare. The effect associated with JOBS is small and just barely outside the 95% confidence limit (approximately the size of the markers). A similar result is seen with the 0 $\to$ 3 event. These two events are the main pathways of welfare recidivism.

Figure 3 shows survival curves for the 3 $\to$ 2 event where the effect associated with the JOBS program is a decrease in the time on welfare without work. There is no effect associated with the level of child support collections. The step nature of the curves is due to the quarterly basis of work status.

Ignoring diagonal transitions ($0 \leftrightarrow 2$ and $1 \leftrightarrow 3$) which are very slow processes, these are the only effects clearly associated with the program indicators.

Hazards for Competing/Repeated Risks

The hazard rate is the negative of the derivative of ln*S* - the slope of a plot of ln*S* vs *t*. None of the twelve unique transitions in this system gave a linear plot; in all cases the hazard appears to be time-dependent. Most plots of ln*S* vs *t* showed a slope smoothly decreasing with time, a situation that can be approximated by the Gompertz hazard given above. The SAS® NLIN procedure for a non-linear least squares regression was used to obtain the Gompertz parameters. In most cases the log survival curves could be fit very well with a sum of two Gompertz functions; in a few cases a sum of three Gompertz functions were needed,

$$\ln S = \sum_{i=1}^{3} a_i(1-e^{b_i t}) \quad .$$

However, our objective is to use the parameters of the Gompertz function to calculate the hazard rate at a particular point in time, and for this only one Gompertz function is justified. As the number of parameters in the fit increases, the parameter errors also increase, leading to a larger error in the hazard rate. In no case was a multiple Gompertz fit hazard rate significantly different than the hazard rate calculated with a single Gompertz.

Using this approach we are able to determine magnitudes of the hazard rates for all transitions and for all program categories. The ratios of the hazards for program categories obtained in this way are very close to those obtained directly from PHREG output. But it is better, in terms of confidence limits, to use the Gompertz fit procedure described above to obtain the magnitude of the reference hazard, and then use the risk ratios output from PHREG to estimate the full set of hazard rates. By determining the hazard at a standard point in time (we have chosen 9 months) we can compare rates across all transitions and across studies. For example, we also are studying other welfare cohorts with shorter lengths of follow-up (Formoso, 1999 & in preparation).

By their magnitudes and error limits the hazards at 9 months fall into four groups, as indicated in Table 2. The Risk Ratios from PHREG output are shown in Table 3. The error in the Risk Ratio is generally lower than 5%, averaging about 3%.The marked Risk Ratios in Table 2 are those we consider strongly significant. A Risk Ratio outside the interval ~0.8 - ~1.25 is above the 95% confidence limit for a real difference in hazard values, assuming a 5% error. Because the diagonal transitions are slow, there is extensive censoring (96 - 98% of diagonal events are censored); lower confidence is necessary for the diagonal events (marked with an open symbol).

Table 4 then shows the strongest and most certain effects associated with the program indicators. An extension of the time off welfare is the most certain effect strongly associated with increasing the level of CSE collections. Work status has little or no effect on the impact associated with the level of CSE collections. The most certain effect strongly associated with the JOBS program is a decrease in the time required to find work while on welfare.

Some Uses of Empirical Hazard Rates

The hazard rates shown in Table 2 offer strong support for the work emphasis of welfare reform. The hazard rates for welfare exit from State 2, 'welfare and work,' are about four times larger than hazard rates for welfare exit from State 3. Hazard rates for welfare exits to employment are about one hundred times faster from State 2, compared with State 3.

In including the JOBS program there is always the question of selection bias. The hazard results indicate that JOBS entrants are not markedly different from other welfare clients in their hazards for exiting welfare. The main effect associated with the JOBS program is an increased hazard for employment while on welfare. This, in fact, does lead to an increased number of welfare exits, because a higher proportion of JOBS entrants reside in State 2 while on welfare.

Once the hazard rates are known they can be used to simulate outcomes in the welfare/work system. Figure 4 shows such a simulation, with observed values and projection forward. With this simulation one can begin to ask about cost outcomes of effecting changes in rates for each separate event in the model. Looking at incremental 10% rate changes, at this point very rough results suggest that the largest marginal cost savings is effected by a 10% reduction in the 0 $\rightarrow$ 3 transition rate.

References

Allison, Paul D., *Survival Analysis Using the SAS® System: A Practical Guide*, Cary, NC: SAS Institute Inc.,1995.

Allison, Paul D., *Event History & Survival Analysis*, Course Notes,1997.

Formoso, Carl, *The Effect of Child Support and Self-Sufficiency Programs on Reducing Direct Support Public Costs*, submitted to the Lewin Group under Subcontract #TLG98-003-1898.02, 1999.

Formoso, Carl, *Cost Avoidance of Child Support Collections Under the TANF Program*, in preparation.

Author

Carl Formoso, Division of Child Support, PO Box 9162, Olympia, WA 98507, (360)664-5090, FAX (360)586-3274, cformoso@dshs.wa.gov.

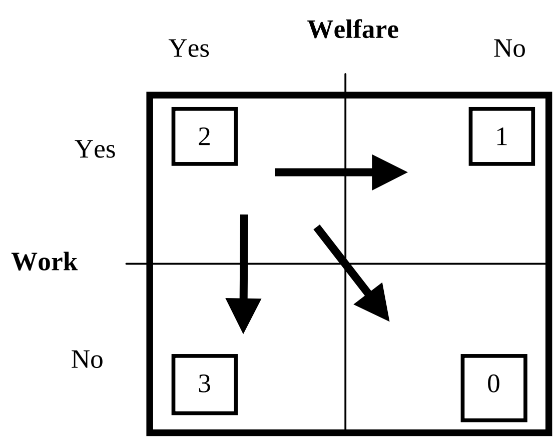

Figure 1
Four-State Welfare/Work Model
Illustrating Competing Risks

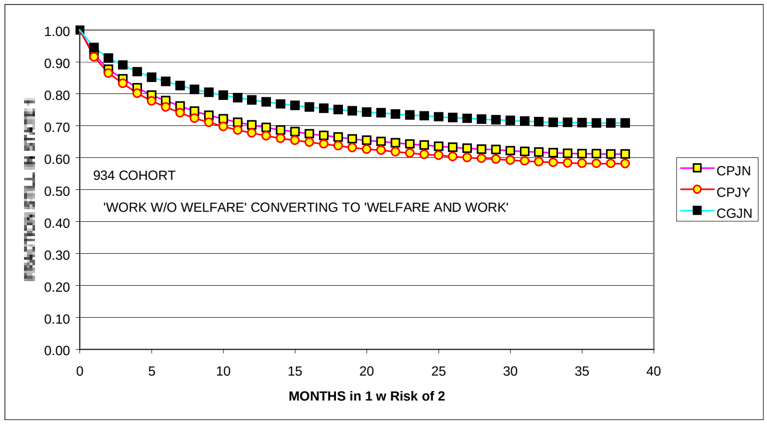

Figure 2
Controlled Survival for Average Welfare Clients
Transition from State 1 to State 2

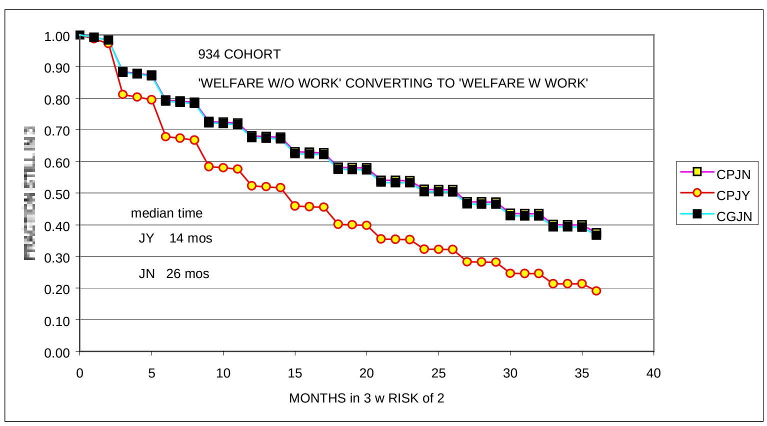

Figure 3
Controlled Survival for Average Welfare Clients
Transition from State 3 to State 2

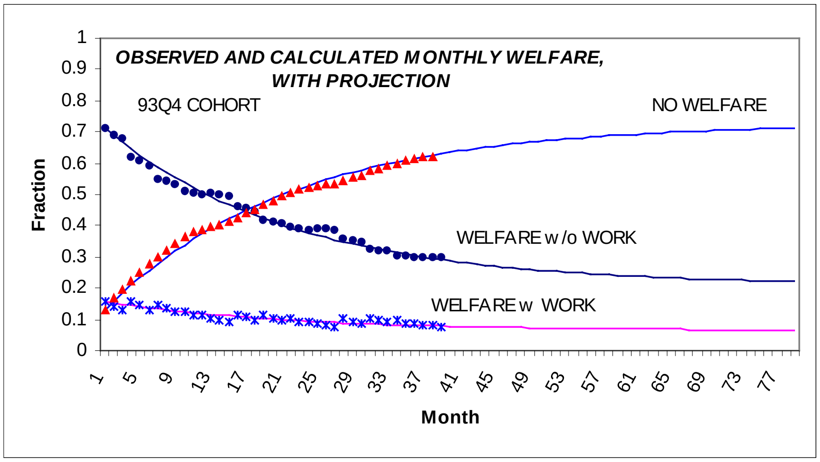

Figure 4
Simulated Outcomes for Average Welfare Clients
Based on Empirical Hazard Rates

Table 1
Explanatory Variables

$\mathbf{x_0}$	= 1, intercept
$\mathbf{x_1}$ - $\mathbf{x_{18}}$	0/1 indicators for demographic characteristics
$\mathbf{x_{19}}$	number in family
$\mathbf{x_{20}}$	0/1 indic. for working in 93Q4
$\mathbf{x_{21}}$	avg qtrly earning in prior two years
$\mathbf{x_{22}}$	months on welfare in prior two years
$\mathbf{x_{23}}$ - $\mathbf{x_{47}}$	0/1 indic. welfare pattern in 93Q4
$\mathbf{x_{48}}$ - $\mathbf{x_{52}}$	0/1 program indic.

Table 2
Reference Hazard Rates at 9 Months

Fast Transitions		
Transition	Hazard Rate @ 9 Mo	Expected Mean Stay
2 to 1	0.11 /Mo	9 Mo
2 to 3	0.08	12.5
1 to 0	0.04	25
Medium Transitions		
3 to 0		
3 to 2	all	all
0 to 1	0.02 - 0.03	~ 40 Mo
0 to 3	/Mo	
1 to 2		
Slow Transitions		
	all	all
2 to 0	0.003 -	~ 300 Mo
1 to 3	0.004/Mo	
Very Slow Transitions		
	all	all
3 to 1	~0.001	~ 1000
0 to 2	/Mo	

Table 3
Risk Ratios from PHREG Output

Risk Ratios		
Relative to CPJN		
Transition	**CPJY**	**CGJN**
0 to 1	**1.138**	**1.149**
0 to 2	° **1.404**	° **0.730**
0 to 3	**0.946**	• **0.769**
1 to 0	**0.812**	**0.944**
1 to 2	**1.103**	• **0.701**
1 to 3	**1.008**	° **0.791**
2 to 0	**1.000**	**1.000**
2 to 1	**1.140**	**1.119**
2 to 3	**0.838**	**0.905**
3 to 0	**1.000**	**1.000**
3 to 1	° **1.368**	**1.000**
3 to 2	• **1.686**	**1.020**

Table 4
Strongest and Most Certain Associations with Program Indicators

Associated with CSE Level				
	CPJN		**CGJN**	
Transition	**Hazard Rate @ 9 Mo**	**Expected Mean Stay**	**Hazard Rate @ 9 Mo**	**Expected Mean Stay**
0 to 3	**0.023 per Mo**	**43 Mo**	**0.018**	**56 Mo**
1 to 2	**0.020**	**50**	**0.014**	**71**
Associated with JOBS				
	CPJN		**CPJY**	
3 to 2	**0.029**	**35**	**0.049**	**20**