Interactive Scientific Processor (ISP) to the Estimation of Multistate Proportional Hazards Model

R. I. Chowdhury[1], K.P. Singh[2], M.A. Islam[3], A.A. Bartolucci[2], S. Bae[2]
1. Department of Health Information Administration, Kuwait University, Kuwait
2. Department of Biostatistics, The University of Alabama at Birmingham, Birmingham, Alabama, USA
3. Department of Statistics, Dhaka University , Bangladesh

Abstract: In longitudinal studies there is an implied ordering of the times of the repeated observations of outcome and covariates. The history of an individual over time shows movement from one state of outcome to another. The histories of a group of individuals produce a complete or partially censored data. The transition from one state to another can be categorized into three distinct types: I. Transition, II. Reverse transition, and III. Repeated transition. To analyze such type of data we can employ models developed by Beck (1979), Kay (1982), Kalbfleisch, Lawless and Vollmer (1983), Islam and Singh (1992), and Islam, et. al. (1997). In this paper an outline for solving equations for estimating parameters of the model developed by Kay (1982) and further extended by Islam, et. al. (1997) for reverse and repeated transition is discussed. For estimating parameters, the Interactive Scientific Processor (ISP), a statistical software package, is used. For application of the model we use longitudinal data on Diabetes mellitus (DM) collected at a diabetes hospital in Bangladesh. Although we use ISP to estimate the parameters of the model, other statistical software which can handle matrix operation (e.g., SAS, S-Plus, etc.) Can be used to estimate the parameters of this model.

1. Introduction

The longitudinal data analysis generally involves repeated observations of outcome and covariates. The history of an individual over time shows movement from one state of outcome to another. The history of a group of individuals produce complete or partially censored data. The transitions from one state to another state can be categorized into three distinct types : I. Transition, II. Reverse transition, and III. Repeated transition. To analyze such type of data we can employ models developed by Beck (1979), Kay (1982), Kalbfleisch, Lawless and Vollmer (1983), Islam and Singh (1992), and Islam, et. al. (1997). In this paper we showed an outline of the estimation procedure of the model extended by Islam (1994) for reverse and repeated transitions.

In this model we have to estimate a large number of parameters on the basis of interactive solution of simultaneous equations. The Interactive Scientific Processor for Personal Computer (PC-ISP, 1985), version 1.12 has been used for the estimation of the parameters of this model. ISP has capability to perform standard matrix operations like addition, multiplication, inverse etc., and the availability of structured statements like (IF . . . THEN . . . ELSE). It also allows loops and the execution of complex programs which reside in single or multiple files.

For application of the model we have used longitudinal data on Diabetes Mellitus (DM) collected at a diabetes hospital in Bangladesh. Although we have used ISP to estimate the parameters of this model, other statistical software which can handle matrix operation (e.g., SAS, S-Plus) can be used with the same outline presented here.

2. Different Types of Transitions of Diabetes Mellitus

In our application the state space is comprised of : state 1 : two hours blood glucose level is less than or equal to 11.1 mmol/l, (S_1), and state 2 : two hours blood glucose level is greater than 11.1 mmol/l (S_2). At the time of first registration a patient is categorized in one of two states, S_1 and S_2, on the basis of their blood glucose level. The Figure .1 shows the possible transitions for those who are detected in state S_2 at the time of first registration.

The transitions comprise of individuals who move from S_2 to S_1 during the follow-up period. A reverse transition is observed for patients who have a transition of the type S_1 -> S_2 and experienced transitions of the type S_2 -> S_1, previously, and a repeated transition occurs if we observe transition from S_2 to S_1 for those who already experienced transitions of the type S_2->S_1->S_2.

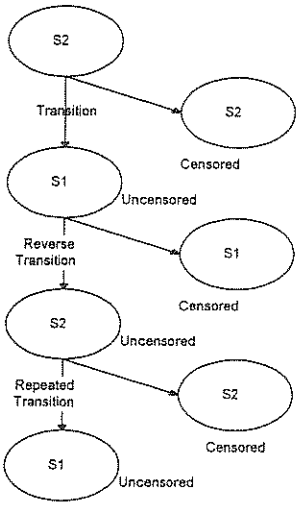

Figure 1. Transitions in Diabetes Mellitus Patients

3. Models for Transitions, Reverse Transitions and Repeated Transitions

The state space of transitions is denoted by S which comprises of states S_1 and S_2. Let I (I=1,2) be the state of origin and j (j=1,2) be the state of destination. The hazard functions are as follows:

$$\lambda(t,1/2,X(t))=\lambda_{021}(t)e^{X(t)\beta_{21}} \qquad (3.1)$$

$$\lambda(t,2/2,1,X(t))=\lambda_{012.2}(t)e^{X(t)\beta_{12.2}} \qquad (3.2)$$

$$\lambda(t,1/2,1,2,X(t))=\lambda_{021.21}(t)e^{X(t)\beta_{12.12}}. \qquad (3.3)$$

Note that (3.1) represents the hazard function of the type S_2 -> S_1, (3.2) shows the hazard function for reverse transition of the type S_1 -> S_2 conditional upon transition of the type S_2 -> S_1, and (3.3) represents a repeated transition of the type S_2 to S_1 conditional upon transition and reverse transition of the type S_2 -> S_1 and S_1 -> S_2 ,respectively. Furthermore, for any transition of the type i->j, X(t) denotes the vector of covariates $X_{ij}(q)$, q=1,2,...,Q and β_{ij} is the corresponding vector of coefficients $\beta_{ij}(q)$.

Let n_1, n_2, and n_3 be the number of individuals at risk of transitions, reverse transitions and repeated transitions respectively. Then the partial likelihood factors for such transitions are given by

$$L_1=L(\beta_{21})=\prod_{m_1=1}^{n_1}\frac{e^{X_{m_1}(t)\beta_{21}}}{\sum_{l_1\in R(t_{12m_1})}e^{X_{l_1}(t_{l_1})\beta_{21}}} \qquad (3.4)$$

where R(t(k)) is the risk set for transitions from S_2 to S_1.

Similarly, the partial likelihood factors for reverse transitions and repeated transitions are as follows:

$$L_2=L(\beta_{12.2})=\prod_{m_2=1}^{n_2}\frac{e^{X_{m_2}(t)\beta_{12.2}}}{\sum_{l_2\in R(t_{12.2m_2})}e^{X_{l_2}(t_{l_2})\beta_{12.2}}}, \qquad (3.5)$$

and

$$L_3=L(\beta_{12.12})=\prod_{m_3=1}^{n_3}\frac{e^{X_{m_3}(t)\beta_{12.12}}}{\sum_{l_3\in R(t_{12.12m_3})}e^{X_{l_3}(t_{l_3})\beta_{12.12}}}, \qquad (3.6)$$

respectively.

We obtain the likelihood estimates using the Newton-Raphson method from the following equations

$$\frac{\partial \ln L_i}{\partial\beta(q)} = 0, \quad i=1,2,3; \qquad (3.7)$$

where β is the vector of parameters for transitions, reverse transitions, and repeated transitions. The estimated covariance matrix can be obtained from

$$\hat{Var}(\hat{\beta}) = [I(\hat{\beta})]^{-1}$$

where (q,v) th element of $I(\hat{\beta})$ is defined as

$$I_{qv}(\hat{\beta}) = -\frac{\partial^2 \ln L_i}{\partial \beta(q)\, \partial \beta(v)}$$

where $q=1,2,...,Q$; $v=1,2,...,Q$. (3.8)

We can test the null hypothesis

$$H_0: \beta = \beta_0$$

by using the asymptotic chi-square

$$\chi^2_Q = (\hat{\beta}-\beta_0)' I(\hat{\beta}) (\hat{\beta}-\beta_0). \quad (3.9)$$

4. Estimation Procedure

The algorithm requires starting values for the parameter vector β and the data matrix. The iterative estimation procedure updates the parameters in every cycle of iteration, starting from initial values and converging to maximum likelihood estimates. As stated earlier Newton-Raphson method is used for convergence and we use ISP for the estimation following are the steps involved in the estimation procedures in our ISP program:

1. Input sample size, number of independent variables, number of transitions, maximum number of iterations and data file name.
2. Initialize variables
3. Starting loop for iteration
4. Starting loop for transition
5. Starting loop for risk set calculation
6. Computation of equations 3.7 and 3.8
7. End of risk set loop
8. Compute score vector and covariance matrix for all the transitions
9. End of transition loop
10. Compute $\hat{\beta}$ and chi-square
11. Check the condition for convergence
12. If converged stop program and print results
13. End loop for Iteration

4.1 Data file

The data file used by our program consists of the variables in the order: First column is for transition type(e.g., transition=1, reverse transition=2, repeated transition=3). Second column is for Censoring (uncensored=0, censored=1). Third column is for duration for a transition and the remaining columns are for independent variables.

4.2 Example of ISP Program

We have written three ISP macro files consisting ISP codes which perform the computations. From our experience it is efficient and easy to program in ISP as compared to other statistical softwares like SAS. We face some problems when we deal with a large number of variables. Following are some ISP codes which compute the score vector and covariance matrix.

```
ZP(*,*)=0
DVD=array($PP)
OVD=ARRAY(ONV)
DO JJ=IOTA(ARRAY(SIZ1))
E1=ARRAY($PP,1)
temp=array($PP,$PP)
ZB=EXP(MPY(Z(JJ:SIZE,*),B0))
TSUM=SUM(ZB)
E1=TRN((mpy(TRN(Z(JJ:SIZE,*)),ZB))/TSU
M)
ZP(JJ:JJ,*)=Z(JJ:JJ,*)-E1
DO LK=IOTA(ARRAY($PP))
DVD(LK)=DVD(LK)+((tsum*sum(Z(JJ:SIZE,L
K)*Z(JJ:SIZE,LK)*ZB))-\
(sum(Z(JJ:SIZE,LK)*ZB)*sum(Z(JJ:SIZE,L
K)*ZB)))/(TSUM**2)
END DO
DELETE LK
COU=0
IF ($PP>1) THEN
   PQ=$PP-1
   DO I1=IOTA(ARRAY(PQ))
   PQ1=$PP-I1
   DO I2=IOTA(ARRAY(PQ1))
   COU=COU+1
   IJ=I2+I1

OVD(COU)=OVD(COU)+((tsum*sum(Z(JJ:SIZE
,I1)*Z(JJ:SIZE,IJ)*ZB))-\

(sum(Z(JJ:SIZE,I1)*ZB)*sum(Z(JJ:SIZE,I
J)*ZB)))/(TSUM**2)
   END DO
   END DO
   DELETE PQ PQ1 IJ COU
END IF
END DO    # end for individuals loop
```

5. An Application to Diabetes Mellitus Data

The Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic Disorders (BIRDEM), keeps record of all the patients since their first registration. Information on the change in the status regarding blood glucose level along with other relevant characteristics are recorded during each follow-up visit. In this study, we have considered 60, 60 and 62 patients being at risk of transition, risk of reverse transition and risk of repeated transition, respectively. The Table 1 presents the distribution of uncensored and censored cases by transitions.

Four variables are included in our models for three types of transitions: Sex, (Male=1, Female=0), BMI, Age and Area (Rural=0, Urban=1). Estimates of parameters of models for transitions, reverse transitions, and repeated transitions are shown in Table 2.

Table 1. Distribution of Censored and Uncensored Cases by Type of Transitions

Transition type	Uncensored	Censored	Total
Transition	33	27	60
Reverse Transition	35	25	60
Repeated Transition	40	22	62
Total	108	74	182

The models are statidtically significant at 5% level (P-value<0.05). For transition from state 2 to state 1 only sex shows significant association indicating a higher rate of transition for male patients than that of females. There is a significant association (P-value<.05) between BMI and the blood glucose level. This implies that a lower BMI accelerates a transition from lower level of blood glucose to a higher level.

Table 2. Estimates of Parameters for Proportional Hazards Model for Transitions, Reverse Transitions and Repeated Transitions

Types of Transitions and Variables	Estimate ($\hat{\beta}$)	SE ($\hat{\beta}$)
Transition		
Sex	-8.8606	2.1179
BMI	0.1011	0.3622
Age	-0.0633	-1.8567
Area	-0.0097	0.0144
Reverse Transition		
Sex	-0.0036	0.0988
BMI	-0.4337	0.3923
Age	-0.0204	0.0455
Area	-0.0013	0.0162
Repeated Transition		
Sex	-0.1806	0.2979
BMI	0.6517	0.3504
Age	0.0433	0.0469
Area	0.0139	0.0157

6. Conclusions

In this paper we made an attempt to show an application of multistate proportional hazards model for repeated observations and the estimation procedure of the model using ISP. The model discussed in this paper has good applications for analyzing data with a number of transient and absorbing states emerging from follow-up studies.

The ISP program we used for the estimation purpose is a DOS-BASED program. The version is also very old. We faced problems when we were dealing with a large number of variables. Now, we are rewriting this program in S- Plus.

7. References

Beck, G.J., Stochastic Survival Models with Competing Risks and Covariates, *Biometrics,* 35, 427-438, 1979.

Islam, M.A. and Singh, K.P., Multistate Survival Models for Partially Censored Data. *Environmetrics,* 3, 223-234, 1992.

Islam, M.A., Multistate Survival Models for Transitions and Reverse Transitions: An Application to Contraceptive Use Data. *JRSS,* 157. Number 3, 441-445, 1994.

Islam, M.A., Singh, K.P, Chowdhury, R.I, Bartoulucci, A.A, Multistate Hazards Model for Transitions and Reverse Transitions, Proceedings of the Third ISSAT International Conference on Reliability & Quality in Design, Anaheim, California, March 12-14, 213-217, 1997.

Kalbfleisch, J.D., Lawless, J.F., and Vollmer, W.M., Estimation in Markov Models from Aggregate Data. *Biometrics,* 39, 907-919, 1983.

Kay, R., The Analysis of Transition Times in Multistate Stochastic Processes Using Proportional Hazard Regression Models. *Communications in Statistics (Theory and Method),* 1743-1756, 1982.

PC-ISP., Interactive Scientific Processor - User's Guide and Command Descriptions. Chapman and Hall, Newyork, 1985.