



# A Critical Review of Multistate Bayesian Models Modelling Approaches in Monitoring Disease Progression: Use of Kolmogorov-Chapman Forward Equations with in WinBUGS

Zvifadzo Matsena Zingoni<sup>\*1,2</sup>, Tobias F. Chirwa<sup>1</sup>, Jim Todd<sup>3</sup>, Eustasius Musenge<sup>1</sup>

<sup>1</sup>Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand and, Johannesburg, South Africa

<sup>2</sup>National Institute of Health Research, Causeway, Harare, Zimbabwe

<sup>3</sup>Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

**\*Corresponding Author:** Zvifadzo Matsena Zingoni, Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand and, Johannesburg, South Africa National Institute of Health Research, Causeway, Harare, Zimbabwe. **Email:** [zmatsena28@gmail.com](mailto:zmatsena28@gmail.com)

**Abstract:** There are numerous fields of science in which multistate models are used, including biomedical research and health economics. In biomedical studies, these stochastic continuous-time models are used to describe the time to event life history of an individual through a flexible framework for longitudinal data which can describe more than one possible time to event outcomes for a single individual. The standard estimation quantities in multistate models are transition probabilities and transition rates which can be mapped through the Kolmogorov-Chapman forward equations. Most multistate models assume the Markov property and time homogeneity; however, if these assumptions are violated an extension to non-Markovian and time-varying transition rates is possible. This manuscript extends reviews in various types of multistate models, assumptions, methods of estimations, data types and emerging software for fitting multistate models. We highlight strengths and limitations in multistate models for different software and emphasis is made on Multistate Bayesian models in Bayes X and Win BUGS software which are underutilized. A partially observed and aggregated dataset from the Zimbabwe national ART program is used to illustrate the use of Kolmogorov-Chapman forward equations in estimating transition rates from a three-state reversible multistate model based on viral load measurements in Win BUGS.

**Keywords:** aggregated data, Bayesian MCMC perspective, Kolmogorov-Chapman forward equations, multistate models, partially observed data, Win BUGS

## 1. INTRODUCTION

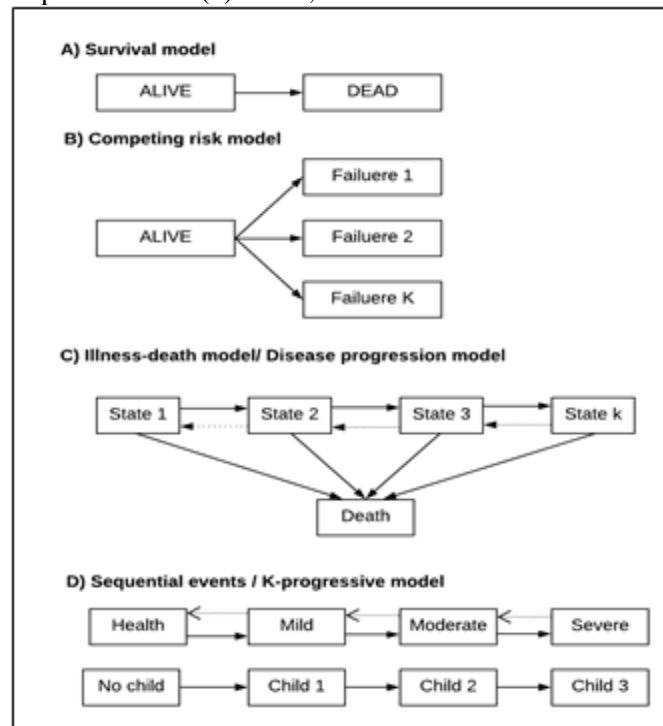
In longitudinal studies, participants are followed-up, and patient's information is collected repeatedly. In such studies, individuals may experience multiple events, and longitudinal failure time data is captured. In such studies, the most commonly used approaches are the deterministic and stochastic models (1). Deterministic models assume that response variables are deterministic functions of time disregarding the randomness of the risk factors completely while stochastic models assume that the response variables are random functions of time with probabilities of moving from one state to another (2). Stochastic models are more realistic than deterministic models since nature is predominantly random; however, statistical models with a stochastic approach are

more complex and challenging than those involving deterministic approach (1,3).

Multistate models are defined as a continuous-time stochastic process which allows participants to move among a finite discrete number of compartments or states which could be clinical symptoms, biological markers, disease stages or disease recurrence in biomedical researches (4,5). Movement from one state to another is called a transition (event has occurred); states can be transient (if a transition can emerge from the state) or absorbing (if no transition can emerge from the state). Movement between transitions can be reversible or irreversible, and these movements contribute to the intricacy of the multistate model in addition to the number of states defined. The transition intensities (hazard rates)

provide the transition specific hazards for movement from one discernable state to another. These transition intensity functions can also be used to compute the mean sojourn time (the average time spent in a single state before death), total length stay in a state (total time spent in a state before making a transitions), the number of transitions made from start to end of study and the transition probabilities (6). Also,

the effect of covariates on each transition can be assessed (7); however, the effects of the covariates on the different transitions may not be the same since the severity of the disease progression differs by each intermediate state. There are different types of multistate models which can be used to answer a different research question, Figure 1.



**Figure1:** Schematic illustration of different types of multistate models

The *mortality model* for survival analysis with only two states and one transition from “alive” state to “dead” state is the simplest multistate model, Figure 1A. These mortality models are useful, mostly in answering etiological research questions (8). The hazard rates are usually estimated using a semi-parametric approach which has a less stringent assumption (9). The hazard function is assumed to be an arbitrary, unspecified, non-negative function of time (10). The incidence or hazard rate is estimated by assuming independence of survival times between distinct individuals in a sample and a constant hazard ratio regardless of time (9).

Another type of multistate model is the *competing risks model* which extends the mortality model depicting a scenario whereby an individual may experience one of the several failure outcomes (11,12), Figure 1B. In such model, competing risk analysis whereby the interest is in the occurrence of primary outcome but have other contesting events which may preclude the occurrence of the primary outcome

or significantly alter the chances of observing the primary outcome; or situations where the different types of events may be relevant, but the analysis focuses on both time and occurrence of the first event (13). The reason why the competing risk analysis is considered to be appropriate over the Kaplan-Meier estimation in such situations described above is that the Kaplan-Meier estimation treats the competing events as censored observations which bring in bias since the independence assumption is violated. The baseline hazard may differ between these competing events (14). The competing risk model also provides an in-depth insight on the effect of interventions on separate outcomes observed are also useful in exploring the relationship between explanatory covariates and the absolute risk which is critical particularly in decision-making and prognostic research work (8).

Partitioning the “alive” state of the mortality model into two or more transient (intermediate) states yields another type of a multistate model

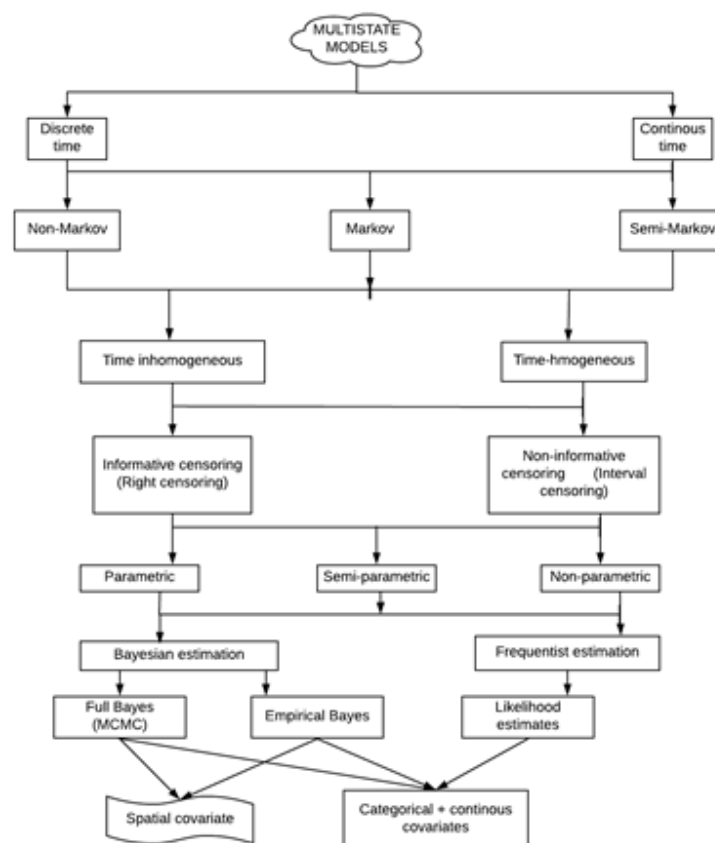
known as the *disease progressive multistate model* of which the simplest is the three-state model (5), Figure 1C. In biomedical research, *illness-death models* or *disability models* which are a special type of a disease progression model, are usually used in estimating disease incidences and the mortality transition intensities (15). The disability model is considered in irreversible models when the disease increases the risk of death. In scenarios whereby the absorbing state is not considered, the *K-progressive models* which follow a sequential process (15) for instance health, mild, moderate and severe sequence or the *fertility model* which is used to describe the reproductive life history of a woman, Figure 1D, where each state is defined by the number of children born are commonly used. Application of multistate models is not limited to biomedical studies like the evaluation of disease progression patterns (16–18) but cuts across various life history data, including health economics. In health economics studies inclined to the monitoring of disease progression, issues on the cost-effectiveness of prevention strategies(19), treatment (20), and diagnosis intervention (21) to inform policy decision-making process(22) can be addressed using multistate models. There is an extensive review of multistate

models in the literature. However, most review papers on multistate models have focused on the frequentist or maximum likelihood estimation (MLE) approach within the multistate model framework (4,5,7). None of these reviews has discussed Bayesian estimation (BE) within the multistate models, which is equally a robust method in multistate statistical modelling. Therefore, this article aims to extend previous reviews on multistate models with primary emphasis on Bayesian inference in multistate models. The rest of the manuscript is structured as follows: Section 2 will highlight the different assumptions within multistate models’ framework, multistate models data features and contrast between MLE and BE methods with their associated software packages; Section 3 will provide a detailed illustration of using Kolmogorov-Chapman forward equation on viral load aggregated data; Section 4 is left for discussion and conclusion. The appendix section provides details information for Section 3 and the WinBUGS code used.

## 2. MULTISTATE MODELS

### 2.1. Assumptions

A flow chart for multistate model assumptions, methods of estimation, possible censoring patterns and covariates types is shown in Figure 2.



**Figure2:** The multistate model assumptions, estimation types and possible covariates flow chart

Multistate models can either be fitted assuming *discrete-time or continuous-time*. In discrete-time models, the movements between states occur at a fixed time, and the transition probabilities are usually reported while in the stochastic continuous-time models the transitions can occur at any time point, and transition rates are standard estimates (23). In biomedical studies, the continuous-time model reflects reality since the transition occurs at random.

Different assumptions can be made on multistate models about the dependency of the hazard rates (transition intensities) on time. The *Markov property* assumes that the transition to a future state is only dependent on the present state occupied not the ones before; hence, the model has a “memory loss”. This type of model is usually used because of its simplicity (24). Alternatively, multistate models can assume a semi-Markov process meaning that the next future transition depends on both the currently occupied state and also the time of entry into the current state. The *semi-Markov model* is considered flexible in most cases; however, there are some drawbacks in using this model. Firstly, the semi-Markov model contains many parameters which make the model much more problematic to fit and the distribution of the sojourn times in each state is a requirement which in most instances might be unavailable (7,18,25,26). Lastly, multistate models can assume a non-Markovian process. This model is dependent arbitrarily on the previously occupied states; hence, there is no “memory loss” in the model. The implementation of *non-Markovian models* has been challenging until the introduction of the “Markov-free” estimators for transition probabilities in the last decade (4).

Another assumption normally made in the multistate model is that of *time homogeneity*. In a time-homogeneous model, the transition intensities are assumed to be constant over time; that is, the rates are independent of time (4,25,27). In such models, the Kolmogorov differential equations can be solved explicitly using the decomposition of the transition matrix into both eigenvalues and eigenvectors (23). Models which assume time homogeneity used more often possibly due to well-developed software at disposal and their less intimidating theoretical framework. However, if time homogeneity assumption is violated, an inhomogeneous time model is used which

assume that the transition intensities change with time (27).

## **2.2. Data Features**

Multistate models can be characterized by the way the data has been captured in a research process. Censoring is a crucial feature in time to event data analysis (28). In observational studies, follow-up studies often end before the outcome occurs leading to right-censoring of observation times (29) while left-censoring occurs when the study begins after the event has occurred but the event times are unknown (28). Frequently, non-informative censoring occurs when participants are follow-up intermittently such that the period between visits is missing. This means that the transition times are not precisely observed, and the states occupied between follow-up time points are unknown. These non-informative censoring observations are considered to be interval-censored (28). The mechanisms which give rise to censoring are essential in statistical inference within the multistate framework, and these data features need to be taken into account during analysis to avoid getting biased estimates since their likelihood functions will be different (30).

Moreover, intermittent follow-up of participants leads to incomplete spaced data points. However, most studies have placed focus on fully observed complete case data. In instances where data is missing (missing at random (MAT) or missing not a random (MNAR)), the multistate model framework allows one to use the likelihood-based method for missing covariates assuming a continuous-time Markov multistate model (31). However, if any covariates are missing, convergence problems are more likely with this method.

Multistate models can also be implemented in scenarios where the disease history of the individual is incomplete. This arises when participants are observed intermittently for a short time, not to completion of their disease history. In such follow-ups, other visits are missed, and the specific time of occurrence of an event is unknown. Work by Kalbfleisch and Lawless equips one to fit a time-homogeneous Markov model with arbitrary transitions structure for such incomplete history data (31).

Besides, within a multistate model, the follow-up data might be partially observed, that is, only the initial state and the final state information is known but the intermediate experiences are



unknown. This happens typically in program data was that the data is usually reported and summarized in an aggregated format. Nonetheless, this data can be used within a multistate framework using the method proposed by Welton (2005) on handling partially observed aggregated data to estimate transition rates (32).

### **2.3. Statistical Inferences and Software**

#### *2.3.1. Frequentist (Maximum Likelihood Estimation) Perspective*

The frequentist approach has been well documented in the literature (4,5). This method strongly relies on the dataset for parameter estimation. In these models, the statistical inference and estimation of the transition rates are based on MLE, and some detailed theoretical steps behind this approach have been provided in Appendix A of the supplementary material. There are various software and packages which can be used to handle MLE multistate models. The commonly used free software and package is R (33) *msm* package (34). The *msm* package can fit continuous-time Markov models (homogeneous time or non-homogeneous time using the piecewise constant models) and hidden Markov multistate models with misclassification error (6). Within the *msm* package, covariates can be included, graphs can be obtained, all data censoring types can be handled, the total length of stay in a state can be estimated, and model diagnostics can be done (6). In these models, the likelihood ratio test (LRT) is used to choose a better fit model between time-homogeneous and time non-homogeneous (piecewise constant) models (4). The goodness-of-fit (GOF) of the model is assessed by comparing the observed and the predicted number of individuals in each state at a specified time (4). The other test is the Pearson-type test which tests if the transition rates depend on several predictors which can be applied to all multistate Markov processes, including those with an absorbing state (4).

The other library within R software is the *tdc.msm* library developed by Meira-Machado *et al.* (2009) which can fit five different multistate models including time-homogeneous and non-homogeneous Markov multistate models; and Cox Markov and Cox semi-Markov multistate models (4). The *tdc.msm* library is a comprehensive package for modelling multistate longitudinal data since different models can be fitted within one library and model comparison

can be done easily (4). For non-parametric estimation, the *msSurv* library within R can be used to estimate state occupation probabilities, initial and exiting time in a state, and the marginal integrated transition rate for the non-Markov multistate process (35). The other R package is the *mstate* developed by Wreede *et al.* (2011) for both the competing risk models and multistate models (36). Additional R libraries are the *etm* library Allignol *et al.* (2011) for empirical transition probabilities and the *change LOS* (change length of hospital stay) library introduced by Wangler *et al.* 2006) for the Aalen-Johansen Estimator is implemented within R software. The limitation of the *change LOS* library is that it does not support the inclusion of covariates in the multistate model and left truncated data (37). However, the *mvna* library can handle both left truncated and right censored multistate data (37).

The STATA software (licensed for use) (38) can fit the MLE multistate models using the *multistate model ado files* developed by Cowther and Lambert (2016) which restructures and declares the multistate data as survival and any survival model within STATA can be used (39). This package can estimate each transition rate by its unique model structure, assuming either a Markov or semi-Markov process (39). Uniquely to the STATA models is the ability to estimate each transition rate assuming different hazard functions which best fit the transition as compared to the R *msm* models which assumes the same hazard function on all the model transition processes. Another option in STATA is using *ill prep* and *stpm2illd* commands which can perform a similar analysis as described elsewhere (40).

#### *2.3.2. Bayesian Perspective*

The BE approach is a flexible method that gives posterior transition estimates from both the likelihood of data and the additional prior information for the unknown parameters (41). Multistate models fitted within Bayesian framework have not been fully implemented compared to the MLE models; however, BE multistate models are much more flexible and can handle most of the data features which are not possible with the MLE multistate models.

The BE multistate models features vary with the type of software used. Free BayesX software can handle multistate models using the *bayesreg* object (42). An illustration of these models was given by Kneib *et al.* (2008) for a continuous-

time semiparametric Markov model and was first demonstrated on human sleep data (43). These models are emerging and have been recently used on the rheumatoid disease progression (44) and HIV disease progression (Matsena Zingoni et al.- accepted paper) which is the first paper to include the spatial effects within BE multistate models framework. The hazard rates are estimated in a multiplicative Cox-type way with the baseline hazard rate simultaneously estimated on a log scale with the other predictors (43). Both empirical Bayes approach which treats unknown hyper-parameters as constants and the fully Bayes approach which provides priors to all unknown parameters in the model can be implemented (43).

The strength of BE multistate models within BayesX is their ability to flexible predictor can handle covariates with time-varying effects, non-linear covariates effects of continuous covariates using penalized splines, non-parametric baseline effects and parametric effects of fixed covariates (45). The MLE multistate models fitted within R *msm* library assume that the functional form of the effects of predictive factors is linear by default (or log-linear) which might not be the case always. This restricts the assessment of non-linear effects on the transitions rates using frequentist multistate models since violation of this assumption may lead to inaccurate statistical conclusions, the increased bias in estimates and decreases statistical power for statistical significance tests (6). However, BE multistate models can fit these non-linear effects of continuous covariates using penalized splines. Another strength of the BE multistate models if the ability to account for frailty terms to explain unobserved heterogeneity in the collected information either at the individual level or spatial level which R *msm* library cannot do (45).

For instance, in HIV disease progression model, it is vital to account for individual-level heterogeneity to estimate transition rates since patients respond differently to ART treatment, and these transitions may also vary by location. Despite all these strengths associated with BE multistate models within BayesX, it is still challenging to perform the model comparison using the deviance information criterion (DIC). However, one may do a prior sensitivity to validate the model by varying prior estimates then check for any changes in the posterior

estimates (coefficients or splines) of the transition rates (45).

The Windows version of Bayesian inference Using Gibbs Sampling (WinBUGS) is another software which can be used to fit multistate models (46). Not only does the WinBUGS platform handle individual data but also aggregated data. Fitting multistate models in WinBUGS makes use of the Kolmogorov's forward equations transcription into a code to estimate the transition probabilities and rates of the multistate model (32). A limitation of the WinBUGS multistate model could be explicitly solving for the Kolmogorov's forward equations solution to be able to transcribe the solutions into the model code. This exercise becomes hectic and tedious if the model has many transitions some of them being reversible, that is, a complex multistate model; however, the WinBUGS Differential Interface (WinDiff) software can be used instead (32). An illustration of this type of multistate modelling approach in WinBUGS is given in detail in Section 3.

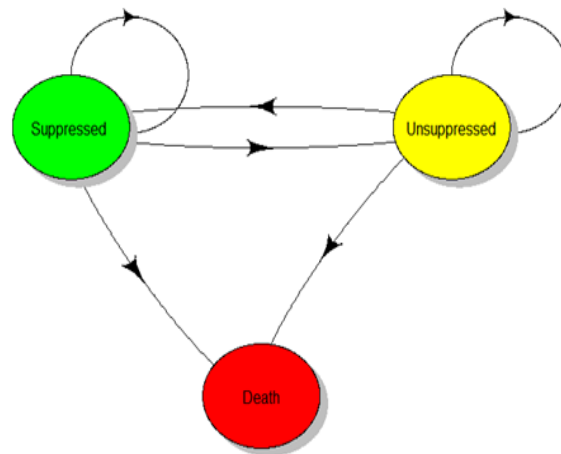
### *2.3.3. Software Challenges in Statistical Modelling*

There are strengths and limitations associated with each software type which handles multistate models, some of which might not have been discussed in detail in this review. Firstly, most of the existing software assumes the Markov property and time-homogeneous by default which makes it difficult if these assumptions are violated. Secondly, the Markov assumption can be difficult to test, and in most cases, studies are silent on pre-model assumption testing of the Markov property; however, *markovchain* library implemented in R is one of the packages which can test for Markov property (47). Thirdly, not all software types are freely available for use like BayesX, WinBUGS and R, some of the software types require a license to be granted access, for instance, STATA and SAS. Fourthly, data argumentation (structure) is different in each of the software, which may be a major drawback if one wishes to make comparison across software. In addition to this, at times convergence is an issue in multistate models depending on the size of the data, the model complexity of the proposed model (number of states, reversible transition and number of covariates included) and the preferred method of estimation (BE or MLE) as the models may take much longer

(hours or days) to converge or may fail to converge at all. Lastly and most importantly, the result outputs vary moving from one software package to another; hence, methodological background theory for each package is crucial to enable results comparison. For instance, in the R *msm* library, one gets the exact transition rate values which can be interpreted directly while in BayesX outputs, the estimates are based on the flexible predictor on a log scale (43).

**3. ILLUSTRATION OF THE KOLMOGOROV-CHAPMAN FORWARD EQUATIONS**

In this example, we implemented the method for estimation transition rates using partially observed data explained elsewhere (23); however, our illustration is based on a three-staged reversible multistate model with states defined based on viral load measurements as shown in Figure 3. In this model, individuals from State 1 (suppressed viral load) may die (State 3) via State 2 (unsuppressed viral load) or directly from State 1. Again individuals may move backwards to State 1 once they are in State 2 (reversible transition).



**Figure3:** The schematic presentation of three states partially observed the multistate model and the corresponding individual-specific transition intensities (State 1= Viral load <50copies/mL; State 2= Viral load ≥50copies/mL)

In general, the transition rates and transition probabilities are mapped using the Kolmogorov-Chapman forward equation, which has the following solution:

$$P(t) = \exp(Q(t)) = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n \quad (1)$$

where  $P(t)$  is the transition probability matrix,

$Q(t)$  is the transition rate matrix and the

$$Q(t) = \begin{pmatrix} -(\gamma_{12} + \gamma_{13}) & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & -(\gamma_{21} + \gamma_{23}) & \gamma_{23} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -\lambda_1 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & -\lambda_2 & \gamma_{23} \\ 0 & 0 & 0 \end{pmatrix}.$$

whose row totals sum to 0. Since State 3 is the absorbing state, the last row entries are equal to zero. With the  $Q(t)$  matrix, the Kolmogorov-Chapman forward equations solutions defined in

$$P(t) = \begin{pmatrix} \pi_{11}(t) & \pi_{12}(t) & \pi_{13}(t) \\ \pi_{21}(t) & \pi_{22}(t) & \pi_{23}(t) \\ 0 & 0 & 1 \end{pmatrix}$$

number of observed transitions is defined by  $n$  while  $t$  defines the time. Let  $\gamma_{jk}$  be the transition rate elements within the 3x3 matrix, the flow rate from a defined state  $j$  be  $\lambda_j$ , for instance,  $(\lambda_1 = \gamma_{12} + \gamma_{13})$ ; therefore, the transition rate matrix is defined as:

equation (1) can be used to map the  $P(t)$  matrix which has the transition probability  $\pi_{jk}(t)$  elements, that is,

where the row totals equal to 1. To fully define each  $\pi_{jk}(t)$  component of the matrix, we modified slightly the solutions given elsewhere (32) to get comparable estimates to those

$$h = \sqrt{(\lambda_1 - \lambda_2)^2 + 4\gamma_{12}\gamma_{21}} \text{ for } \lambda_1 = \gamma_{12} + \gamma_{13} \text{ and } \lambda_2 = \gamma_{21} + \gamma_{23} \text{ and}$$

$$e1 = \exp\left(-\frac{1}{2}(\lambda_1 + \lambda_2 - h) \times (2 \times t)\right) \text{ and } e2 = \exp\left(-\frac{1}{2}(\lambda_1 + \lambda_2 + h) \times (2 \times t)\right).$$

Then the Kolmogorov-Chapman solution for each  $\pi_{jk}(t)$  element is simplified to:

$$\pi_{11}(t) = \frac{e1 \times (-\lambda_1 + \lambda_2 + h) + e2(\lambda_1 - \lambda_2 + h)}{2h}$$

$$\pi_{12}(t) = \frac{(-\lambda_1 + \lambda_2 + h) \times (\lambda_1 - \lambda_2 + h) \times (e1 - e2)}{4h\gamma_{21}}$$

$$\pi_{13}(t) = 1 - \pi_{11}(t) - \pi_{12}(t)$$

$$\pi_{21}(t) = \frac{\gamma_{21}(e1 - e2)}{h}$$

$$\pi_{22}(t) = \frac{e1 \times (\lambda_1 - \lambda_2 + h) + e2 \times (-\lambda_1 + \lambda_2 + h)}{2h}$$

$$\pi_{23}(t) = 1 - \pi_{21}(t) - \pi_{22}(t)$$

obtained in R using *msm* library, assuming one has individually observed data. We considered this as a sensitivity approach to validate proper transcription of the code in WinBUGS. Let

In this example, we assumed a time-homogeneous Markov model and the data were aggregated having been observed over two-time points in a single year follow-up. This means for each individual, the initial state at ART initiation and the final state a year later at the end of follow-up is known but the route to the last observed state after a year is unknown. This means virally suppressed individuals (State 1) at time 0 who later died after a year of follow-up may have died directly from the State 1 or may have died via State 2 (unsuppressed viral load). This type of data is said to be partially observed since the intermediate transitions are unknown. The used aggregated data is displayed in Table 1.

**Table1:** Partially observed data for viral load suppression among adult ART patients from the Zimbabwe national ART program after a year time cycle for a three-state model

	Number of participants at baseline	Initial state, <i>j</i>	Final state after a year of follow-up, <i>k</i>		
			State 1	State 2	State 3
	2490	State 1	2269	143	78
	3106	State 2	137	2882	87
	0	State 3	0	0	0
<b>Totals</b>	5596		2406	3025	165

Let  $x_{jk}$  be the number of transitions, which is equal to the number of individuals over states, observed after a year from state *j* to *k* for  $j \neq k$ . Since we employed the Bayesian inference, the likelihood function of this data was assumed to follow a multinomial distribution with probabilities:

$$(x_{j,1}, x_{j,2}, \dots, x_{j,m}) \square \text{Multinomial}(\pi_{j,1}, \pi_{j,2}, \dots, \pi_{j,m}; n_j) \quad (3)$$

The priors were assumed to be non-informative exponential priors for the unknown transition rate parameters  $\gamma_{12}, \gamma_{13}, \gamma_{21}$  and  $\gamma_{23}$  with parameter values  $\beta = 0.001$ . We also performed a prior sensitivity by varying the prior distributions for the transition rate to follow a Gamma distribution with a parameter of 0.1 for both the

shape and scale. However, the estimates were not similar meaning the choice of prior distribution did not influence the posterior transition rate estimates.

$$P(\gamma_{jk}) \propto \beta \exp(-\beta\gamma_{jk}) \quad (4)$$

The Bayesian MCMC method was implemented in WinBUGs. We set 10,000 MCMC simulations after a burn-in period of 1,000 simulations and thinning of 10. The code used is provided in **Appendix B**, and this code may be viewed at the transcription of  $P(t)$  matrix each element fully defined (Equation 2), the likelihood function (Equation 3) and prior function (Equation 4). The posterior distribution fitted is given by Equation (5) below:

$$P(\pi_{jk}, \gamma_{jk} / x_j) \propto \left[ \frac{\Gamma\left(\sum_j x_j + 1\right)}{\prod_j \Gamma(x_j + 1)} \prod_{j=1}^m \pi_{jk}^{x_j} \right] \times [\beta \exp(-\beta\gamma_{jk})] \quad (5)$$



The posterior estimates are shown in **Table 2**.

**Table2:** Posterior estimates and correlations for the transition rates for the three-state model using partially observed data from the adult ART patients in the Zimbabwe national ART data after a year of follow-up.

Parameter	Parameter estimates		Correlations			
	Mean estimate	95% Credible interval	$\gamma_{12}$	$\gamma_{13}$	$\gamma_{21}$	$\gamma_{23}$
$\gamma_{12}$	0.0633	0.053 to 0.074	1.0	0.068	0.124	0.0985
$\gamma_{13}$	0.0323	0.026 to 0.040		1.0	0.121	0.046
$\gamma_{21}$	0.0485	0.041 to 0.056			1.0	0.015
$\gamma_{23}$	0.0285	0.023 to 0.035				1.0

Since this data is partially observed, there is uncertainty as to the exact route, an individual who reached State 3 followed. However, there are four possibilities to describe this:

- Participant arrived directly from State 1 having not visited State 2 during the follow-up period (State 1 to State 3)
- Participant arrived directly from State 2 having not visited State 1 during the follow-up period (State 2 to State 3)
- Participant arrived from State 2 via State 1 (State 1 to State 2 to State 3)
- Participant arrived from State 1 via State 2 (State 2 to State 1 to State 3)

However, the correlations between the transition rates are positive, which means the data on reaching State 3 is compatible with an increase in each of the transition. This is also evident in the bivariate scatter plots for these four transitions in Appendix C. The transition from State 1 to State 2 was 0.0633 (95% incredible interval (CI): 0.053-0.074), and this was statistically significant and was the highest observed rate.

The transition probabilities are shown in Table 3 for the three-cycle times, 3 months, 6 months and 1years cycles.

**Table3:** Posterior estimates for the transition probabilities during three months, 6months and a one-year cycle for the three-state model using partially observed data from the adult ART patients in the Zimbabwe national ART data after a year of follow-up.

Parameter	3 months		6 months		1 year	
	Mean estimate	95% Credible interval	Mean estimate	95% Credible interval	Mean estimate	95% Credible interval
$\pi_{11}(t)$	0.9765	0.973 to 0.980	0.9530	0.945 to 0.960	0.9103	0.898 to 0.922
$\pi_{12}(t)$	0.0155	0.013 to 0.018	0.0302	0.026 to 0.036	0.0580	0.048 to 0.068
$\pi_{13}(t)$	0.0081	0.006 to 0.010	0.0160	0.012 to 0.020	0.0318	0.026 to 0.039
$\pi_{21}(t)$	0.0119	0.010 to 0.014	0.0231	0.019 to 0.027	0.0442	0.037 to 0.052
$\pi_{22}(t)$	0.9810	0.979 to 0.983	0.9626	0.958 to 0.967	0.9273	0.918 to 0.936
$\pi_{23}(t)$	0.0071	0.006 to 0.009	0.0142	0.011 to 0.017	0.0285	0.023 to 0.035

From these estimates, the probability of leaving the initial state increases as the cycle time increases, that is, for a transition from State 1 to State 2, the probability at 3 months cycle was smaller than the probability at 6 months cycles, and both were smaller than the probability at 1-year cycle:  $(\pi_{12}(t=3months) = 0.0155 < \pi_{12}(t=6months) = 0.0302 < \pi_{12}(t=1year) = 0.058)$ . This pattern was similar across other transitions and was as anticipated for such types of models.

**4. DISCUSSION**

In this manuscript, we have discussed the various types of multistate models disposable to use in different science fields to answer different

kinds of research questions from longitudinal time to event data. The advantage of multistate models is the ability to draw schematic diagrams of mutually exclusive states which help to understand the model better. We have highlighted different assumptions for multistate models, existing software which can implement multistate models, strength and weaknesses of these software.

We have also discussed two methods of statistical inference with emphasis on Bayesian estimation, which has not been fully utilized in the literature. We also put forward the strengths of BayesX multistate model, that is, the ability to handle spatial, non-linear covariates effects in addition to adjusting for individual

heterogeneity which models from other software cannot do and WinBUGS models which can handle aggregated data.

This manuscript further put forward the advantage of using a multistate modelling as this approach may bring out new and important biological insights in understanding the intermediate processed of a disease which ordinary regression models like Cox proportional hazard model may be ignoring. We have also used the Kolmogorov-Chapman forward equations on a new dataset on viral load measurement in HIV studies. This method affirms how important aggregated data could be in instances where there is no individual-level data. This becomes more important as it lays the foundation on the importance of collecting individual-level data and further research formulations. However, this method needs to be improved so that it covariates effects on the aggregated transition rates can be accounted for.

The advantages of Bayesian inference over the frequentist perspective is that the Bayesian approach is theoretically simpler, more robust, more flexible and easier to implement with minimum support (1). Though Bayesian estimation can give comprehensive posterior estimates of the transition rates, these models are generally computationally intensive since they require more time (hours or days) to converge. Despite requiring technical expertise in fitting the model to achieve convergence and make the correct inferences; the main disadvantage of Bayesian models in the specification of the priors, which is normally subjective. This is quite a debatable issue in Bayesian modelling since the observed posterior estimates are heavily dependent on the prior specification, especially for informative prior choices. However, to override this argument, assigning diffuse (vague/non-informative) priors is encouraged, and results are comparable with the frequentist estimates(32,43). Also, prior sensitivity analysis is encouraged as an excellent modelling courtesy to validate that the choice of the prior used does not influence the results obtained.

In conclusion, multi-state modelling offers a flexible tool for the study disease progression and estimate transition rates using various forms of assumptions, data and estimation methods. Multistate models bring out significant disease progression understandings which the traditional naïve regression models may ignore. Therefore,

multistate models should be used as a supplement to the traditional naïve regression models to gain additional information.

#### **ACKNOWLEDGEMENT**

Special thanks to the Ministry of Health and Child Care, AIDS/TB Units department for the support and compilation of the data used in this study. We also acknowledge the Division of Epidemiology and Biostatistics at the School of Public Health for their assistance in the getting ethical approval of this study.

#### **REFERENCES**

- [1] Akpa OM, Oyejola BA. Review article Modeling the transmission dynamics of HIV / AIDS epidemics: an introduction and a review. *J Infect Dev Ctries.* 2010;4(10):597–608.
- [2] Dehghan M, Nasri M. Global stability of a deterministic model for HIV infection in vivo. *Chaos, Solitons and Fractals.* 2007; 34:1225–38. doi: 10.1016/j.chaos.2006.03.106
- [3] Djordjevic J, Silva CJ, Torres DFM. A stochastic SICA epidemic model for HIV transmission. *Applied Mathematics Letters.* 2018; 84:168–75. doi: 10.1016/j.aml.2018.05.0 05
- [4] Meira-Machado L, Uña-álvarez J De, Cadarso-suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Statistical methods in Medical Research.* 2009;18(2):195–222.
- [5] Hougaard P. Multi-state Models: A Review. *Lifetime Data Analysis.* 1999; 264:239–64.
- [6] Jackson C. Multi-state modelling with R: the msm package. 2016;1–57.
- [7] Commenges D. Multi-state Models in Epidemiology. *Lifetime Data Analysis.* 1999; 5:315–27.
- [8] Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondre K, et al. Competing risks analyses: Objectives and approaches. *European Heart Journal.* 2014;35(42):2936–41.
- [9] Cox DR. Regression Models and Life-Tables. *Journal of Royal Statistical Society Series B(Methodological).* 1972;34(2):187–220. doi: 10.1007/978-1-4612-4380-9\_37
- [10] Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, Safaee A, Abadi A, Zali MR. Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific journal of cancer prevention: APJCP.* 2007;8(3):412–6.
- [11] Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association.* 1999;94(446):496–509.
- [12] Andersen PK, Z SA, Susanne Rosthøj. Competing risks as a multi-state model.

- Statistical methods in Medical Research. 2002; 11:203–15.
- [13] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in medicine*. 2007; 26:2389–430.
- [14] Coviello V, Boggess M. Cumulative Incidence Estimation in the Presence of Competing Risks. *Stata J*. 2004;4(2):103–12.
- [15] Eulenburg C, Mahner S, Woelber L, Wegscheider K. A Systematic Model Specification Procedure for an Illness-Death Model without Recovery. *PLoS*. 2015;(August): 1–16. doi: 10.1371/journal.pone. 0123489
- [16] Matsena Zingoni Z, Chirwa TF, Todd J, Musenge E. HIV Disease Progression Among Antiretroviral Therapy Patients in Zimbabwe: A Multistate Markov Model. *Frontiers in Public Health*. 2019; 7:1–15. doi: 10.3389/fpubh.2019. 00326
- [17] Shoko C, Chikobvu D. A superiority of viral load over CD4 cell count when predicting mortality in HIV patients on therapy. *BMC Infectious Diseases*. 2019;19(1):1–10. doi: 10.1186/s12879-019-3781-1
- [18] Dessie ZG. Multistate models of HIV / AIDS by homogeneous semi-Markov process. *American Journal of Biostatistics*. 2014;4(2): 21–8. doi: 10.3844/ajbssp.2014.21.28
- [19] Ford N, Roberts T, Calmy A. Viral load monitoring in resource-limited settings. *AIDS*. 2012;26(13):1719–20.
- [20] Oyugi JH, Byakika-Tusiime J, Ragland K, Laeyendecker O, Mugerwa R, Kityo C, et al. Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS*. 2007;21(8):965–71.
- [21] Commenges D. Inference for multi-state models from interval-censored data. *Statistical methods in Medical Research*. 2002; 11:167–82. doi: 10.1191/0962280202sm279ra
- [22] Edlin R, McCabe C, Hulme C, Hall P, Wright J, Edlin R, et al. Finding the Evidence for Decision Analytic Cost Effectiveness Models. *Cost Effectiveness Modelling for Health Technology Assessment*. 2015. 15–40 p. doi: 10.1007/978-3-319-15744-3\_2
- [23] Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: Uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making*. 2005;25(6):633–45. doi: 10.1177/0272989X05 282637
- [24] Jackson CH, Sharples LD, Thompson SG, Duffy SW. Multistate Markov models for disease progression with classification error. *The Statistician*. 2003; 52:193–209.
- [25] Goshu AT, Dessie ZG. Modelling Progression of HIV/AIDS Disease Stages Using Semi-Markov Processes. *Journal of Data Science*. 2013;11(2):269–80.
- [26] Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. 2015;1–15.
- [27] Jackson CH. Multistate Markov models for disease progression with classification error. *The Statistician*. 2003;52(2):193–209. doi: 10.3844/ajbssp.2014.21.28
- [28] Leung K-M, Elashoff RM, Afifi AA. Censoring Issues in Survival. *Annu Rev Public Health*. 1997; 18:83–104.
- [29] Siannis F, Copas J, Lu G. Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*. 2005;6(1):77–91. doi: 10.1093/biostatistics/kxh019
- [30] Macdonald A. An Actuarial Survey of Statistical Models for Decrement and Transition Data: Multiple state, POisson and Binomial Models. *British Actuarial Journal*. 1996;2(1):129–55.
- [31] Lindsey JC, Yan P. Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in medicine*. 1994; 13:805–21.
- [32] Welton NJ, Ades AE. Estimation of markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Supplement. Medical Decision Making*. 2005;25(6):633–45.
- [33] R Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2011.
- [34] Jackson CH. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*. 2011;38(8):1–28.
- [35] Ferguson N, Brock G. msSurv: An R Package for Nonparametric Estimation of Multistate Models. *Journal Statistics Software*. 2012;50(14):1–24.
- [36] Wreede LC de, Fiocco M, Putter H. state: An R Package for the Analysis of Competing Risks and Multi-State Models. *Journal of Statistical Software*. 2011;38(7):1–30. doi: 10.18637/jss.v038.i07
- [37] Allignol A, Schumacher M, Beyersmann J. Empirical Transition Matrix of Multi-State Models: The etm Package. *Journal of Statistical*. 2011;38(4):1–15.
- [38] StataCorp. *Stata Statistical Software: Release 15*. College Station, Texas, USA. Texas: STATA; 2017.

[39] Crowther MJ, Lambert PC. Parametric multi-state survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences *Statistics. Statistics in medicine.* 2016;1–40. doi: 10.1002/sim.0000

[40] Hinchliffe SR, Scott DA, C.Lambert P. Flexible parametric illness-death models. *The Stata Journal.* 2013;13(4):759–75.

[41] Castro M de, Chen M-H, Zhang Y. Bayesian Path Specific Frailty Models for Multi-State Survival Data with Applications. *Biometrics.* 2015;33(4):395–401. doi: 10.1038/nbt.3121. ChIP-nexus

[42] Belits C, Brezger A, Klein N, Kneib T, Lang S, Umlauf N. *BayesX - Software for Bayesian Inference in Structured Additive Regression Models.* Boston: Free Software Foundation; 2015.

[43] Kneib T, Hennerfeind A. Bayesian semiparametric multi-state models. *Statistical Modelling.* 2008;8(2):169–98.

[44] Musenge E. Rheumatoid arthritis disease progression in a South African cohort: Bayesian multistate chronic disease, dynamic modelling. In: South African Center for Epidemiological Modelling and Analysis, editor. *South African Center for Epidemiological Modeling and Analysis.* Cape Town, South Africa: SACEMA; 2013.

[45] Fahrmeir L, Klinger A. A nonparametric multiplicative hazard model for event history analysis. *Biometrika.* 1998;85(3):581–92.

[46] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – A Bayesian modelling framework:

Concepts, structure, and extensibility. *Statistics and Computing.* 2000; 10:325–37. doi: 10.1023/A:1008929526011

[47] Spedicato GA, Signorelli M. The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R.

## 5. APPENDIX

### 5.1. Appendix A: The Maximum Likelihood Estimation

The transition rates of a multistate model can be estimated theoretically. Considering the three-state model shown in Figure B1, the aim is to estimate the four transition parameters shown on the model. More a maximum likelihood estimation (MLE), the first step is to get the product of the distribution function of the parameters. Let the transition function be

$$f(\gamma_{i,jk}) = \frac{(\gamma_{i,jk})^{n_{jk}} \exp(-\gamma_{i,jk}(t_j))}{n_{jk}!} \quad \text{for } j \neq k \quad (5)$$

where  $\gamma_{i,jk}$  if the transition rate from state  $j$  to state  $k$  for an individual  $i$ . The number of the observed movements between the states is represented by  $n_{jk}$  and  $t$  is the total observed waiting time in the state  $j$  for  $j=1, 2$ .

To get the likelihood function:

$$L(\gamma_{i,jk}) = \prod_{j \neq k} \left[ \frac{(\gamma_{i,jk})^{n_{jk}} \exp(-\gamma_{i,jk}(t))}{n_{jk}!} \right] \quad \text{for } j, k = 1, 2, 3 \quad (6)$$

$$= \frac{(\gamma_{i,12})^{n_{12}} \exp(-\gamma_{i,12}(t_1))}{n_{12}!} \times \frac{(\gamma_{i,13})^{n_{13}} \exp(-\gamma_{i,13}(t_1))}{n_{13}!}$$

$$\times \frac{(\gamma_{i,21})^{n_{21}} \exp(-\gamma_{i,21}(t_2))}{n_{21}!} \times \frac{(\gamma_{i,23})^{n_{23}} \exp(-\gamma_{i,23}(t_2))}{n_{23}!}$$

Treating the factorial part in the equation as a constant and taking log both sides yields:

$$\ln L(\gamma_{i,jk}) \propto \ln(\gamma_{i,12})^{n_{12}} - \gamma_{i,12}(t_1) + \ln(\gamma_{i,13})^{n_{13}} - \gamma_{i,13}(t_1) \quad (7)$$

$$+ \ln(\gamma_{i,21})^{n_{21}} - \gamma_{i,21}(t_2) + \ln(\gamma_{i,23})^{n_{23}} - \gamma_{i,23}(t_2)$$

which simplifies to

$$\ln L(\gamma_{i,jk}) = n_{12} \ln(\gamma_{i,12}) - \gamma_{i,12}(t_1) + n_{13} \ln(\gamma_{i,13}) - \gamma_{i,13}(t_1)$$

$$+ n_{21} \ln(\gamma_{i,21}) - \gamma_{i,21}(t_2) + n_{23} \ln(\gamma_{i,23}) - \gamma_{i,23}(t_2)$$

Re-arranging the terms gives

$$= n_{12} \ln(\gamma_{i,12}) + n_{13} \ln(\gamma_{i,13}) - [\gamma_{i,12}(t_1) + \gamma_{i,13}(t_1)]$$

$$+ n_{21} \ln(\gamma_{i,21}) + n_{23} \ln(\gamma_{i,23}) - [\gamma_{i,21}(t_2) + \gamma_{i,23}(t_2)]$$



The summation of transition rates from the same state is defined at the *flow rate*,  $\lambda_{i,j}$ , which defines the probability of transition from the state  $j$ . This means:

$$\lambda_{i,j} = \sum_{j \neq k} \gamma_{i,jk} \quad (8)$$

Let  $\rho_{i,jk}$  be the conditional probability that

Substituting equation (1.4) into equation (1.3) yields, that is,  $\lambda_1(t_1) = [\gamma_{i,12}(t_1) + \gamma_{i,13}(t_1)]$  and

$$\lambda_2(t_2) = [\gamma_{i,21}(t_2) + \gamma_{i,23}(t_2)]:$$

$$\ln L(\gamma_{i,jk}) = n_{12} \ln(\gamma_{i,12}) + n_{13} \ln(\gamma_{i,13}) - \lambda_1(t_1) + n_{21} \ln(\gamma_{i,21}) + n_{23} \ln(\gamma_{i,23}) - \lambda_2(t_2)$$

The next step is differentiating equation (1.3) with respect to each transition rate in the equation:

For transition  $\gamma_{i,12}$

$$\frac{\partial \ln L(\gamma_{i,12})}{\partial (\gamma_{i,12})} = \frac{\partial}{\partial (\gamma_{i,12})} (n_{12} \ln \gamma_{i,12} - \gamma_{i,12}(t_1))$$

$$= \frac{n_{12}}{\gamma_{i,12}} - t_1$$

Equating the solution to zero and solve to get the MLE of that transition yields:

$$\hat{\gamma}_{i,12} = \frac{n_{12}}{t_1}$$

Similarly, the other MLE for the other transitions becomes

$$\hat{\gamma}_{i,13} = \frac{n_{13}}{t_1}; \hat{\gamma}_{i,21} = \frac{n_{21}}{t_2}; \hat{\gamma}_{i,23} = \frac{n_{23}}{t_2}$$

From this working, if one knows the number of the transitions from state  $j$  to  $k$  and the total exposure time in state  $j$ , the transition rates can be estimated using equation (6). Similarly, if the waiting time is unknown, the transition rate and the number of observation can give an estimate of the time.

**5.2. Appendix B: Winbugs Three State Reversible Multistate Model Code**

```
model {
#Multinomial likelihood for observed data
for (i in 1:2)
```

the next destination is state  $k$  given that the transition from state  $j$  to  $k$  occurs. Then the flow rate can be defined in terms of the conditional probability as

$$\gamma_{i,jk} = \begin{cases} \lambda_j \rho_{i,jk} & \text{for } j \neq k \\ -\lambda_j & \text{for } j = k \text{ since } \rho_{i,jj} = \rho_{i,kk} = 1 \end{cases}$$

```

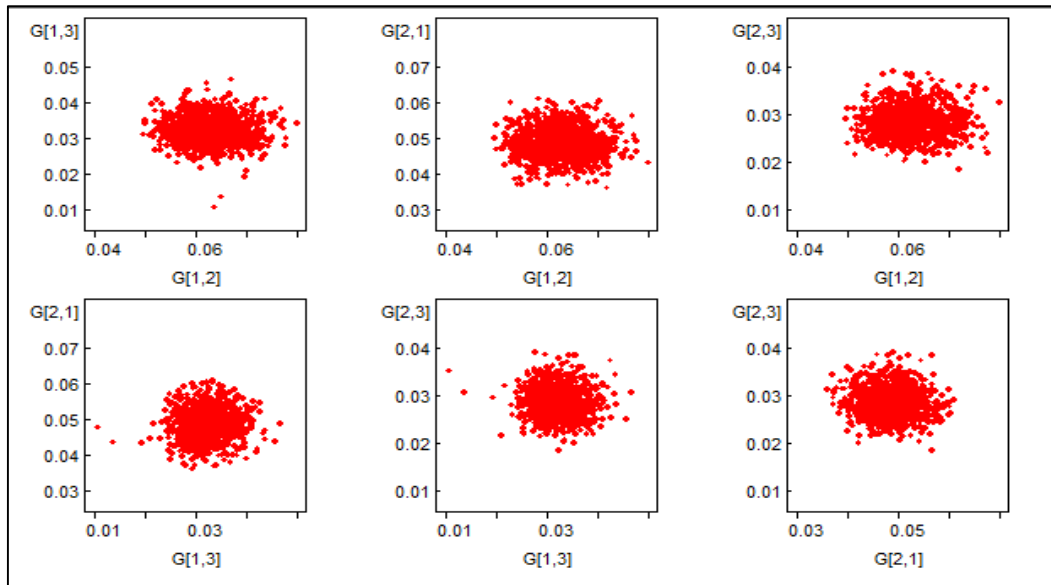
r[i,1:3] ~ dmulti(P[i,1:3],n[i]) }
#Find transition probabilities (for a given time)
in terms of rates
h<- sqrt(pow(lambda[1]-lambda[2], 2)
+ 4*G[1,2]*G[2,1])
e1<-exp(-.5*(lambda[1] + lambda[2] -
h)*t.obs)
e2<-exp(-.5*(lambda[1] + lambda[2]
+ h)*t.obs)
P[1,1]<-((-
lambda[1]+lambda[2]+h)*e1+
(lambda[1]-
lambda[2]+h)*e2)/(2*h)
P[1,2]<-((-
lambda[1]+lambda[2]+h)*(lambda[1]-
lambda[2]+h)*(e1-e2))/(4*h*G[2,1])
P[1,3]<-1 - P[1,1] - P[1,2]
P[2,1]<- G[2,1]*(e1-e2)/h
P[2,2]<-
((lambda[1]-
lambda[2]+h)*e1+ (-lambda[1] + lambda[2] +
h)*e2)/(2*h)
P[2,3]<-1 - P[2,1] - P[2,2]
#Give exponential priors for unknown transition
rate parameters
for (i in 1:2)
{
for (j in (i+1):3)
{G[i,j] ~ dexp(.001)}
}
for (i in 2:2)
{
for (j in 1:(i-1))
{G[i,j] ~ dexp(.001)}
}
```

```

    }
lambda[1]<- G[1,2] + G[1,3]
lambda[2]<- G[2,1] + G[2,3]
#Find P(t.new) for a given new time of interest
e1.new<-exp(-.5*(lambda[1] + lambda[2] - h)*t.new)
e2.new<-exp(-.5*(lambda[1] + lambda[2] + h)*t.new)
Pt[1,1]<-((-lambda[1]+lambda[2]+h)*e1.new+
(lambda[1] -lambda[2]+h)*e2.new)/(2*h)
Pt[1,2]<-((-lambda[1]+lambda[2]+h)*(lambda[1]-
lambda[2]+h)*(e1.new-e2.new))/(4*h*G[2,1])
Pt[1,3]<-1 - P[1,1] - P[1,2]
Pt[2,1]<- G[2,1]*(e1.new-e2.new)/h
Pt[2,2]<- ((lambda[1]-lambda[2]+h)*e1.new+ (-
lambda[1] + lambda[2] + h)*e2.new)/(2*h)
Pt[2,3]<-1 - P[2,1] - P[2,2]
}

```

**5.3. Appendix C: The bivariate transition rates scatter plots**



**FigureA1:** Bivariate scatter plots rates for the four transition rates parameters from the partially observed data

**Citation:** Zvifadzo Matsena Zingoni, et al, A Critical Review of Multistate Bayesian Models Modelling Approaches in Monitoring Disease Progression: Use of Kolmogorov-Chapman Forward Equations with in WinBUGS ARC Journal of Public Health and Community Medicine. 2019; 4(4):9-22. DOI: [dx.doi.org/10.20431/2456-0596.0404003](https://doi.org/10.20431/2456-0596.0404003).

**Copyright:** © 2019 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.