IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Estimation of Transition Probabilities in Multi-State Survival Data

Haider Kareem Raheem[1] , Hayder Yahya Mohammed[2]

[1,2]AL- Karkh University of science

ARTICLE INFO

Article history:

Received	6 July 2025
Revised	7 July 2025
Accepted	18 July 2025
Available online	2 August 2025

Keywords:

Survival Data;
Multi – Marcov State model;
Nelson-Aalen Estimator;
Aalen-Johansen Estimator;
Kaplan-Maeier Estimator;
Multi-State Model; Hazard Function.

ABSTRACT

This study calculates nonparametric estimators including Kaplan-Meier, Nelson-Aalen and Aalen - Johnson for applying the Multi-State Model (MSM) to analyze type IIIness-death transitions between alive and death followed by different duct cancer types (Breast, Lung , Brain , Ovary) through reduction techniques by using diminutions (k+1)*(k+1) for matrix , The package software program was used to generate elementary Nelson-Aalen and Kaplan-Meier estimator, reach a conclusion that the two procedure are producing result that are very close..

1. Introduction

Studying multistate models requires answering two mainframe questions about what probability exists for random persons to occupy stage i at time t and what rate (hazard) measures persons leaving stage j for stage j^*.

These questions take two different forms in more formal contexts: first, determining the stage occupation probabilities and second evaluating the transition hazard for transitions from stage j to stage j^*.

Multinomial models are commonly observed in Right censoring affects biomedical data collections frequently.[26] Aalen and Johansen (1978) presented an estimator based on product limits for transition probability matrices from finite-state homogeneous Markov processes that extended Kaplan Meier single failure estimation and Aalen multiple decrement models.[2] The analysis of system components requires marginal information including transition hazards and stage probabilities which can be calculated through (Aalen, 1978) even for Markovian systems.[1][24] This paper presents survival data estimation through Aalan-Johansen and Nelson-Aalen using finite state homogeneous Markov chains and translate probability matrices that determine stage j^* hazard rates for individuals in stage j.

2. Multi-State Model (MSM)

The MSM serves as a continuous time stochastic model which enables people to transition between finite states. An individual moves from one state

Corresponding author E-mail address: dr.haider.yahya@kus.edu.iq

https://doi.org/10.62933/gn808t81

to another through time. Transition intensities serve to determine both future destinations and timing of state changes in the model according to its type in MSM. [19] :

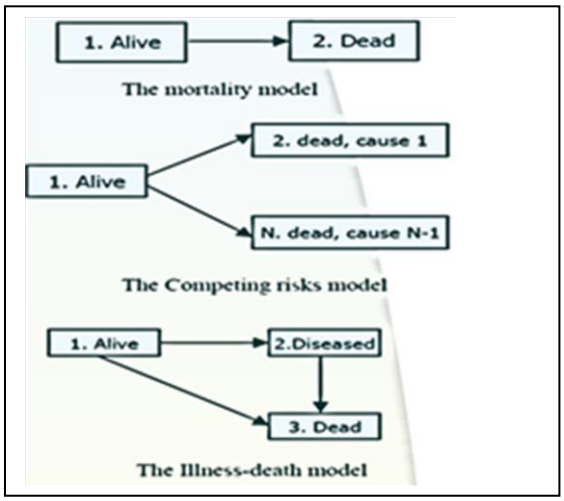

Figure 1 : Multi-State Model for Data

3. Survival Data

Survival analysis represents a statistical framework that includes concepts and methods and models to study time-dependent event occurrences in multiple study subjects through survival curves.. [25]

We start by examining survival data where we analyze how long The survival of a homogeneous population depends on two functions: hazard rate function $\psi(t)$ and cumulative hazard rate function

$$\chi(t) = \int_0^t \psi(s)\, ds \qquad (1)$$

Assumingly we have n samples.

Observations of survival durations for members of this population commonly end up right-censored because some participants leave the study before their survival times become known.[12][8]

The censoring process meets the independence requirement because additional knowledge about censoring events does not affect the failure risk at time t. Risk of failure remains unchanged for any time period t when no additional information about the lifetime duration becomes known.

We denote by $t_1 < t_2 < t_3 < t_4$ ….., Death observations occur at multiple times where d_j represents the number of individuals dying at t_j. [21]

4. Markov Chain

A Markov chain follows this description: The states form a set S={s1,s2,...,sr}. The process begins its journey from one of these states before moving between states one after another, Each move called a step. The chain adopts position si first before transition occurs to position sj using the transition probability pij no matter the earlier states The chain remains in its current condition with probability pii while an initial distribution on S determines the starting position. A particular state usually becomes the initial condition for starting the process when defining the system. The current state of the procedure dictates both previous and future occurrences to remain independent of each other. [9][6]

5. Multi - Markov State Model(MMSM)

Random individual movements fall under multi state modeling when people shift between limited states. A transition represents a state change since states that enable movement to different states are considered transient. The state structure identifies all possible states alongside specifying which state transitions are permissible. An absorbing state represents a terminal stage where no more state changes become possible. Death as a state represents an instance of absorbing state, transient state is not absorbing. [26]

The important targets

• estimation of the transition intensities αij (t)

- At time t the risk for transition i to j occurs instantaneously

• assess covariate effects

• The model needs to establish methods for measuring transition probabilitiesPi,j (s, t)

- A state j at time t becomes probable for a patient who starts in state i at time s < t

• prognostic forecasting. [9]

From figure 1 can be have two models as follow :

MODEL 1 : Competing risk

An investigation of death-specific mortality rate includes examination of individuals sharing exposure to multiple death causes : one transient state: "Alive" The final absorbing conditions include Death from cause c = 1 through k observations: time of death T > 0 and cause of

death D ∈ [1, . . ., k] Markovian: all transitions from "Alive" The transition hazard functions α0c (t) equals the corresponding cause specific hazard rates αc (t). [15][7]

$$\alpha c\,(t) = \lim_{h \to 0} \frac{Pr(t \le T < t + h, D = c \mid T \ge t)}{h} \qquad (2)$$

6. MODEL2 : iIIness-death

Study occurrence of the chronic diseases as well as death, The approach proves appropriate for conditions that become permanent beyond medical possibility of recovery. There is no possible route to return from state 1 (Diseased) to state 0 (Healthy). The model displays disease development patterns for research on cancer and similar studies. • Markovian if 1 → 2 transition probability independent on state 1 duration.[17][9]

7. Nelson-Aalen Estimators

The Nelson Aalen estimator serves as anonparametric estimator The method explores estimates of cumulative hazard rate functions extracted from survival data with censoring while avoiding distributional requirements. Its main purpose is model fit inspection through graphical analysis making it useful for determining parametric model suitability hence its initial development by Nelson (1972).[22]

The same estimator emerged independently from Nelson and Altshuler (1970) when studying animal experiments with competing risks.

Aalen (1978) later generalized its application through counting process methods while investigating about small and large sample properties using martingales, Nelson Aalen estimator serves as the current designation for this statistical method despite other names including Nelson estimator and Altshuler estimator and Nelson Aalen estimator and empirical cumulative hazard estimator. The section demonstrates diverse Nelson-Aalen estimator applications before moving on to an illustration and discussion of counting processes as a unified framework for these diverse situations. The most critical characteristics of the Nelson-Aalen estimator are subsequently listed. [22][21]

The Nelson- Aalen estimator of the cumulative hazard rate function becomes:

$$\hat{A}(t) = \sum_{t_j \le t} \frac{d_j}{r_j} \qquad (3)$$

The nelson estimator contains an instant jump ($\frac{d_j}{r_j}$) at each failure time t_j when the number at risk during the previous period is r_j. The Aalen estimator derives its shape from increasing right continuous step functions which receive ($\frac{d_j}{r_j}$) increases at failure time observations. [27]

8. Aalen-Johnsen Estimator

The medical research field extensively uses Aalen-Johansen estimator for calculating state probabilities in progressive illness-death Markov model applications. Survival data situations become feasible through the Markov process as long as the system has alive and dead states with distinct dead states. [16]

A Markov model approves the analysis of death competition between different causes of mortality. A Markov model known as the illness death model plays a vital role in bio statistical research since states are defined as ‘healthy’ diseased and ‘dead’ as one minus the Kaplan-Meier estimator for general cases. The product integral method of Aalen and Johansen demonstrates how the estimator commonly referred to as the Aalen-Johansen estimator functions [2], The matrix version of Kaplan-Meier estimator appears below in two chronic disease settings: competing risks model and the Markov illness-death model. We show explicit forms of the Aalen-Johansen estimator elements in these simple scenarios ,The general Aalen-Johansen estimator follows with an explanation of its derivation through product integration of Nelson-Aalen cumulative transition estimators.[1]

The method of intensities operates to extract information about its statistical attributes while we briefly show its use in such applications, we let r_{ig}be the number of individuals in state g just prior to time t_j Then the Aalen Johansen estimator takes the form

$$\hat{P}(s,t) = \prod_{s<t_j\leq t} \left[I + \hat{\psi}_j\right] \dots\dots (4)$$

Here I is the (k+1)*(k+1) identity matrix , $\hat{\psi}_j$ is the (k + 1)*(k + 1) matrix with entry (g , h), equal to:

$\hat{\psi}_{ghj} = \frac{dgh_j}{r_{ghj}}$ for $g \neq h$ and entry (g , g), equal to:

$\hat{\psi}_{ggj} = -\frac{dg_j}{r_{gj}}$, and the matrix product taken in

The order of increasing t_j, The explicit mathematical solution exists for basic models including competing risks models and illness-death models we analyzed previously.

expressions for the elements of (4) A direct implementation of (4) exists through matrix multiplications enabled software systems although this solution remains out of reach in general situations.[4] [5][18].

9. Kaplan-Meier Estimator

The Kaplan Meier estimator provides a nonparametric method for estimating survival distributions from censored data by employing the classical actuarial estimator as its limiting expression. Today the estimator bears the authors' names but sometimes people call it the product limit estimator. This paper explains Kaplan-Meier estimation then demonstrates application with sample data to show median and mean survival calculations,[13] The paper demonstrates how the Kaplan-Meier estimate can be expressed as a product integral of the Nelson-Aalen estimator and explains its statistical properties. Since almost four decades the Kaplan Meier estimator has become the main instrument to evaluate censored survival data while all books on survival analysis mention it. Fleming and Harrington present all technical derivations about this estimator's statistical properties. The Kaplan Meier method generates the survival distribution function through this expression:

$$\hat{S}(t) = \sum_{t_j\leq t} \left(1 - \frac{d_j}{r_j}\right) \qquad (5)$$

The number of individuals alive without censoring at t_j equals r_j and the number of deaths occurring at this time is d_j. A number of death at t_j.[5][25][27]

Application With Censed Data

In this section, we compute Nelson-Aalen and Kaplan-Maier estimator for censored data at four

Disuse cancer types (i.e breast ,lung ,brain ,ovary), table blew represent censored data :

ALIVE				
Years	Breast	Lung	Brain	Ovary
2011	1284	370	33	96
2012	1325	355	312	96
2013	1316	365	281	94
DEATH				
Years	Breast	Lung	Brain	Ovary
2011	7	1	1	1
2012	6	1	2	1
2013	8	5	1	2

Then compute Aalen-Johnsen estimator at both Nelson-Aalen and Kaplan-Maier represent as $\hat{\psi}_j$ is the (k + 1)*(k + 1) matrix by using R

10. Main Result

KAPLAN-MAEIR ESTMATOR				
Years	Breast	Lung	Brain	Ovary
2011	0.001783	0.000917	0.003496	0.001076
2012	0.003312	0.002752	0.010489	0.002152
2013	0.004840	0.006422	0.017482	0.004305
NELSON-AALEN ESTMATOR				
Years	Breast	Lung	Brain	Ovary
2011	0.001783	0.002700	0.006197	0.007273
2012	0.001528	0.002446	0.009439	0.010515
2013	0.002038	0.006625	0.010121	0.012274

AALEN-JOHNSEN ESTMATOR WITH KAPLAN-MAEIR

$$\begin{pmatrix} 0.009906 & 0 & 0 & 0 \\ 0 & 0.010066 & 0 & 0 \\ 0 & 0 & 0.031188 & 0 \\ 0 & 0 & 0 & 0.007519 \end{pmatrix}$$

AALEN-JOHNSEN ESTMATOR WITH NELSON-AALEN

$$\begin{pmatrix} 0.005341 & 0 & 0 & 0 \\ 0 & 0.011732 & 0 & 0 \\ 0 & 0 & 0.025543 & 0 \\ 0 & 0 & 0 & 0.029771 \end{pmatrix}$$

11. CONCLUSION

The Hazard rate of AALEN-JOHNSEN ESTMATOR WITH KAPLAN-MAEIR are little bit higher than that of AALEN-JOHNSEN ESTMATOR WITH NELSON-AALEN , Hence it may be conclude that both procedure are yielding almost same result.

REFRENCES

[1] Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, *6*(3), 534-545.

[2] Aalen, O. & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 141-150.

[3] Andersen.P.K,Borgan Gill R. D. and Keiding (1993)" Statistical Models Based on Counting Processes" Springer Verlag, New York.

[4] Balboa, V., & De Uña Alvarez, J. (2018). Estimation of transition probabilities for the illness-death model: Package tp. idm. *Journal of Statistical Software*.

Biometrics,34,571-580.

[5] Borgan, Ø. (1997). *Three Contributions to the Encyclopedia of Biostatistics: The Nelson-Aalen, Kaplan-Meier, and Aaalen-Johansen Estimators*. Department of Mathematics, University of Oslo.

[6] Fleming, T. R. (1978a). Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *The Annals of Statistics*, 1057-1070.

[7] Fleming, T. R. (1978b). Asymptotic distribution results in competing risks estimation. *The Annals of Statistics*, 1071-1079.

[8] Fleming.T.R.and Harrington. D. P.(1991) "Counting Processes and Survival Analysis",Wiley ,New York.

[9] Heggland, T. (2015). *Estimating transition probabilities for the illness-death model* (Master's thesis).

[10] Hornung . R. and Meinhardt .T (1987) "Quantitative risk assessment of lung cancer" in (US) uranium Miners Health Physics,52,417-30.

[11] Johansen. S. (1978) "The product limit estimator as maximum likelihood estimator" Scandinavian Journal of Statistics,5,195-199.

[12] Kalbfeisch,J.D.and Prentice. R. L.(1980) "The Statistical Analysis of Failure Time Data ",Wiley,New York.

[13] Kaplan. E. L. and Meier.P (1958) "Non_parametric estimation from incomplete observations" Journal of the American Statistical Associatio.53,451-481.

[14]Keiding.N.andAndersen.P.K.(1989)"Nonparametrc estimation of transition intensities and transition probabilities a case study of a two state Markov process",Applied Statistics,38.

[15] Llopis-Cardona, F., Armero, C., & Sanfélix-Gimeno, G. (2021). Reflection on modern methods: competing risks versus multi-state models. *arXiv preprint arXiv:2104.03671*.

[16] Ma, Y. (2022). *Nonparametric Estimation of Transition Probabilities in Illness-Death Model Based on Ranked Set Sampling* (Doctoral dissertation, University of South Florida).

[17] Machado, L. F. M., & Roca-Pardiñas, J. (2011). p3state. msm: Analyzing survival data from an illness-death model. *Journal of Statistical Software*, *38*, 1-18.

[18] Maltzahn, N., Hoff, R., Aalen, O. O., Mehlum, I. S., Putter, H., & Gran, J. M. (2021). A hybrid landmark Aalen-Johansen estimator for transition probabilities in partially non-Markov multi-state models. *Lifetime Data Analysis*, *27*, 737-760.

[19] Manevski, D., Putter, H., Pohar Perme, M., Bonneville, E. F., Schetelig, J., & de Wreede, L. C. (2022). Integrating relative survival in multi-state models—a non-parametric approach. *Statistical Methods in Medical Research*, *31*(6), 997-1012.

[20] Nair.V. N (1984) "Confidence bands for survival functions with censored data,A comparativestudyTechnometrics",26,265-275.

[21] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, *14*(4), 945-966.

[22] Nelson. W. (1969)"Hazard plotting for incomplete failure data" Journal of Quality Technology,1,27-52.

[23] Nießl, A., Allignol, A., Beyersmann, J., & Mueller, C. (2023). Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*.

[24] Nyinawajambo, M. R. (2018). Survival Analysis of Time to Event Data: An application to child mortality in Sub-Saharan Africa Region using Their Demographic and Health Surveys.

[25]Schliching.P.Christensen.E.Andersen.P.K.Fauerholdt.L. Juhl.E.Poulsen.H.and Tygstrup .N (1983)"for The Copenhagen Study Group for Liver Diseases Prognostic factors in cirrhosis identied by Cox;s regression model " Hepatology ,3,889-893.

[26] Snijders, M. (2017). *Prediction for Transition Probabilities in Multi-State Models* (Bachelor's thesis).

[27] Soutinho, G. U& Meira-Machado, L. U (2020). Estimation of the transition probabilities in multi-state survival data: New developments and practical recommendations. *WSEAS Transac Math. 2020; 19: 353-366*.