

Análisis de supervivencia en presencia de riesgos competitivos, un enfoque no paramétrico

Luz Bullón C. ¹

Resumen

En muchas investigaciones hay interés en estudiar el tiempo de vida (o de sobrevivida) con el objetivo de estimar la función de supervivencia. Particularmente en investigación clínica o biomédica, es muy frecuente encontrar datos de tiempo de vida asociados a múltiples eventos causas de falla. Dichos eventos pueden ser concurrentes, ordenados o aleatorios. En esta situación se dice que el análisis es de riesgos competitivos y la metodología que aquí se presenta es con enfoque no paramétrico de Kaplan-Meier. El método se ilustra con datos presentados en Molinero (2004) y se efectúa una comparación de esta metodología frente a la de Kaplan-Meier clásico, concluyendo que en presencia de eventos competitivos no es conveniente usar este último.

Palabras clave: riesgos competitivos, análisis no paramétrico de tiempo de vida.

Abstract

Many researchs have the interest in studying the life time (or survival time) with the objective of estimate the survival function. Both, clinics essays or biomedical studies is often to find data about multiple events of failure. In this case we are in a case named competing risk. The methodology is going to be showed in this paper, is the nonparametric Kaplan-Meier estimator. The competing risk analyse is explained using the dates taken of the paper published by Molinero (2004). A comparison between the classic methodology by Kaplan-Meier and competing risk analyse is also presented, concluding what in presence of competing events isn't good to use classical Kaplan Meier estimator.

Key words: competing risks, Nonparametric estimation from failure times.

1. Introducción¹

El análisis del tiempo de supervivencia t , implica el registro del tiempo desde un instante t_0 , claramente definido, hasta que ocurra un evento, por lo general falla, hasta un tiempo preestablecido τ , o hasta que el sujeto sea retirado del estudio. La inclusión en el análisis, de estos últimos denominados censuras, requiere de especializaciones de carácter teórico por parte de la estadística. La censura es identificada por $\delta = 1$, si ocurre falla, o $\delta = 0$ si ocurre censura. Asimismo, en diversos estudios se incluyen covariables fijas z , recolectadas al tiempo t_0 .

Los objetivos más generales de los métodos de análisis de supervivencia considerando datos del tipo (t_j, δ_j, z_j) , $j = 1, 2, \dots, n$ son estimar las tasas de supervivencia, esto es, estimar $P(T > t | z)$ y verificar la significación de factores de pronóstico z . Un área específica del análisis de supervivencia estudia los tiempos de vida (o de sobrevivida) asociados a las causas de falla y se denomina análisis de riesgos competitivos. Los datos que provienen de una situación de riesgos competitivos son del tipo (t_j, i, δ_j, z_j) , $j = 1, 2, \dots, n$ donde i identifica la causa de falla C_i , $i = 1, 2, \dots, k$.

Por ejemplo, al evaluar la eficacia de trasplantes de También, en estudios de mortalidad, puede ser más interesante estudiar separadamente las muertes

debido a enfermedades del corazón, diabetes, cáncer y otros, que combinar todas las causas. Estas diferentes causas de falla son consideradas como eventos competitivos, los cuales introducen el concepto de riesgos competitivos.

Asimismo, en las condiciones anteriormente descritas, suponga que se quiere estimar la función de supervivencia sin sida de los pacientes infectados por el virus de la inmunodeficiencia humana (VIH) y tratados con medicamentos antiretrovirales en un país subdesarrollado (con mayor mortalidad por otras causas), entonces la curva de supervivencia se debe de ajustar por riesgos competitivos.

Los objetivos con datos de este tipo son estimar las tasas de supervivencia totales (sin considerar las causas de falla) además de las tasas de supervivencia por causas específicas.

A manera de resumen podemos decir que los métodos estadísticos para datos de supervivencia en presencia de riesgos competitivos, hacen uso de los modelos paramétricos y semi paramétricos de Cox. Sin embargo, también hay uso de métodos no paramétricos. En el presente trabajo presentamos estos últimos aplicándolos además, a un conjunto de datos.

La función de supervivencia es definida como la probabilidad de que una observación no presente un evento de interés hasta el tiempo t . En términos probabilísticos, $S(t) = P(T \geq t)$. En presencia de observaciones censuradas, las mismas que son

¹ Facultad de Economía y Planificación, Universidad Nacional Agraria La Molina. E-mail: luz@lamolina.edu.pe.

frecuentes en estudios clínicos, es común usar el estimador no paramétrico de Kaplan-Meier para estimar la probabilidad de supervivencia. Si se asume n pacientes el estimador de Kaplan-Meier clásico, en el tiempo t , viene dado por la siguiente fórmula,

$$KM = \prod_{t_j < t} \left(1 - \frac{d_j}{n_j} \right)$$

donde n_j es el número de pacientes en riesgo en el tiempo t_j y d_j el número de pacientes que presentan el evento principal al tiempo t_j .

En investigaciones biomédicas es muy frecuente encontrar estudios cuando dos o más eventos de falla compiten por ocurrir en el individuo. Therneau (1997), sugirió para propósitos de análisis, que los eventos de falla deben ser clasificados de acuerdo a si los eventos ocurren en un orden o si ellos son recurrentes. Los eventos ordenados pueden resultar de un estudio que registra el tiempo hasta la ocurrencia del primer evento, luego el segundo, etc. Por ejemplo, el tiempo hasta la ocurrencia del primer infarto de miocardio, luego se registra el tiempo hasta la ocurrencia del segundo infarto de miocardio y así sucesivamente. Estos son eventos ordenados en el sentido que el segundo no puede ocurrir antes del primero. Por otro lado, en eventos no ordenados, la ocurrencia de estos es aleatoria, pudiendo ser del mismo tipo o diferentes. Un ejemplo de eventos recurrentes es el desarrollo de cáncer de mamas en familias genéticamente propensas.

Las investigaciones de eventos de falla no ordenados de diferentes tipos son frecuentemente llamados estudios de eventos competitivos. En estos estudios, un paciente puede presentar un estado o evento de un conjunto de posibles estados finales en forma aleatoria.

La forma más simple de analizar datos de falla múltiple es examinar el primer evento e ignorar el resto tomándolo como censura. Esta aproximación, sin embargo, no es adecuada porque desecha posible información relevante.

El estimador de Kaplan-Meier se ha convertido en un estándar en el análisis de supervivencia en la epidemiología poblacional. En cambio el análisis de supervivencia en presencia de riesgos competitivos tiene un mayor impacto en la epidemiología clínica que la poblacional.

Un importante supuesto del método de Kaplan-Meier en el análisis de supervivencia es, que la probabilidad de que una observación sea censurada es independiente del tiempo de supervivencia. Esta asunción puede ser admitida en presencia de censuras no informativas, es decir, si esta se produce porque terminó el estudio o por un límite en el número de fallas. En cambio si la censura se produce por otras causas que puedan estar relacionadas con el tiempo de supervivencia el método clásico no sería conveniente.

3. Metodología

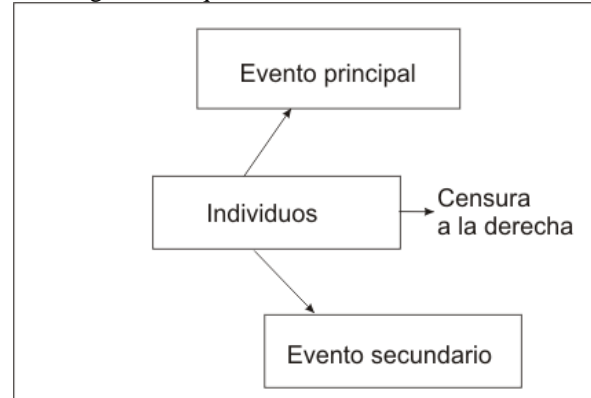
En riesgos competitivos, la variable dicotómica δ indica censura cuando $\delta = 0$ y falla (por causa

$C_i, i = 1, \dots, k$) cuando $\delta = 1$. Por lo tanto, cada observación es de la forma (t_j, i_j, δ_j) , $j = 1, 2, \dots, n$ sin presencia de covariables. Donde i identifica la causa de falla C_i .

Prentice et al (1978) definió la función de riesgo para cada causa de falla como:

$$\lambda_i(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt, C = i / T \geq t)}{dt}$$

Para entender mejor la forma de trabajar con riesgos competitivos, se parte de un modelo de dos eventos competitivos mutuamente excluyentes representados en el siguiente esquema:



Como se puede observar en la figura anterior el individuo al final del estudio puede presentar cualquiera de los dos eventos (estados finales) o simplemente no presentar ninguno convirtiéndose por tanto en una observación censurada. Si se tienen n pacientes en un estudio y se denota como t_j el tiempo de seguimiento del individuo j bajo el esquema presentado anteriormente, el modelo para el evento 1, utilizando el estimador de Kaplan-Meier en presencia de dos eventos competitivos sería,

$$\hat{KM}_1(t) = \prod_{t_j < t} \left(1 - \frac{f_j}{n_j} \right)$$

mientras que para el evento 2,

$$\hat{KM}_2(t) = \prod_{t_j < t} \left(1 - \frac{r_j}{n_j} \right)$$

Por lo tanto, el estimador de la distribución de las fallas asociadas a los dos eventos es:

$$\begin{aligned} \hat{KM}_{12} &= \hat{KM}_1(t) \times \hat{KM}_2(t) \\ &= \prod_{t_j < t} \left(1 - \frac{f_j}{n_j} \right) \prod_{t_j < t} \left(1 - \frac{r_j}{n_j} \right) \\ &= \left(1 - \frac{f_1}{n_1} \right) \left(1 - \frac{f_2}{n_2} \right) \dots \left(1 - \frac{f_j}{n_j} \right) \left(1 - \frac{r_1}{n_1} \right) \left(1 - \frac{r_2}{n_2} \right) \dots \left(1 - \frac{r_j}{n_j} \right) \\ &= \left(1 - \frac{f_1}{n_1} \right) \left(1 - \frac{r_1}{n_1} \right) \left(1 - \frac{f_2}{n_2} \right) \left(1 - \frac{r_2}{n_2} \right) \dots \left(1 - \frac{f_j}{n_j} \right) \left(1 - \frac{r_j}{n_j} \right) \\ &= \prod_{t_j < t} \left(1 - \frac{f_j}{n_j} \right) \left(1 - \frac{r_j}{n_j} \right) \end{aligned}$$

donde, n_j es el número de pacientes en riesgo en el tiempo t_j , f_j el número de pacientes que presentan el evento principal en tiempo t_j y r_j el número de

pacientes que presentan el evento secundario en el tiempo t_j .

Si se considera que dos fallas no se producen simultáneamente en un individuo, es decir, $f_j = 0$ o

bien $r_j = 0$ y $\frac{f_j r_j}{n_j^2} = 0$, entonces

$$\widehat{KM}_{12} = \prod_{t_j < t} \left(1 - \frac{f_j + r_j}{n_j} \right)$$

Para estimar las probabilidades de los tiempos de falla para los dos eventos respectivamente, se utiliza el siguiente estimador para el evento 1,

$$\widehat{R}_1(t) = \sum_{t_j < t} \widehat{KM}_{12}(t_{j-1}) \times \frac{f_j}{n_j}$$

mientras que para el evento 2,

$$\widehat{R}_2(t) = \sum_{t_j < t} \widehat{KM}_{12}(t_{j-1}) \times \frac{r_j}{n_j}$$

Como se puede observar, las probabilidades de los tiempos de falla, se calculan como la suma de los estimadores de Kaplan-Meier en cada tiempo ponderados por la proporción de fallas en ese mismo tiempo. Prentice *et al* (1978), entre otros autores, denominan a las probabilidades de falla como incidencia del evento principal o del secundario.

Para el cálculo de la probabilidad de supervivencia estimada se utilizará la siguiente expresión

$$S(t) = 1 - IC(t)$$

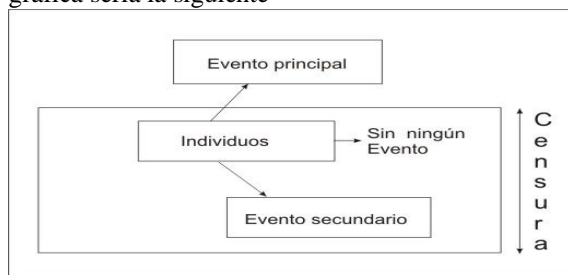
donde $IC(t) = R_1(t) + R_2(t)$.

Para calcular el error estándar para la probabilidad de supervivencia se realizará mediante la formula de Greenwood (1926).

$$Desv = S^2(t_j) \sum \frac{d_j}{n_j(n_j - d_j)}$$

donde $d_j = f_j + r_j$

Suponga que el modelo anterior fuese tratado en presencia de eventos competitivos mediante el análisis común de Kaplan Meier, la representación gráfica sería la siguiente



Como se puede observar en la figura anterior los individuos que no sufrieron ningún evento y aquellos que presentaron el segundo evento, se unen para formar un solo estado que en el análisis serían considerados como datos censurados a la derecha y donde el evento principal es el único de interés. Por lo tanto, en el análisis común de supervivencia mediante los estimadores de Kaplan-Meier, en presencia de eventos competitivos trata al evento 2 como un dato censurado. Luego, esto conlleva a la sobrestimación de la probabilidad de supervivencia.

Para un modelo de riesgos competitivos que considera un estado inicial y k posibles eventos finales se tiene $t_{i1} < t_{i2} < \dots < t_{in_i}$ los n_i tiempos de falla por causa $C_i, i = 1, 2, \dots, k$, el estimador de la distribución de fallas asociada a la causa C_i es

$$\widehat{KM}_{12\dots k}(t) = \prod_{t_{ij} < t} \left(\frac{n_{ij} - d_{ij}}{n_{ij}} \right)$$

donde n_{ij} es el número de sujetos en riesgo antes de t_{ij} . Las fallas por causas diferentes de C_i son tratados como censuradas en la estimación. El estimador de la distribución de probabilidades de los tiempos de falla por causa C_i está dado por la ecuación que sigue y los gráficos de las funciones de supervivencia y riesgo son funciones escalonadas.

$$\widehat{R}_i(t) = \sum_{t_{ij} < t} \widehat{KM}_{12\dots k}(t_{j-1}) \times \frac{d_{ij}}{n_{ij}}$$

2. Resultados y discusión

La metodología previamente descrita, ha sido aplicada a un conjunto de datos provenientes del contexto que describimos a continuación. Se tiene interés en estimar la probabilidad de supervivencia al trasplante de un órgano, en donde el evento a investigar es la muerte del paciente por cualquiera de dos sucesos alternativos, el rechazo al órgano transplantado (evento principal) o la muerte (antes de sufrir el rechazo), asociada a otras causas relacionadas con el trasplante (evento secundario o competitivo). Luego, estamos frente a un problema de estimación en presencia de eventos competitivos. Los datos corresponden a 390 pacientes en los cuales se realizó el trasplante y presentan la estructura siguiente,

ID	ESTADO	DIAS
1	1	1805
2	0	628
3	2	10
4	1	31
5	1	463
...		
389	1	25
390	1	2

La covariable estado se define como 0 para los pacientes vivos (datos censurados), 1 para los que fallecen por rechazo del órgano transplantado y 2 para los que presentan el suceso alternativo el cual está relacionado al no correcto trabajo del corazón. El tiempo se ha medido en días. El análisis exige la presentación de los tiempos de seguimiento ordenados de menor a mayor.

El programa estadístico especializado para la estimación no paramétrica de la curva de supervivencia en presencia de eventos competitivos, usado en la presente aplicación fue el SAS 7.0 mediante la macro **competingrisk.sas**, presentada en el ANEXO. Las tablas siguientes presentan parte de la salida al usar el correspondiente paquete estadístico.

Tabla 1. Resultados con la metodología de Kaplan-Meier.

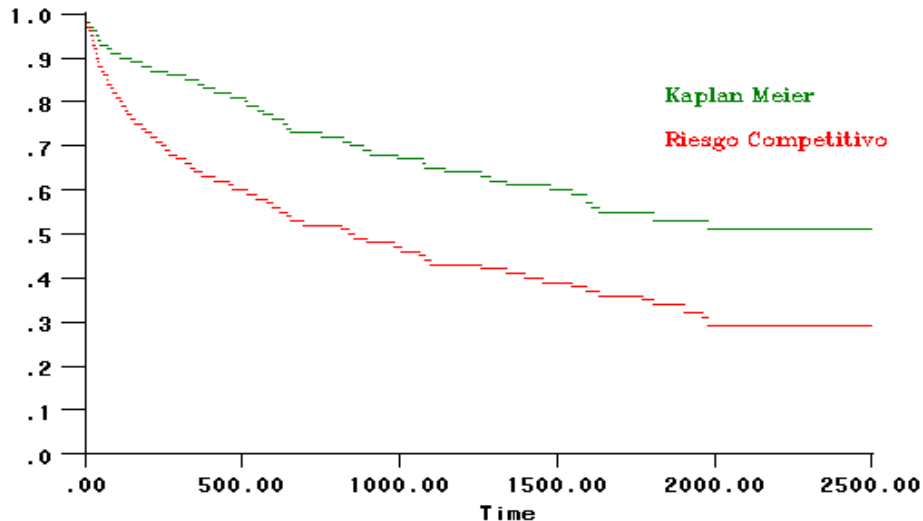
Obs	tiempo	Frecuencia del evento principal	Observaciones censuradas	Pacientes en riesgo	Supervivencia libre de eventos	Error estándar libre de todos los eventos
1	1	2	0	390	0.99487	0.003617
2	2	1	1	388	0.99231	0.004424
3	3	1	0	386	0.98974	0.005105
4	5	2	1	385	0.98460	0.006240
5	6	1	1	382	0.98202	0.006735
6	7	0	1	380	0.98202	0.006735
7	10	0	1	379	0.98202	0.006735
8	11	1	0	378	0.97942	0.007201
9	13	0	1	377	0.97942	0.007201
10	14	1	1	376	0.97682	0.007639
...						
...						
...						
321	2080	0	1	11	0.51010	0.047437
322	2084	0	1	10	0.51010	0.047437
323	2087	0	1	9	0.51010	0.047437
324	2100	0	1	8	0.51010	0.047437
325	2136	0	1	7	0.51010	0.047437
326	2137	0	1	6	0.51010	0.047437
327	2140	0	1	5	0.51010	0.047437
328	2154	0	1	4	0.51010	0.047437
329	2156	0	1	3	0.51010	0.047437
330	2168	0	1	2	0.51010	0.047437
331	2171	0	1	1	0.51010	0.047437

Tabla 2. Resultados con la metodología de Riesgos Competitivos.

Obs	tiempo	Fi del evento principal	Fi del evento secundario	Oi censuradas	Pacientes en riesgo	Incidencia del evento principal	Incidencia del evento secundario	Supervivencia libre de eventos	Error estándar libre de todos los eventos
1	1	2	0	0	390	0.005128	0.000000	0.99487	0.003617
2	2	1	0	1	388	0.007692	0.000000	0.99231	0.004424
3	3	1	0	0	386	0.010263	0.000000	0.98974	0.005105
4	5	2	1	0	385	0.015405	0.002571	0.98202	0.006733
5	6	1	0	1	382	0.017975	0.002571	0.97945	0.007189
6	7	0	1	0	380	0.017975	0.005148	0.97688	0.007618
7	10	0	1	0	379	0.017975	0.007726	0.97430	0.008022
8	11	1	0	0	378	0.020553	0.007726	0.97172	0.008405
9	13	0	1	0	377	0.020553	0.010303	0.96914	0.008769
10	14	1	1	0	376	0.023130	0.012881	0.96399	0.009450
...									
...									
...									
321	2080	0	0	1	11	0.369640	0.334800	0.29557	0.037215
322	2084	0	0	1	10	0.369640	0.334800	0.29557	0.037215
323	2087	0	0	1	9	0.369640	0.334800	0.29557	0.037215
324	2100	0	0	1	8	0.369640	0.334800	0.29557	0.037215
325	2136	0	0	1	7	0.369640	0.334800	0.29557	0.037215
326	2137	0	0	1	6	0.369640	0.334800	0.29557	0.037215
327	2140	0	0	1	5	0.369640	0.334800	0.29557	0.037215
328	2154	0	0	1	4	0.369640	0.334800	0.29557	0.037215
329	2156	0	0	1	3	0.369640	0.334800	0.29557	0.037215
330	2168	0	0	1	2	0.369640	0.334800	0.29557	0.037215
331	2171	0	0	1	1	0.369640	0.334800	0.29557	0.037215

La probabilidad estimada de que un paciente que fue sometido a un transplante de corazón hasta el tiempo de 2080 días (aproximadamente 5 años) no haya muerto por el rechazo de este ni por otras causas es de 0.29557 mediante el análisis en presencia de eventos competitivos y 0.51010 mediante el estimador de Kaplan-Meier.

Curvas de Supervivencia



3. Conclusiones

Como se puede observar en ambas tablas, hasta el tiempo 6, las probabilidades de supervivencia tanto por el método de Kaplan-Meier como por el de riesgos competitivos coinciden. A partir del tiempo 7 dichas probabilidades empiezan a ser diferentes. Así, la probabilidad de supervivencia estimada para el tiempo de 10 días mediante el estimador de Kaplan-Meier en presencia de dos eventos competitivos es 0.97682, mientras que por el análisis de riesgos competitivos es 0.96399, es decir, hay una pequeña diferencia entre ambas probabilidades.

A medida que el tiempo es mayor, estas diferencias se hacen más grandes como se puede observar en el resto de filas de las tablas. Dichas diferencias, se encuentran mejor representadas en la figura 1.

En consecuencia, el análisis de supervivencia común con métodos actuariales y con estimadores de Kaplan-Meier no es aplicable en presencia de riesgos competitivos de muerte ya que *sobreestima siempre la probabilidad de supervivencia*, Sá Carvalho *et al.* (2005).

Recomendaciones

Si el objetivo es comparar tratamientos para determinar el mejor, como por ejemplo, la eficacia de dos antiretrovirales, será más adecuado utilizar el estimador de Kaplan-Meier clásico.

Es necesario explorar casos de eventos múltiples de falla como es el de eventos paralelos y/o ordenados así como diferenciarlos de los eventos competitivos, debido a que de ello depende la forma de ingresar los datos y obviamente las interpretaciones de los resultados.

Se recomienda explorar la metodología de la regresión de Cox bajo la presencia de covariables para la situación de riesgos competitivos, así como el enfoque de cadenas de Markov.

4. Referencias bibliográficas

- ANDERSON, B. (1999). *Algorithms for competing risks*. September; 1-9.
- KAPLAN E.L., MEIER, P. (1958). *Nonparametric estimation from incomplete observations*. J Am Stat Assoc; **53**:457-81.
- LLORCA J, DELGADO-RODRÍGUEZ M. (2000). Competing risks in absence of independence. Impact of AIDS on liver function failure mortality, and lung cancer on ischemic heart disease mortality. J Clin Epidemiol; **53**:1145-9.
- LLORCA, J., DELGADO-RODRÍGUEZ, M. (2001). Competing risks analysis using Markov chains: impact of cerebrovascular and ischaemic heart disease in cancer mortality. International Journal of Epidemiology **30**:99-101.
- MOLINERO, L. (2004). Introducción al análisis de supervivencia cuando existen riesgos "competitivos" o cuando existen diferentes estados posibles (modelos multi-estado). Asociación de la Sociedad española de hipertensión.
- PRENTICE, R.L., KALBFLEISCH, J.D., PETERSON, A.V., FLOURNOY, N., FAREWELL, V.T., BRESLOW N.E. (1978). *The analysis of failure times in the presence of competing risks*. Biometrics; **34**:541-54.
- SÁ CARVALHO, M., et al (2005) Análise de sobrevivência, Teoria e aplicações em Saúde. Caderno de respostas. UFRGS, Brasil.

ANEXO

```

/***** Algorithms for competing risks
Analysis *****/

```

```

/*****
*****
/* References: */
/* Andersen, PK, Borgan, O, Gill, R., and Keiding,
N, Statistical Models based on Counting Processes,
Springer, 1993 */
/* Anderson, William N., Algorithms for Actuarial
and Actual Analysis, pages 128 - 133 in Proceedings
of 8th Annual */
/* Western Users of SAS Software (WUSS),
Scottsdale, AZ, 2000 */
/* Gaynor, JJ, et al., On the use of Cause-Specific
Failure and Conditional Failure Probabilities:
Examples for */
/* Clinical Oncology, JASA, vol 88, 1993, pages 400
- 409. */
/* Grunkemeier, Gary L., W. R. Eric Jamieson,
Craig D. Miller, and Albert Starr, Actuarial vs.
Actual Risk of Porcine */
/*****
*****/

```

```

%macro Greenwood(inset =, outset =, timevar =,
eventvar =, lastevnt =);

```

```

data tempset (keep = &timevar &eventvar); set
&inset;
/* Kill undefined times, and reclassify events */
if &timevar >= 0;
if &eventvar < 1 | &eventvar > &lastevnt then
&eventvar = 0; /* Censor these events */
run;

```

```

/* Count total at risk */
/* Add dummy observations to ensure that all
designated events are reflected by variables in the
output data set */
/* These will have missing time, and will be killed in
a couple of steps */

```

```

data tempset; set tempset nobs = numpats end = last;
output;
if last then do; &timevar = .; do &eventvar = 1 to
&lastevnt; output; end; end;
call symput('numpats', put(numpats, 10.));
run;

```

```

proc sort data = tempset;
by &timevar;
run;

```

```

proc freq noprint data = tempset;
by &timevar;
tables &eventvar/out = temp;
run;

```

```

proc transpose data = temp out = &outset;

```

```

by &timevar;
id &eventvar;
var count;
run;

```

```

data &outset; set &outset; if &timevar ^= .; run;

```

```

/* kill the dummy observations */

```

```

/* Now add an at risk column and put missing counts
to zero */
/* Also make the variable names as promised */

```

```

data &outset; set &outset;
rename _0 = censored;
%do event = 1 %to &lastevnt; rename _&event =
event&event; %end;
attrib atrisk label = 'Patients at Risk';
retain atrisk newrisk &numpats;
drop newrisk;
atrisk = newrisk;
if _0 = . then _0 = 0; newrisk = atrisk - _0;
%do event = 1 %to &lastevnt;
if _&event = . then _&event = 0;
newrisk = newrisk - _&event;
%end;
run;

```

```

/* Now do the actual computations */

```

```

data &outset; set &outset;
/* Some labels for variables coming in */
attrib
censored label = 'Censored Observations'
%do event = 1 %to &lastevnt; event&event label =
"Incidents of Event &event" %end;
;

```

```

/* Components of present incidence vector */
/* (In the general model, this is the first row of the
transition matrix) */
/* Only the first row is variable */

```

```

attrib
pk0_0 label = 'Freedom from all events'
%do event = 1 %to &lastevnt;
pk0_&event label = "Incidence of event &event"
%end;;

```

```

/* Components of previous incidence vector */
attrib
pkb0_0 label = 'Previous freedom from all events'
%do event = 1 %to &lastevnt;
pkb0_&event label = "Previous incidence of event
&event"
%end;;

```

```

/* Components of present covariance matrix */
attrib
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;

```

```

ck&row._&col label = "Variance component
(&row, &col)"
%end; %end;;

/* Components of previous covariance matrix */
attrib
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
ckb&row._&col label = "Previous variance
component (&row, &col)"
%end; %end;;

/* Standard Errors */
attrib error0 label = 'Standard Error of Freedom from
All Events'
%do event = 1 %to &lastevnt; error&event label =
"Standard Error of Incidence of Event &event"
%end;;

/* Components of variance addition due to new
terms */
/* These are from the multinomial distribution */
attrib
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
va&row._&col label = "Additional variance
component (&row, &col)"
%end; %end;;

/* Components of differential transition Matrix */
/* All but the first row are constant, but are made
explicit for clarity of code */
attrib
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
da&row._&col label = "Differential transition
(&row, &col)"
%end; %end;;
/* retain variables needed for the recursion */
/* also initialize, for use the first time around */
retain
pkb0_0 1 %do event = 1 %to &lastevnt;
pkb0_&event 0 %end;
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
ckb&row._&col 0
%end; %end;;

/* Now to the general case */
/* Compute terms needed for the update */
/* Differential transition matrix */
da0_0 = 1;
%do event = 1 %to &lastevnt;
da0_&event = event&event/atrisk;
da0_0 = da0_0 - da0_&event;
%end;
/* The rest of this section is logically redundant, and
kept in to clarify the matrix multiplication */
%do row = 1 %to &lastevnt; %do col = 0 %to
&lastevnt;
%if &row = &col %then %do; da&row._&col = 1;
%end;
%else %do; da&row._&col = 0; %end;

%end; %end;

%end; %end;

/* Additional variance component */
/* The matrix is symmetric; for clarity all terms are
explicitly created */
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
%if &row = &col %then %do; va&row._&col =
da0_&row*(1 - da0_&row)/atrisk; %end;
%else %do; va&row._&col =
da0_&row*da0_&col/atrisk; %end;
%end; %end;

/* Updating the incidence vector is one matrix
multiplication */
%do col = 0 %to &lastevnt;
pk0_&col = 0;
%do event = 0 %to &lastevnt;
pk0_&col = pk0_&col +
pkb0_&event*da&event._&col;
%end; %end;

/* Updating the Variance matrix is more complicated
*/
/* There is the sum of two products of three matrices
each */
/* The first part of the first product has the
temporary name t1 */
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
t1&row._&col = 0;
%do event = 0 %to &lastevnt;
t1&row._&col = t1&row._&col +
da&event._&row*ckb&event._&col;
%end; %end; %end;

/* Finishing the first product */
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
ck&row._&col = 0;
%do event = 0 %to &lastevnt;
ck&row._&col = ck&row._&col +
t1&row._&event*da&event._&col;
%end; %end; %end;

/* Now add in the second term */
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
ck&row._&col = ck&row._&col +
va&row._&col*pkb0_0**2;
%end; %end;

/* Finally fix up the terms to be retained for next
time around */
%do event = 0 %to &lastevnt; pkb0_&event =
pk0_&event; %end;
%do row = 0 %to &lastevnt; %do col = 0 %to
&lastevnt;
ckb&row._&col = ck&row._&col;
%end; %end;

/* And the output standard errors */

```

```

%do event = 0 %to &lastevnt; error&event =      2  1  3  6  4  2
sqrt(ck&event.&event); %end;                    5  0  1  3  2  1
run;                                             8  2  5  1  3  2
;
data &outset /* keep only those variables likely to be
of interest */
(keep = &timevar atrisk censored
%do event = 1 %to &lastevnt; event&event
%end;
%do event = 0 %to &lastevnt; pk0_&event
%end;
%do event = 0 %to &lastevnt; error&event
%end;);
set &outset;
run;

%mend;
options mprint;
data myset;
attrib etime label = 'Event Time' evnt label = 'Event
Class';
input etime evnt @@;
cards;
1 1 1 2 1 5

```

```

Title 'Data Set';
proc print label data = myset; run;

Title 'Ordinary Kaplan-Meier Analysis';
%Greenwod(inset = myset, outset = kset, timevar =
etime, eventvar = evnt, lastevnt = 1)
proc print label data = kset; run;

Title 'Standard Actual Analysis';
%Greenwod(inset = myset, outset = actset, timevar =
etime, eventvar = evnt, lastevnt = 2)
proc print label data = actset; run;

Title 'Analysis with 4 terminal events';
%Greenwod(inset = myset, outset = fourset, timevar
= etime, eventvar = evnt, lastevnt = 4)
proc print label data = fourset; run;

```