ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 5, Issue 2, pp: (734-739), Month: April - June 2017, Available at: www.researchpublish.com

Generation of Mean Transition (Sojourn) Time in Transient States of HIV Disease Progression in Kenya

Stephen Kinyanjui Mwambura[1], Prof. Samuel Karoki[2]

Department of Mathematics and Computer Science, Catholic University of Eastern Africa

Abstract: **Many studies have been carried out in this field and different models have been proposed with a view to better understanding of this disease. State transition models were monitored by serial measurements of CD4 cells per unit volume (mm^3) of blood in an HIV patient as a tool for modeling HIV disease progression. WHO disease staging system for HIV infection was used to analyze data. The methods used in this dissertation are applied to HIV data sourced from one health facility. HIV progression is analyzed through the application of a four state Markov model with reversible transitions such that state 1: CD4 count $\geq$ 500, state 2: 350$\leq$ CD4 count $<$499, state 3: 200$\leq$ CD4 count $<$349, state 4: CD4 count $<$ 200.**

Keywords: **Mean Transition, Transient state, Progression.**

1. INTRODUCTION

The staged Markov model is a useful way of describing a process in which an individual moves through a series of states in continuous time. Survival analysis is the simplest two-state model where individuals remain alive until an observed or censored time of death. Staged models based on Markov processes are a well-established method of estimating rates of transition between stages of disease. Staged models are a sub-discipline of survival analysis. They are the most common models for describing longitudinal failure time data. Staged models are models with definitive states that an individual in a study may visit. The most convenient model in staged models is a Markov model. Research has reported the application and usefulness of Markov chains in a wide range of topics such as medicine, game theory, internet applications, social sciences, and statistics among others. One may examine the occurrence of clinical events which occur early in HIV infection or the values of biological markers as markers of disease progression. The most widely used marker of HIV progression is the CD4 lymphocyte count, which plays a crucial role in the immune system and is used to determine when ART should commence. When an individual loses CD4 cells he or she is more vulnerable to opportunistic infections. In this project, a model with state structure based on intervals of CD4 count and death as an absorbing state was used.

2. STATEMENT OF THE PROBLEM

Longini *et al* (1989) modeled stages of HIV infection with irreversible progression. This model examined the mean sojourn times in each of the states and the length of the AIDS incubation period but did not take into account the effect of cofactors on rates of HIV progression. Tarylee (2011) applied multistate Markov model to HIV progression using CD4 count intervals with ARV initiation as an absorbing state. This study has analyzed HIV progression using CD4 counts intervals of six months since enrolment on ART with the objective of investigating the probabilities of transitions to lower CD4 counts and estimating the average stay in the CD4 count states. Previous researchers in Kenya have mainly applied Markov processes to manpower systems. Because of this a study of the survival of HIV positive patients in Kenya is needed to understand the disease progression.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 5, Issue 2, pp: (734-739), Month: April - June 2017, Available at: www.researchpublish.com

3. LITERATURE REVIEW

The desirability of a descriptive tool for censored survival data, free of parametric assumptions, had been recognized for decades. During the 1950s, well established demographic and actuarial techniques were presented to the medical-statistical community in influential surveys such as those by Berkson and Gage (1952). In this approach, time was grouped into discrete units for example one year intervals and the chain of survival frequencies from one interval to the next were multiplied together to form an estimate of the survival probability across several time periods. The difficulty was in the development of the necessary approximations due to the discrete grouping of the intrinsically continuous time and somewhat oblique observation fields in cohort studies and more complicated demographic situations. The penetrating study of Kaplan and Meier (1958), the fascinating genesis of which was chronicled by Breslow et al (1991), eliminated in principle the need for these approximations in the common situations in medical statistics where all survival and censoring times are known precisely. Actuaries and demographers have used parametric survival models for decades but these have never dominated the medical uses of survival analysis. However, in early period, important contributions to the statistical theory of survival analysis were based on simple parametric methods.

Cox (1972) revolutionized survival analysis by his semi-parametric regression model for the hazard, depending arbitrarily non-parametrically on time and parametrically on covariates. The homogenous Markov staged models are time-continuous models for which the transition probability only depends upon the current state and has constant transition hazards for the state. In medical research, discrete observation times are generally used. The state that the patient is in at the observation is the only thing known. The researcher may know that there has been a transition to a new state but does not know when in that interval or time it occurred. Thus we consider homogeneous Markov models with interval censoring. The Markov model is a type of stochastic process and has long played a role in bio statistical modeling. A more recent example of major importance is the Armitage-Doll model for the development of cancer (i.e. carcinogenesis), which appeared in 1954 (Armitage-Doll, 1954). The so-called multistage model is a Markov chain which described how a cell moved through a number of different stages before becoming cancerous. The model has been a considerable inspiration for understanding the development of cancer. We consider a population of individuals, each having a failure time T, where $T \geq 0$ is a random variable of interest. A vital function of T referred to as the survivor function is given by; n$S(t) = P(T > t) = 1 - Pr(T \leqslant t) = 1 - F(t)$where $F(t)$ is the cumulative distribution function of the random variable T.The lifetime distribution function, conventionally denoted by F, is defined as the complement of the survival function,$F(t) = Pr(T \leqslant t) = 1 - S(t)$and the derivative of "$F$", which is the density function of the lifetime distribution, is conventionally denoted f,

$$f(t) = F'(t) = \frac{d}{dt}F(t)$$

The function f is sometimes called the event density; it is the rate of death or failure events per unit time.The survival function is often defined in terms of distribution and density functions

$$S(t) = Pr(T>t) = \int_t^{\infty} f(u)du = 1 - F(t)$$

Thus the survival function gives us the probability that the individual does not fail (i.e. survives) within the time interval (0, t). Similarly, a survival event density function can be defined as

$$s(t) = S'(t) = \frac{d}{dt}S(t) = \frac{d}{dt}\int_t^{\infty} f(u)du = \frac{d}{dt}[1 - F(t)] = -f'(t)$$

4. RESEARCH METHODOLOGY

For a continuous time stochastic process {X (t), t ≥0} whose state space is S, we say it has the Markov property if

$$P(X(t) = j | X(s) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1) = P(X(t) = j | X(s) = i)$$ where

$0 \leq t_1 \leq \ldots \leq t_{n-1} \leq s \leq t$ is any nondecreasing sequence of $n + 1$ state occupation times and $i_1 \ldots, i_{n-1}, i, j \in S$.

In other words the state of the process at time t depends only on the most recent state occupied prior to time t.A continuous time stochastic process {X (t), t ≥ 0} is called a continuous time Markov process (hereafter CTMC) if it has the Markov property.

A CTMC is said to be time homogeneous if for any s ≤t and any states $i, j \in$ S

$P(X(t) = j | X(s) = i) = P(X(t - s) = j | X(0) = i)$

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 5, Issue 2, pp: (734-739), Month: April - June 2017, Available at: www.researchpublish.com

That is, dependence on time is only through the length of time elapsed between events. The time homogeneous property means that whenever state *i* is entered, the way the process evolves is equivalent to having started in state *i*at time 0.When the process enters state *i*, the time it spends there before moving to another state is called the holding time in state *i* or the sojourn time. The sojourn time is of great interest in disease modeling as it gives us an indication of how rapidly the disease is progressing. Longer sojourn times in a disease state mean a slow progressing disease and shorter sojourn times mean a rapidly progressing disease. With the assumption of time homogeneity, it should be evident that the sojourn time in state *i* would be the same every time state *i* is entered. Hence we can speak of a holding time or sojourn time distribution. For a time homogeneous continuous time Markov chain, T_i(the sojourn time in state *i*) is exponentially distributed. The theorem has been adapted from Random Processes, Statistic (2007) notes. The proof is based on the memoryless property which is unique for the exponential distribution. By time homogeneity we assume that the process starts in state *i*. Then,

$$\begin{aligned} P(T_i > s+t | T_i > s) &= P(X(u) = i \; for \; 0 \le u \le s+t | X(u) = i \; for \; 0 \le u \le s \\ &= P(X(u) = i \; for \; s \le u \le s+t | X(u) = i \; for \; 0 \le u \le s \\ &= P(X(u) = i \; for \; s \le u \le s+t | X(s) = i \;) \\ &= P(X(u) = i \; for \; 0 \le u \le t | X(0) = i \;) \\ &= P(T_i > t) \end{aligned}$$

This is the unique memoryless property for an exponentially distributed random variable, therefore T_i must be exponentially distributed with corresponding state *i* mean sojourn time given by μ_i . This implies that $Var(T_i) = {\mu_i}^2$ is not independent of the mean.

5. FINDINGS

The transition probability matrix $P(t)$ is evaluated below at times 1, 2, 3, 4, 5, 6, and 7 years. As time progressed the probability for transitions to the absorbing state increased and the probability of immune recovery decreases.

$$P(1) = \begin{bmatrix} 0.5114 & 0.3032 & 0.171 & 0.01434 \\ 0.3146 & 0.3494 & 0.305 & 0.03098 \\ 0.1945 & 0.3343 & 0.4075 & 0.06373 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P(2) = \begin{bmatrix} 0.3902 & 0.3182 & 0.2496 & 0.04196 \\ 0.3301 & 0.3194 & 0.2847 & 0.06576 \\ 0.2839 & 0.312 & 0.3013 & 0.1029 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P(3) = \begin{bmatrix} 0.3492 & 0.3129 & 0.2655 & 0.07333 \\ 0.3247 & 0.3069 & 0.2699 & 0.09854 \\ 0.3019 & 0.2958 & 0.2665 & 0.1358 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 5, Issue 2, pp: (734-739), Month: April - June 2017, Available at: www.researchpublish.com

$$P(4) = \begin{bmatrix} 0.3282 & 0.3037 & 0.2632 & 0.1049 \\ 0.3151 & 0.2959 & 0.2591 & 0.1299 \\ 0.2993 & 0.284 & 0.2505 & 0.1663 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P(5) = \begin{bmatrix} 0.3146 & 0.2936 & 0.256 & 0.1358 \\ 0.3046 & 0.2855 & 0.2497 & 0.1601 \\ 0.2911 & 0.2737 & 0.2399 & 0.1953 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P(6) = \begin{bmatrix} 0.303 & 0.2835 & 0.2477 & 0.1658 \\ 0.2942 & 0.2756 & 0.241 & 0.1892 \\ 0.2816 & 0.2641 & 0.231 & 0.2233 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P(7) = \begin{bmatrix} 0.2923 & 0.2737 & 0.2392 & 0.1947 \\ 0.284 & 0.266 & 0.2326 & 0.2173 \\ 0.272 & 0.2549 & 0.229 & 0.2502 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Interpretation of the transition probability matrix:

A patient presenting at the health facility with a CD4 count greater than 500, has a 25% chance of having a CD4 count between 200 and 349 and 4% chance of being absorbed within 2 years. Seven years later, however this patient has a much higher probability of being absorbed (20%); A patient presenting at the health clinic with a CD4 count between 350 and 499 has a 32% chance of experiencing immune recovery (an increase in CD4 count) within the first year. As time progresses, this chance of immune recovery decreases to 28% after seven years; Transition probabilities from states 1, 2, 3 into state 4 (absorbing state) increases as time increases. For example

$p_{24}(1) = 0.03098$ while $p_{24}(7) = 0.2173$

$p_{34}(1) = 0.06373$ while $p_{34}(7) = 0.2502$

A patient presenting with $200 \leqslant$ CD4 count<349 has a 25% chance of being absorbed within seven years.

6. CONCLUSION AND RECOMMENDATION

Based on the findings of the study the following conclusions were made at 95% confidence limits:The study reveals that a patient presenting with $200 \leqslant$ CD4 count<349 has a 25% chance (0.1966, 0.4042) of being absorbed within seven years. This model reveals that CD4 cell count is a good indicator for gauging the strength of the immune system and for

determining whether a person is at risk of infection with certain organisms. The higher the CD4 count, the stronger the immune system. Nationally, investigative efforts are needed to ascertain why patients pin emotions and hopes to a single laboratory test of the CD4 cells count for proper monitoring.

REFERENCES

[1] Aalen,O.O.&Johansen,S.(1978).An *empirical transition matrix form on homogeneous Markovchains based on censored observations*.Scandinavian Journal of Statistics 5, 141–150.

[2] Armitage, P. and Doll, R. (1954). *The age distribution of cancer and a multistage theory ofcarcinogenesis*. Br. J. Cancer, 8, 1-15. United States of America.

[3] Beck, J.R., Pauker, S.G (1983). *The Markov process in medical prognosis*. Med Decision Making 3,419-458.

[4] Berkson and Gage (1952). *The Simplest cure rate model.* Wikipedia. The Free Encyclopedia. Binquet, C .,Rey,D.,Mahboubi,A. (2009). *Markov modelling of HIV infection evolution in Highly*

[5] *Active Antiretroviral Therapy.* France.

[6] Breslow K. & McLaren (1991). *Thick Malignant Melanomas with good clinical outcome: ahistological study and survival analysis*. East of Scotland.

[7] Bwayo, J., Nagelkerke, N., Moses, S., Embree, J., Ngugi, E., Mwatha, A., Kimani, J., Anzala, A., Choudhri, S., and Achola, J. (1995). *Comparison of the declines in CD4 counts in HIV-1-seropositive female sex workers and women from the general population in Nairobi, Kenya*. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 10,457–461.

[8] Christopher Jackson (2011). *Multi-state modeling with R: the msm package*. Journal of Statistical Software Volume 38, Issue 8, 1-25. Medical Research Council Biostatistics Unit, Cambridge, U.K.

[9] Cox, R.D & Miller, H.D (1965). *The Theory of Stochastic Processes*. Chapman and Hall, London. Cox,D.R.(1972). *Regression models and life-tables* (with discussion). Journal of the Royal

[10] Statistical Society Series B 34, 187-220.

[11] Ira M. Longini, Jr, Michael G.Hudgens (2003). *Lecture notes on stochastic processes inbiostatistics: Applications to infectious diseases*. Pp 31-61. Atlanta

[12] Jensen, P.A., Bard, J.F., 2002b. *Sensitivity analysis, duality and interior point methods. In:Operations Research Models and Methods*, John Wiley & Sons, Inc, Danvers, MApp.111-145

[13] Kaplan, E.L & Meier, P. (1958). *Nonparametric estimation from incomplete observations.* Journal of the American Statistical Association 53(282), 457-481.

[14] Kay, R. (1986). *A Markov model for analyzing Cancer Markers and Disease States in SurvivalStudies*. Biometrics 42:4, 855-865.

[15] Longini, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., George F.L, and Herbert W.H. (1989). *Statistical analysis of the stages of HIV infection using a Markov model.Statistics in Medicine* 8,831–843.

[16] Longini, I. M., Clark, W. S., Gardner, L. I., and Brundage, J. F. (1991). *The Dynamics of CD4+ T-Lymphocyte decline in HIV-Infected Individuals: A Markov Modeling Approach*. Journalof Acquired Immune Deficiency Syndromes 4, 1141–1147.

[17] Meeker, W.Q., Escobar, L.A. (1998b). *Models for continuous failure time processes. In StatisticalMethods for Reliability Data*, John Wiley and Sons, Inc, New York pp.26-45.

[18] Mullins, D. C. (1996). *A Simplified Approach to Teaching Markov Models*. American Journal of pharmaceutical Education 60, 42–47.

[19] Redfield, R.R., Wright D. and Tramont,E(1986). The *Walter Reed staging classification for HTLV-III/LAV infection*. New England Journal of Medicine, 314, 131-132

[20] Simpson K.N, Baran R.W (2011). *Markov models to assess the cost-effectiveness of switching to asecond-line protease inhibitor regimen of lopinavir plus ritonavir for HIV-infected patients who failed a first-line treatment in sub-Saharan Africa.* University of New York.

[21] Simwa, R.O. &MugishaJ .T,(2003c). A *stochastic model for the CD4 cells counts in an HIV/AIDSpatient and its application in treatment interventions*. In: Proceedings of the 1st Pan-African Biomathematics Congress 8th-12th, December, 2003, Kampala, Uganda.

[22] Simwa, R.O.,(2003a). *On empirical modeling of HIV/AIDS pandemic with application to EastAfrica*. African J.Sci. Technol. 4, 34-43.

[23] Statistics (2007). *Random processes*. University of KwaZulu-Natal, Durban, South Africa.

[24] Tarylee, R. (2011). *The Application of Multistate Markov Models to HIV Disease Progression*, University of KwaZulu-Natal,MSc thesis, South Africa.

[25] Titman,A.C (2007). *Model diagnostics in multi-state models of biological systems.* PhD thesis, University of Cambridge.

[26] World Health Organization (2005). *Interim Clinical Staging of HIV/AIDS& HIV/AIDS case definitions for surveillance*, African Region.