ORIGINAL RESEARCH

Health Science Reports *Open Access* WILEY

A Markov jump process approach to modeling blood donor status: Donor retention and attrition rates at a blood service center in Zimbabwe

Delson Chikobvu | Coster Chideme

Department of Mathematical Statistics and Actuarial Sciences, University of the Free State, Bloemfontein, South Africa

Correspondence
Coster Chideme, Department of Mathematical Statistics and Actuarial Sciences, University of the Free State, Bloemfontein, South Africa.
Email: 2019532015@ufs4life.ac.za

Abstract

Background: Blood service agencies depend upon the availability of regular blood donors for sustainability. The knowledge and understanding of the stochastic behavior of donors is the first step toward sustaining the blood supply. Analyzing the changes in the donor status within the donor pool will help the blood service authorities to manage the blood donation process.

Objectives: The study presents a multistate Markov jump model in analyzing the changes in blood donor status during their blood donation career. Relevant covariates are used to aid in explaining the transitions.

Materials and Methods: The status of a blood donor i that can be in one of four states $S = \{1; 2; 3; 4\}$. A new donor ($s = 1$), repeat/regular donor ($s = 2$), occasional donor ($s = 3$), and lapsed donor ($s = 4$). A Continuous-time Markov model was used to estimate blood donor progression during their blood donation career. Frequencies of blood donations made in a given time interval determines the state occupied.

Results: In the early years of blood donation career, first-time donors have a higher likelihood of becoming regular donors. Donor attrition increases with time whilst donor retention decreases with time. The results show that when the jump process is currently in an occasional state, the probability that it moves into lapsed state when it leaves the occasional state is given as 69.06%. Similarly, donors are forecasted to spend 21.193 months (1.8 years) in the occasional state before lapsing. Repeat donors can spend 39.342 months (3.3 years) in the regular state before the transition to other states. The study established that donor-specific demographic factors such as age and gender are critical in donor status transitions.

Conclusions: With the passage of time, donor status evolves, with trend inclined towards reduction in the frequency of blood donations as more donors become inactive or lapsed. The transition of donors in various states can be described by a time homogeneous Markov model.

KEYWORDS
blood donor status, continuous-time Markov model, Kolmogorov's forward equations, Markov jump process, multistate model, retention and attrition rates.

1 | INTRODUCTION

The blood supply chain is facing a dual problem of ever-increasing demand for blood and the sometimes dwindling number of blood donors.[1,2] Many countries have continued to experience a decrease in number of blood donors yearly resulting in blood inventories instability. In the Netherlands, for example, approximately 10% of blood donors lapse each year.[3] In Australia, studies have shown that nearly a third of first-time blood donors never return to donate blood.[4] As alluded by Schreiber et al.,[5] there is need to establish reasons why approximately half of first-time donors give blood once and never return. For example, in the United States, 8% of donors give blood regularly while an estimated 62% of first-time donors never return to donate at the same center of their first donation.[6]

Previous studies indicated that regular or repeat donors are crucial for sustainability of the blood industry when compared to new donors who are associated with uncertainty in the quantity and safety of the donated blood.[7] Regular donors are a key resource in maintaining a safe and sufficient blood supply as they contribute immensely to blood donations during their lifetime.[5,6] Therefore, blood collection agencies should strive to retain regular donors as they are safer and more cost effective than new or unregular donors.[8] It is essential to continuously recruit new donors to replace those who drop or retire from the donation career.[9]

The frequency of donations and intentions to donate is not easily predictable and can be regarded as stochastic random variables. Blood collection agencies could gain insights into the stochastic characteristics of blood donors and their stochastic behavioral patterns in blood donation. To predict the dynamics of the blood donor status in the donor pool, a prediction and modeling mathematical tool is necessary. The movement of blood donors in their donor career possesses stochastic characteristics and that can be modeled as a Markov jump process. Markov jump process modeling is a typical example of stochastic modeling which describes a sequence of transitions from one state to another in continuous time.

Blood donation is regarded as a career with a blood donation cycle, a donor stays in a given state for a certain period, moves to the next state or leaves the donation system on retirement or due to some other circumstances. The status i of a blood donor can be in one of four states $S = \{1; 2; 3; 4\}$ at any one time point. A new donor ($s = 1$), repeat/regular donor ($s = 2$), occasional donor ($s = 3$), and lapsed donor ($s = 4$). Blood donors move between these states with transition rates that depend on donor population characteristics and factors that motivates blood donation. Analyzing the dynamics of the blood donors in each category or status/state within the donor pool will help the blood authorities to plan on donor recruitment drives, donor retention, and making estimations of the blood units they are likely to collect.[10]

There is hardly any literature on the application of Markov jump processes in modeling transitions in status of blood donors in Zimbabwe's blood service system. The theoretical and practical framework for applying a Markov jump process model in this study helps to explain the migration and dynamics of blood donors from one state to another and how it impacts on the blood supply in a blood service center.

A study by Dyantikapm[11] applied a hidden Markov model in analyzing blood donation behavior of blood donors. The hidden Markov model approach was suitable for the study because it accommodated the time-varying feature on the data. Results showed that significant numbers of both males and females did not return to donate blood after their first-time donation. The results further showed a high dropout rate in blood donation over time. A study to establish the variation in blood supply and demand was conducted by Abubakar et al.[12] They developed a stochastic model based on transitions among the various states. The model incorporated three donor groups namely; first time, sporadic and regular donors. The results showed that, of the first-time donors who did not leave the system, only 28% became regular donors and the rest became sporadic donors.

A study by Bar-Lev et al.[13] focused on the application of a Markov model with a first order dependence in blood transfusion in a city hospital. The blood groups formed the state space of the model. Their model sought to optimize the donation and transfusion of blood considering the different blood group types and the various cross-matching permutations. Results showed that blood group AB was the most demanded blood type, followed by blood groups B, A, and O in that order. Transition probabilities from the results showed little change in future demand for blood transfusion at the hospital. A stochastic model for a blood bank in which the quantities of blood are supplied and demanded according to stochastic processes was developed by Hosseinifard et al.[14]

The goal of this study is to present/apply the multi-state Markov modeling as an option analyzing donor behavior in Zimbabwe. The model involves estimating blood donor progression during their blood donation career using the frequencies of blood donations made in each time interval to determine the Markov states. The outcome of the research is expected to provide Zimbabwe blood service managers with a decision support system to manage blood donor recruitment, attrition, and retention. It could enable the prediction of future proportions of blood donors in different categories and add new knowledge to existing literature on blood donors' attrition and retention, not only in Zimbabwe, but the region and similar countries throughout the world.

The sections in this study are organized as follows, Section 2 presents methods in data collection and the methodology. Section 3 focuses on data analysis and results. Section 4 presents discussions and conclusions are in Section 5.

2 | MATERIALS AND METHODS

This section describes the sampling, data gathering tools, data set, and the data analyzing tools.

A retrospective study was conducted on the blood donors selected from the National Blood Service Zimbabwe's (NBSZ) head

office database in Harare. The head office data is assumed to contain a diversity of characteristics of the blood donors in Zimbabwe's donor population.

2.1 | Description of the data

The study is based on blood donations data collected in Harare, Zimbabwe over a period of four years, from January 1, 2014 to December 31, 2017. The sample size was calculated by using the Taro Yamane formula, stated as $\left(n = \frac{N}{(1 + N * e^2)}\right)$, where n is the sample size, N is the population of the study and e is the error in the calculation (95% or 0.05). A total of all the 8312 new voluntary and non-remunerated blood donors in 2014 were retrieved from the donors' database thus forming the study population. From the Taro Yamane formula, the sample size was found out to be 382 and was increased by an additional 68 donors to 450 donors to improve the accuracy of the estimates. Random sampling was then used to select the objects of the study. The donors' specific data on donor identification number, age, sex, number of donations each year, interval between whole blood donations and blood group were extracted as secondary data from the NBSZ database. Of the 450 new blood donors, 15 lapsed blood donors were excluded from the analysis because they had resumed donation after lapsing. The analysis assumed lapsing to be an absorbing state. The donors were classified according to the number of times they donated blood within a cycle of 12 months for a period of four years from 2014 to 2017. The donors were analyzed after every 12 months to establish their donation frequencies in each year. Once donation enter the lapsed state, no further donations were made thus resulting in 1812 observations.

Statistical analysis was performed using the *R* software which is an inbuilt *msm package* version 1.4 developed by Jackson.[16]

2.2 | Definition of states in the Markov chain

The blood donor status is defined according to the number of times each donor gave blood within a given period as per the World Health Organisation standards.[15]

The four identified major types of donors in the National Blood Service Zimbabwe blood system are, new or first-time donors, regular or repeat donors, occasional or sporadic and lapsed or inactive donors.

New donor (State 1): donor giving blood for the first time.

Regular/repeat/returning donor (State 2): donor who gives two or more donations within 12 months.

Occasional/sporadic donor (State 3): donor who gives blood sporadically once in 12 months and skips other cycles.

Lapsed donor (State 4): donor who has not donated or presented to donate blood for a period of 24 months from the date of their last donation.

2.3 | Model formulation and structure

The model is based on a continuous time Markov process with four states. Consider the status i of a blood donor that can be in one of four states $S = \{1; 2; 3; 4\}$. The donor categories are described as new donor ($s = 1$), repeat/regular donor ($s = 2$), occasional donor ($s = 3$), and lapsed donor ($s = 4$). A blood donor enters the blood donation system as first-time donor, then either becomes a repeat donor or occasional donor before lapsing.

The transitions from state i to state j are governed by a set of transition intensities.

Let a_{ij} be defined by

$$a_{ij} = \lambda_i \times P_{ij} \quad \forall \ i, j.$$

Since λ_i is the rate at which the process leaves state i and P_{ij} is the probability that it goes to state j, it follows that a_{ij} is the rate when in state i, that the process makes a transition into state j. Hence a_{ij} is called the transition rate from i to j.

Since $\sum_{j=i}^{r} P_{ij} = 1$, it follows that

$$\lambda_i = \sum_{j \neq i}^{r} a_{ij}.$$

Let T_{ij} be the time the process spends in state i before entering state $j (\neq i)$. The time T_{ij} is exponentially distributed with rate λ_i.

Let $\tilde{T}_i$ be the sojourn time in state i.

Consider a short time interval Δt and since T_{ij} and $\tilde{T}_i$ are exponentially distributed, we have,

$$P_{ij}(\Delta t) = P(\tilde{T}_i > \Delta t) = e^{-\lambda_i \Delta t} \approx 1 - \lambda_i \Delta t,$$

$$P_{ij}(\Delta t) = P(T_{ij} \leq \Delta t) = 1 - e^{-\lambda_i \Delta t} \approx \lambda_i \Delta t,$$

when Δt is small.

Therefore,

$$\lim_{\Delta t \to 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(\tilde{T}_i < \Delta t)}{\Delta t} = \lambda_i,$$

$$\lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(T_{ij} < \Delta t)}{\Delta t} = a_{ij} \text{ for } i \neq j.$$

Then the transition rates a_{ij} can be presented as a matrix:

$$Q = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rr} \end{bmatrix},$$

where the diagonal elements are $a_{ii} = -\lambda_i = -\sum_{j \neq i}^{r} a_{ij}$.

The entries of row i of the transition rate matrix A are the transition rates out of state i ($j \neq i$).

These are called departure rates from state i. The rate $-a_{ii} = \lambda_i$ is the sum total departure rate from state i. The sum of the entries in row i is always equal to 0, $\forall \ i \in S$.

By using the Markov property and the law of total probability, then

$$P_{ij}(t + s) = \sum_{k \in S}^{r} P_{ik}(t)P_{kj}(s) \ \square \ i, j \in S, t, s > 0.$$

These equations are known as the Chapman–Kolmogorov equations.

To find $P_{ij}(t)$ we start by considering the Chapman–Kolmogorov equations

$$P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(t)P_{kj}(\Delta t).$$

We consider

$$P_{ij}(t + \Delta t) - P_{ij}(t) = \sum_{k \in S} P_{ik}(t)P_{kj}(\Delta t) - [1 - P_{jj}(\Delta t)]P_{ij}(t).$$

By dividing by Δt and then taking the limit as $\Delta t \to 0$, we obtain

$$\lim_{\Delta t \to 0} \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} = \lim_{\Delta t \to 0} \left[\sum_{k \in S} P_{ik}(t) \frac{P_{kj}(\Delta t)}{\Delta t} - \frac{1 - P_{jj}(\Delta t)}{\Delta t} P_{ij}(t) \right].$$

Since the summing index is finite, and by interchanging the limit and summation, we obtain the Kolmogorov forward equations.

$$P'_{ij}(t) = \sum_{k \in S} a_{kj}P_{ik}(t) - \lambda_j P_{ij}(t) = \sum_{k \in S} P_{ik}(t)a_{kj},$$

where $a_{jj} = -\lambda_i$.

The transition between the states is shown in the transition diagram in Figure 1 below:

The problem under study is formulated as a multi-state model with four states as illustrated in Figure 1. At time t, the blood donor is in state $S(t)$ and the arrows show possible transitions between the states.
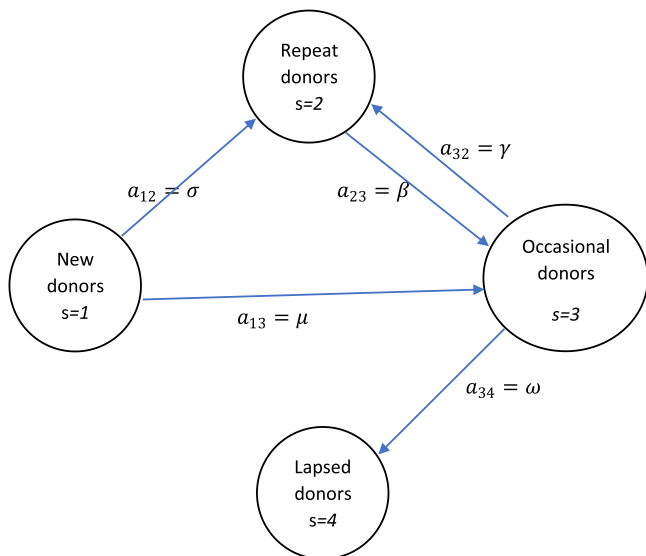

FIGURE 1 Transition rates diagram among the states

The frequency of blood donation within a cycle by a blood donor determines the status or type of blood donor.

The transitions of blood donors among the states can be expressed as a generator matrix Q given below.

$$Q = \begin{bmatrix} -\sigma - \mu & \sigma & \mu & 0 \\ 0 & -\beta & \beta & 0 \\ 0 & \gamma & -\gamma - \omega & \omega \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A blood donor can move from state 3 to 4. However, one could still have a transition intensity from some states being equal to zero in the Q matrix. It is impossible for a donor to return to state 1.

The intensity matrix Q is then used to fit the multi-state model to the data.

The transition rate matrix can be used to establish the transition equations among the blood donor type state i at time t using stochastic differential equations (SDEs).

The SDEs are derived from the Chapman–Kolmogorov equations.

SDEs can be formed using the generator matrix Q. The SDEs in turn give rise to formulas used for calculating transition probabilities.

Theorem 1. *Kolmogorov forward equations*

For a continuous-time Markov jump process, $\{X(t)\}_{t \geq 0}$, with a transition intensity matrix $Q = \{a_{ij}\}_{i,j \in S}$ and transition probability $\{P_{ij}(t)\}_{i,j \in S}$, it always holds that the solution to:

$$P'_{ij}(t) = \sum_{k \in S} P_{ik}(t)a_{kj} = P(t)Q,$$

is given by:

$$P(t) = \exp(-\lambda_i t)\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(t - v) \times a_{kj} \times \exp(-\lambda_i(v))dv.$$

The theorem is proven by starting from the solution and show that the stochastic differential equation holds.

Proof.

Since $X(s) = j$, then by minimal construction, the waiting time to the final jump follows an exponential distribution with the rate $\lambda_i = -a_{jj}$ and the solution is written as:

$$P(t) = \exp(-\lambda_i t)\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(t - v) \times a_{kj} \times \exp(-\lambda_i(v))dv.$$

By change of variables, $u = t - v$ the equation becomes:

$$P(t) = \exp(-\lambda_i t)\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(u) \times a_{kj} \times \exp(-\lambda_i(t - u))du$$

$$= \exp(-\lambda_i t)\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(u) \times a_{kj}\exp(\lambda_i u) \times \exp(-\lambda_i t)du$$

$$= \left[\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(u) \times a_{kj} \exp(\lambda_i u)du \right] \times \exp(-\lambda_i t),$$

where $\delta_{ij} = 0$, $i \neq j$, and $\delta_{ii} = 1$. The integral is continuous in t since its integrand is bounded on finite intervals, and so $P_{ij}(t)$ is continuous. Then the integrand is continuous and so the derivative of the integrand exists implying that $P_{ij}(t)$ is differentiable in t. Taking the derivatives gives:

$$P'_{ij}(t) = \left[\delta_{ij} + \int_0^t \sum_{k \neq j} P_{ik}(u) \times a_{kj} \times \exp(\lambda_i u) du\right] \times [-\lambda_i \exp(-\lambda_i t)]$$

$$+ \left[\sum_{k \neq j} P_{ik}(t) \times a_{kj} \times \exp(\lambda_i t)\right] \exp(-\lambda_i t)$$

$$= -\lambda_i \times P_{ij}(t) + \sum_{k \neq j} P_{ik}(t) a_{kj} = P_{ij}(t) a_{jj} + \sum_{k \neq j} P_{ik}(t) a_{kj}$$

$$= \sum_{k \in S} P_{ik}(t) a_{kj} = P(t)Q.$$

Here are four examples of the SDEs, $P'_{11}(t)$, $P'_{12}(t)$, $P'_{23}(t)$ and $P'_{34}(t)$

$$a_{21} = a_{31} = a_{41} = a_{14} = a_{24} = a_{42} = a_{43} = a_{44} = 0$$
$$P'_{ij}(t) = \sum_{k \in S} P_{ik}(t) a_{kj} P'_{11}(t) = \sum_{k \in S} P_{1k}(t) a_{k1} = P_{11}(t) a_{11} + P_{12}(t) a_{21}$$
$$+ P_{13}(t) a_{31} + P_{14}(t) a_{41} = -(a_{12} + a_{13}) P_{11}(t) \text{ since } a_{21}$$
$$= a_{31} = a_{41} = 0, and \ a_{11} = -(a_{12} + a_{13})$$
$$= -\lambda_1 P_{11}(t), \text{ since } \lambda_1 = a_{12} + a_{13}.$$

$$P'_{12}(t) = \sum_{k \in S} P_{1k}(t) a_{k2} = P_{11}(t) a_{12}$$
$$+ P_{12}(t) a_{22} + P_{13}(t) a_{32} + P_{14}(t) a_{42} = a_{12} P_{11}(t)$$
$$- \lambda_2 P_{12}(t) + a_{32} P_{13}(t) \text{ since } a_{22} = -\lambda_2, a_{42} = 0.$$

$$P'_{23}(t) = \sum_{k \in S} P_{2k}(t) a_{k3} = P_{21}(t) a_{13} + P_{22}(t) a_{23} + P_{23}(t) a_{33}$$
$$+ P_{24}(t) a_{43} = a_{13} P_{21}(t) + a_{23} P_{22}(t) + a_{33} P_{23}(t)$$
$$+ a_{43} P_{24}(t) = a_{13} P_{21}(t) + a_{23} P_{22}(t) - (a_{32} + a_{34}) P_{23}(t)$$
$$\text{since } a_{43} = 0, a_{33} = -(a_{32} + a_{34}) = a_{13} P_{21}(t)$$
$$+ a_{23} P_{22}(t) - \lambda_3 P_{23}(t), \text{ since } \lambda_3 = a_{32} + a_{34}$$

$$P'_{34}(t) = \sum_{k \in S} P_{3k}(t) a_{k4} = P_{31}(t) a_{14} + P_{32}(t) a_{24} + P_{33}(t) a_{34}$$
$$+ P_{34}(t) a_{44} = P_{33}(t) a_{34} \text{ since } a_{14} = a_{24} = a_{44} = 0 \qquad \square$$

2.3.1 | Probability that the process goes into state j when it leaves state i

The estimated value of transition intensities will enable the determination of the probabilities that the donor will be in state j next given that the blood donor was in state i. If the jump process is currently in state i, the probability that it moves into state j when it leaves the state i is expressed as:

$p_{ij} = \frac{a_{ij}}{\lambda_i}$, where $\lambda_i = \sum_{j \in S} a_{ij}$ for each i and j, such that $i \neq j$ is the total force of transition out of state i.

For example, $p_{12} = \frac{a_{12}}{a_{12} + a_{13}} = \frac{a_{12}}{\lambda_1}$, since $\lambda_1 = a_{12} + a_{13}$.

2.3.2 | Expected holding times

The expected holding time in each state, also known as the mean sojourn time, is the average time an individual donor spends in each state in a single stay before making a transition to another state. This is estimated by $\frac{1}{\lambda_i}$, where $\lambda_i = \sum_{i \neq j} a_{ij}$ is the sum total force of transition out of state i.

2.3.3 | Total time spent in each state

The total waiting time is useful in calculating the estimates of the parameters $\hat{\sigma}$, $\hat{\mu}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\omega}$ using the maximum likelihood method.
    Let:

T_{Ni} = Waiting time of the ith donor in new state
T_{Ri} = Waiting time of the ith donor in regular state
T_{Oi} = Waiting time of the ith donor in occasional state
U_i = Number of transitions New → Regular by the ith donor
V_i = Number of transitions New → Occasional by the ith donor
Z_i = Number of transitions Regular → Occasional by the ith donor
X_i = Number of transitions Occasional → Regular by the ith donor
Y_i = Number of transitions Occasional → Lapsed by the ith donor
We also define totals as:

$$T_N = \sum_{i=1}^N T_{Ni}, \ T_R = \sum_{i=1}^N T_{Ri}, \ T_O = \sum_{i=1}^N T_{Oi},$$

$$U = \sum_{i=1}^N U_i, \ V = \sum_{i=1}^N V_i, \ Z = \sum_{i=1}^N Z_i, \ X = \sum_{i=1}^N X_i,$$

$$Y = \sum_{i=1}^N Y_i.$$

Using lower case symbols for the observed samples, it can be shown that the likelihood given the parameters, $\sigma, \mu, \varepsilon, \gamma, \beta, \omega$ for the model is proportional to:

$$L(\sigma, \mu, \gamma, \beta, \omega) = \prod_{i=1}^N e^{-(\sigma+\mu)T_{Ni}} e^{-(\gamma)T_{Ri}} e^{-(\beta+\omega)T_{Oi}} \sigma^{u_i} \mu^{v_i} \beta^{x_i} \gamma^{z_i} \omega^{y_i}$$

$$= e^{-(\sigma+\mu)\sum T_{Ni}} e^{-\gamma \sum T_{Ri}} e^{-(\beta+\omega)\sum T_{Oi}} \sigma^{\sum u_i} \mu^{\sum v_i} \beta^{\sum x_i} \gamma^{\sum z_i} \omega^{\sum y_i}$$

$$= e^{-(\sigma+\mu)T_N} e^{-\gamma T_R} e^{-(\beta+\omega)T_O} \sigma^u \mu^v \beta^x \gamma^z \omega^y.$$

Where $u = \sum u_i$, $v = \sum v_i$, $z = \sum z_i$, $x = \sum x_i$, $y = \sum y_i$
    The likelihood function $L(\sigma, \mu, \gamma, \beta, \omega)$ for the ith donor reflects:

- The probability of the donor remaining in the new state for total time T_{Ni}, in the regular state for total time T_{Ri} and in the occasional state for total time T_{Oi}, giving the factors $e^{-(\sigma+\mu)T_{Ni}}$, $e^{-(\gamma)T_{Ri}}$ and $e^{-(\beta+\omega)T_{Oi}}$ respectively.
- The probability of the donor making the relevant number of transitions between states, giving factors $\sigma^{u_i}, \mu^{v_i}, \beta^{x_i}, \gamma^{z_i}$ and ω^{y_i}.

Factorizing the likelihood into functions of each parameter gives:

$$L(\sigma, \mu, \gamma, \beta, \omega) = (e^{-\sigma T_N}\sigma^u) \times (e^{-\mu T_N}\mu^v) \times (e^{-\gamma T_R}\gamma^z) \times (e^{-\beta T_O}\beta^x)$$
$$\times (e^{-\omega T_O}\omega^y).$$

Taking logarithms, gives the log-likelihood:

$$\log L = -(\sigma + \mu)T_N - \gamma T_R - (\beta + \omega)T_O + u\log\sigma + v\log\mu$$
$$+ x\log\beta + z\log\gamma + y\log\omega).$$

Differentiating with respect to each of the five parameters gives:

$$\frac{\partial \log L}{\partial \sigma} = -T_N + \frac{u}{\sigma}, \frac{\partial \log L}{\partial \mu} = -T_N + \frac{v}{\mu},$$

$$\frac{\partial \log L}{\partial \beta} = -T_O + \frac{x}{\beta}, \frac{\partial \log L}{\partial \gamma} = -T_R + \frac{z}{\gamma},$$

$$\frac{\partial \log L}{\partial \omega} = -T_O + \frac{u}{\omega}.$$

Equating each derivative to zero and solving the equations gives:

$$\hat{\sigma} = \frac{u}{T_N}, \hat{\mu} = \frac{v}{T_N}, \hat{\beta} = \frac{x}{T_O}, \hat{\gamma} = \frac{z}{T_R}, \hat{\omega} = \frac{y}{T_O}.$$

To obtain the maxima, the Hessian matrix should be negative definite, that is the eigenvalues of the Hessian matrix are all negative. The Hessian matrix is obtained from second derivatives, thus:

$$\frac{\partial^2 \log L}{\partial \sigma^2} = \frac{-u}{\sigma^2}, \frac{\partial^2 \log L}{\partial \mu^2} = \frac{-v}{\mu^2}, \frac{\partial^2 \log L}{\partial \beta^2} = \frac{-x}{\beta^2}, \frac{\partial^2 \log L}{\partial \gamma^2}$$
$$= \frac{-z}{\gamma^2}, \frac{\partial^2 \log L}{\partial \omega^2} = \frac{-y}{\omega^2}.$$

Hence, we consider the matrix

$$\begin{bmatrix} \frac{-u}{\sigma^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{-v}{\mu^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{-x}{\beta^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{-z}{\gamma^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{-y}{\omega^2} \end{bmatrix}.$$

Since this is a negative definite matrix, the maximum likelihood estimates of $\sigma, \mu, \gamma, \beta, \omega$ are:

$$\hat{a}_{12}^{(0)} = \hat{\sigma} = \frac{u}{T_N}, \hat{a}_{13}^{(0)} = \hat{\mu} = \frac{v}{T_N}, \hat{a}_{23}^{(0)} = \hat{\beta} = \frac{x}{T_O}, \hat{a}_{32}^{(0)} = \hat{\gamma} = \frac{z}{T_R},$$

$\hat{a}_{34}^{(0)} = \hat{\omega} = \frac{y}{T_O}$. The $\boldsymbol{0}$ in the superscript has now been added to denote baseline parameter estimates in the absence of covariates. When the covariates are added, the notation becomes $\boldsymbol{a_{ij}}$.

3 | RESULTS

The data is recorded and analyzed as a series of observations for each donor. The variables in the data frame are: donor identification number (donor_ID), time of observation and donor state (state), defined by reference to frequency of donation per annum. In fitting the multistate Markov models, data analysis was conducted using R version 4.0.3 and the *msm* package.[16] This section gives a summary of the results obtained from the analysis of the data based on the model given in Figure 1. The donor transitions, starting from first time or new donors and their evolution overtime is studied.

Donor retention rates are expressed as the percentage of blood donors who return to donate in each subsequent period while attrition is the percentage of blood donors that the blood service loses from one period to another. The donor attrition has a negative impact on the blood inventory, since a decline in numbers of donors has a ripple effect on the volume of blood collections. Furthermore, acquiring new blood donors cost more than retaining the current pool.

Figure 2 below illustrates the donor retention and attrition rates over the 4-year period.

Figure 2 shows a rise in the number of donors who are lost (attrition or solid line) from the initial pool of donors. The graph also shows that the donors who returned for further donations declined with time (retention or dashed line).

According to data sequences from the four states namely: new, regular, occasional, and lapsed, transition frequency or transition count matrix was first calculated and are shown in Table 1.

Table 1 lists the observed transition frequencies of the blood donors. The results show that 248 of the new donors transitioned to regular state, 477 of regular donations repeated their donations as regular donors and 251 of the occasional donors ceased their donations and transitioned to lapsed state by the end of the study.

Regular donors are a reliable source of blood, and their high frequency of blood donations is critical in blood centers. However, with passage of time, frequency of donations declines resulting in high occasions of observed lapsed donors.

The model was then specified as governed by the intensity matrix Q. This was done by calculating the maximum likelihood estimators.[17] Estimates of the parameters: $a_{12}^{(0)}, a_{13}^{(0)}, a_{23}^{(0)}, a_{32}^{(0)}$, and $a_{34}^{(0)}$ with their confidence Intervals (CI) are shown in Table 2. The $\boldsymbol{0}$ in the superscript denotes baseline parameter estimates in the absence of covariates.

However, before calculating vital functions such as transition probabilities matrices, the effect on the parameter estimates for certain individuals was established. Individuals with larger score residuals had a greater effect on the estimates. The score residuals for each blood donor to establish the assertion is presented in Figure 3 below.

Figure 3 shows that there are four donors with residual score above 0.25 and these have large influence on the likelihood estimates.
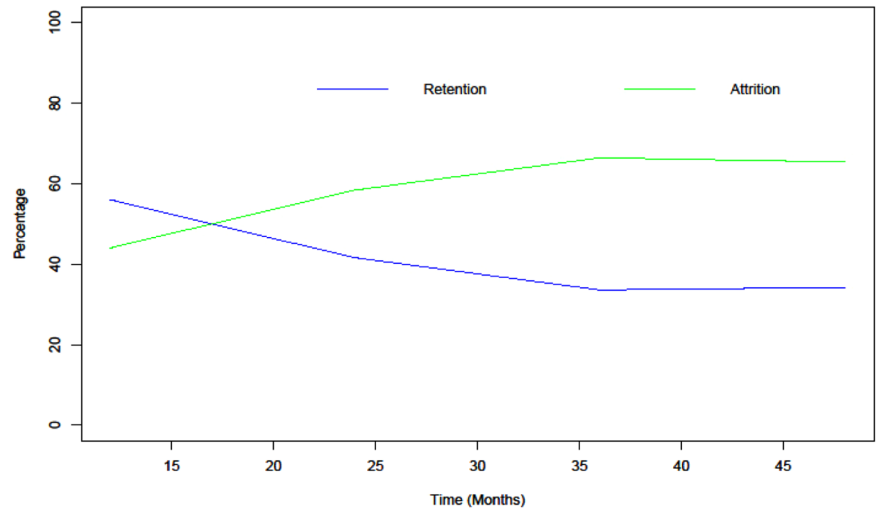
FIGURE 2 Donor retention and attrition rates

TABLE 1 Transition frequency matrix for blood donors' donations

From (i)	To (j)			
	New	**Regular**	**Occasional**	**Lapsed**
New	0	248	187	0
Regular occasional	0	477	170	0
	0	64	0	251

TABLE 2 Time homogeneous Markov model parameter estimates

Transitions	Estimate of $a_{ij}^{(0)}$ (with 95% CI)
State 1–2	1.68394 (9.180e−02, 30.88793)
State 1–3	0.01176 (2.275e−06, 60.75562)
State 2–3	0.03791 (3.334e−02, 0.04311)
State 3–2	0.02412 (1.849e−02, 0.03146)
State 3–4	0.04637 (4.071e−02, 0.05281)
	−2 × log-likelihood: 3751.135

The four donors showing large influence on the estimates were removed and the model refitted. Table 3 below shows a comparison of the estimates of the parameters between the initial model and the refitted model.

Results from Table 3 indicate that the removal of the four most influential donors had a significant effect on the transition intensities.

Using the refitted model, the results show that for first-time blood donors (state 1), there is a higher rate of transition from a new state to a repeat state (1.57118) compared to rate of transition from new state to occasional state (0.01457). Having higher rates for repeat donations is an ideal scenario for blood centers. The results further show that the odds of lapsing from occasional state is more than double (0.04718/0.02114 = 2.23) that of transition to regular state from occasional state. The odds ratio of a transition from state 2 (regular) to state 3 (occasional)

is (0.03657/0.02114 = 1.73) times more than a transition from state 3 (occasional) to state 2 (regular) indicating a loss in the pool of voluntary donors who are a reliable source of blood.

The maximum likelihood estimates are also used to calculate the transition probabilities that the process goes into state j when it leaves state i. The estimated value of transition intensities enabled us to determine the probabilities that the donor will move to state j given that the blood donor is in state i. For example, the probability that state 3 (Occasional donor) is next on condition that the blood donor was in state 2 (Regular donor) is given by:

$$p_{32} = \frac{a_{32}^{(0)}}{a_{32}^{(0)} + a_{34}^{(0)}} = \frac{0.02114}{(0.02114) + (0.04718)} = 0.3094 \text{ as in Table 4.}$$

From Table 4, the first column indicates possible transitions of blood donors from state i to state j. The second column indicates the transition intensities.

Transition intensities are only estimated for the transitions away from the origin state and, hence, for the off-diagonal entries. Larger rates are related to larger transition probabilities away from a state. The results also show a high conditional probability (0.6906) of a donor in state 3 moving to state 4 next and a low probability (0.3094) of a donor in state 3 moving to state 2. This means that blood donors who donate blood occasionally have a higher likelihood of lapsing.

The mean sojourn time gives the holding time or average time in a single stay in a state. Results from Table 5 show that a donor just entering the regular state is expected to spend 27 months (2.25 years) in that state before becoming a sporadic donor. Similarly, a donor just entering the occasional state can expect to spend 14.6 months (1.2 years) at that level before moving to either regular or lapsed state.

The total length of stay in Table 6 gives the expected amount of time spent in each transient state between two future time points. Thus, a blood donor is forecasted to spend 21.193 months (1.8 years) as an erratic blood donor before lapsing. Repeat donors can spend 39.342 months (3.3 years) in the regular state before the transition to other states.
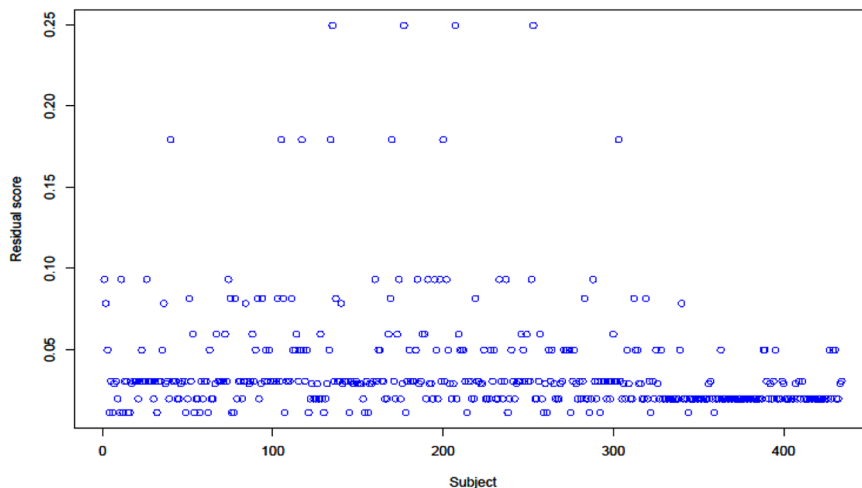
FIGURE 3 Plot of score residuals to establish individual donor effect on likelihood

TABLE 3 Comparison of parameter estimates

Transitions	Model 1 estimates of $a_{ij}^{(0)}$	Refitted model estimates of $a_{ij}^{(0)}$
State 1-2	1.68394	1.57118
State 1-3	0.01176	0.01457
State 2-3	0.03791	0.03657
State 3-2	0.02412	0.02114
State 3-4	0.04637	0.04718
	−2 × log-likelihood: 3751.135 −2 × log-likelihood: 3697.661	

TABLE 4 Maximum likelihood estimates of transition intensities and probability that state j is next

Transitions	Intensities $a_{ij}^{(0)}$	p_{ij}
State 1-2	1.57118	0.9908
State 1-3	0.01457	0.009189
State 2-3	0.03657	1
State 3-2	0.02114	0.3094
State 3-4	0.04718	0.6906
	−2 × log-likelihood: 3697.661	

3.1 | Effects of covariates on transition intensities

This subsection analyzed the effects of donor-specific demographic covariates on blood donation transition intensities. The effects of covariates on the transition intensity for a blood donor i at time j is estimated by a proportional intensities model:

$$a_{ij}(Z) = a_{ij}^{(0)} \exp\left(\beta_{ij}^T Z\right), i \neq j,$$

TABLE 5 Estimates of the sojourn times

Transition	Sojourn time (months)	Standard error	95% Confidence interval
New	0.631	1.023	(0.03, 10.44)
Regular	27.341	1.886	(23.19, 29.99)
Occasional	14.635	0.967	(12.49, 16.12)

TABLE 6 Estimates of the total length of stay in each state

Transition	Total length of stay (months)
New	0.631
Regular	39.342
Occasional	21.193
Lapsed	Infinity

where:

a_{ij} is the estimated transition intensity with covariate Z.

Z is a vector of explanatory variables.

$a_{ij}^{(0)}$ is the baseline intensity function between states i and j.

β_{ij} is a vector of regression parameters.

The covariates are defined and coded as:

$$\text{Gender} = \begin{cases} 1 = \text{Female} \\ 0 = \text{Male} \end{cases},$$

$$\text{Age} = \begin{cases} 0, <40 \text{ years} \\ 1, \geq 40 \text{ years} \end{cases}.$$

The gender male, represented by a zero, acted as the reference category for gender. The reference age category is donors less than 40 years.

The effects of the covariates on the transition intensities, a_{ij}, for blood donor d is given by the following model:

TABLE 7 Estimated covariate effects and their confidence intervals

Transitions	Transition intensities (a_{ij}) (with 95% Confidence interval [CI])	Age coefficient $\beta_{ij}^{(Age_d)}$ (with 95% CI)	Gender coefficient $\beta_{ij}^{(Gender_d)}$ (with 95% CI)
State 1-2	1.403	0.696	0.488
	(0.171, 11.55)	(0.042, 11.54)	(0.017, 14.41)
State 1-3	0.00055	1.874	0.823
	(5.83e−22, 5.11e14)	(1.80e−20, 1.95e20)	(1.98e−25, 3.42e24)
State 2-3	0.037	0.562	1.651
	(0.032, 0.042)	(0.422, 0.748)	(1.259, 2.164)
State 3-2	0.022	1.207	1.435
	(0.016, 0.029)	(0.641, 2.27)	(0.793, 2.594)
State 3-4	0.047	0.907	0.963
	(0.041, 0.054)	(0.657, 1.25)	(0.734, 1.265)

Note: −2 × log-likelihood: 3657.677

TABLE 8 Likelihood ratio test for model models with and without covariates

Model	−2logLR	Degrees of freedom	p
With covariates	39.997	10	0.000017

$$a_{ij} = a_{ij}^{(0)} \exp\left(\beta_{ij}^{(Gender_d)} Gender_d + \beta_{ij}^{(Age_d)} Age_d\right).$$

For this model, the baseline transition intensities, $a_{ij}^{(0)}$, refer to a blood donor with age category = 0 (<40 years old), gender = 0 (Male).

A continuous time Markov model for the effects of covariates, Z = [age, gender], is fitted. The covariates blood group and interval between donations were not statistically significant and hence dropped from the model, leaving age and gender as important demographic factors in this cohort of blood donors. Transition intensities incorporating the effects of the covariates are estimated and the results are shown in Tables 7 and 8.

Results from Table 7 show that new recruits female donor transition rates are 1.6291 [exp (0.488)] times (higher) than their male counterparts rates in transiting to the regular donor status. Similarly, the new recruits older donor (above 40 years of age) transition rates are 2.0057 [exp (0.696)] times (higher) than the rates of their new recruits younger counterparts (below 40 years old). This is quite worrisome to the blood service managers since the sustainability of any blood bank is anchored on young regular donors, yet the older donors had more repeated blood donation episodes. This trend is further shown by occasional older donor transition rates which are 3.3434 [exp (1.207)] times (higher) than the rates of their younger

counterparts (below 40 years of age) in transiting to regular donor status.

Also, female occasional donor transition rates are 4.1996 [exp (1.435)] times (higher) than the male rates in transiting to the regular donor status, thus making female donor a more reliable partner in blood donation drives. However, the occasional older donor transition rates are 2.4769 [exp (0.907)] times (higher) the younger donors' rates in transiting to lapsed donor status. Similarly, the occasional female donor transition rates are 2.6195 times (higher) than their male counterparts in transiting to lapsed donor status. Generally, occasional donors have a higher likelihood of lapsing regardless of the demographic characteristic.

A likelihood ratio test was performed to confirm the improvement. The model with covariates has a log-likelihood function of −2*log-likelihood = 3657.674 which shows an increase of 39.987 compared to −2*log−likelihood = 3697.661 for the model without covariates.

Table 8 confirms the result that the model with covariates fits significantly better than the model without covariates (p = 0.000017 < 0.05).

For the continuous-time homogeneous model with covariates of age and gender estimated above, the percentage prevalence in each of the states are as shown in Figure 4 below.

Figure 4 shows plots of blood donor percentage prevalence in blood donation from the time the donors entered the blood donation system up to the end of the study. At t = 0, the whole system is in state 1, which is the initial state (new donors). From 12 months onwards, none of the blood donors would have remained in state 1 as new donors. All the donors would have moved to other states hence the percentage prevalence of 0. The model underestimates the expected prevalence from the point of entry as new donors up to 12 months.

In state 2 (regular donors), the model gives a good fit of the observed data beyond 12 months up to 48 months. The average percentage prevalence is around 50%, indicating a fair blood donor retention rate. However, there are slight yearly drops in the prevalence in state 2, an indication that regular blood donors reduce their donation frequency with time or cease to donate with the passage of time.

An average of 45% prevalence in state 3 (occasional donors) is seen between 12 and 24 months. The rise can be attributed to new donors from state 1 who do not return to donate after their first-time donation. From 24 months onwards, the decline in percentage prevalence could be attributed to occasional donors in state 3 returning to regular donation or lapsing in their donation, thus moving to state 4. Expected frequencies from 24 months and above gives a good fit of the observed prevalence.

State 4 (lapsed donors) plot indicates a good fit of the observed data. The stepped and steep increase in percentage prevalence for state 4 is a worrisome trend to blood managers. This means that, with the passage of time, more blood donors become inactive or lapse from donating blood due to various reasons over this very short donation period considered (4 years only).
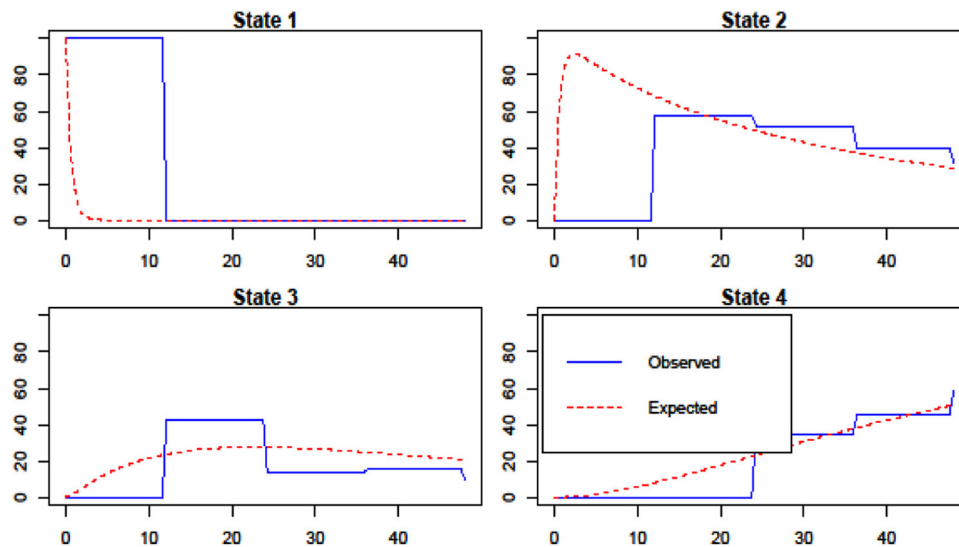
FIGURE 4 Observed and expected prevalence plot for model with age and gender as covariates

4 | DISCUSSION

The study analyses the changes in status of new blood donors who start their blood donation career for the first time until the end of the study. Continuous-time Markov models are fitted on the states: from new to: regular, occasional and lapsed states. The states are defined by the frequency of blood donations in each annual cycle of time. With the passage of time, donor status evolves, more blood donors become inactive or lapsed.

Donor attrition has a negative impact on the blood inventory since a decline in the numbers of donors has a ripple effect on the volume of blood collections. Time in the blood donation career had a noticeable effect on the frequency of blood donation. Results from the study show that the donors who returned for further donations declined with time. In other words, the retention rate decreased while the attrition rate increased. This trend is in agreement with conclusions made by other researchers.[4,5,18,19] This is further evidenced by an increasing prevalence of donors in the lapsed state. To maintain a stable donor pool that would ensure adequate blood supply, blood bank authorities should not focus on the recruitment of new donors only, but also retaining first time donors.[20] The analysis of donor-specific factors is vital in helping blood bank authorities to understand variables to focus on when recruiting potential blood donors in supporting continued blood donation.[21]

Occasional donors have a higher likelihood of lapsing (69.06%) than becoming repeat donors (30.94%). Blood service authorities should target this group with motivational strategies that encourage them to donate blood regularly and prevent lapsing. It is cost effective in time, safety, and resources to motivate this group of occasional donors than to recruit totally new donors.[11,19] In other words, acquiring new blood donors cost more than retaining the current donors.

In consonance with previous literature, this study established the donor-specific factors linked with blood donation intensities namely, age and gender.[22,23] Female donors have higher transition rates from occasional donor status to regular donor status when compared to their male counterparts. Older donors (above 40 years of age) were more consistent in donation than younger donors for both males and females. A study by Veldhuizen et al.[24] investigated the correlation between sociodemographic factors and blood donor prevalence. The conclusion was that, the prevalence in blood donation varied among the age subgroups between males and females. Similarly, donation rates were also particularly higher among the blood group O donors who remained regular donors during the donation career than the other blood groups.

Blood donor recruitment campaigns should be targeted at potential donors with a higher probability of donating and becoming regular donors. The salient characteristics of such donors should be matched with those of current regular donors in the pool.[25,26]

5 | CONCLUSIONS

In this study, a multistate continuous-time Markov model is fitted to blood donation frequencies data. The maximum likelihood parameter estimation method was used to estimate the parameters. The model was improved further with the addition of covariate, age and gender. The 4-year period of study showed that blood donors who return for repeat donations declined with time. Furthermore, the study findings also showed that the odds of donor lapsing were more than double for the occasional state when compared to other states indicating a loss in the pool of voluntary donors who are a reliable source of blood. The analysis and results show that the Markov model is an alternative technique to study and model blood donor status data in Zimbabwe. Furthermore, the study showed that the less studied variables such as blood group and donation time intervals had no statistically significant effects on blood donor transition intensities.

AUTHOR CONTRIBUTIONS

Delson Chikobvu: Supervision; validation; writing – review & editing. **Coster Chideme**: Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; writing – original draft.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author and the National Blood Service Zimbabwe upon reasonable request.

TRANSPARENCY STATEMENT

The lead author Coster Chideme affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ORCID

Coster Chideme http://orcid.org/0000-0002-0624-9136

REFERENCES

1. Gilcher RO, McCombs S. Seasonal blood shortages can be eliminated. *Curr Opin Hematol*. 2005;12(6):503-508. doi:10.1097/01.moh.0000180436.98990.ce

2. Kasraian L, Maghsudlu M. Blood donors' attitudes towards incentives: influence on motivation to donate. *Blood Trans*. 2012;10(2):186-190. doi:10.2450/2011.0039-11

3. Merz EM, Ferguson E, van Dongen A. Psychosocial characteristics of blood donors influence their voluntary nonmedical lapse. *Transfusion*. 2018;58(2596):603-2603. doi:10.1111/trf.14891

4. Gemelli CN, Hayman J, Waller D. Frequent whole blood donors: understanding this population and predictors of lapse. *Transfusion*. 2017;57:108-114. doi:10.1111/trf.13874

5. Schreiber GB, Sanchez AM, Glynn SA, Wright DJ. Increasing blood availability by changing donation patterns. *Transfusion*. 2003;43:591-597. doi:10.1046/j.1537-2995.2003.00388.x

6. Schreiber GB, Sharma UK, Wright DJ, et al. First year donation patterns predict long-term commitment for first-time donors. *Vox Sang*. 2005;88(2):114-121.

7. Vermeulen M, Lelie N, Sykes W, et al. Impact of individual-donation nucleic acid testing on risk of human immunodeficiency virus, hepatitis B virus, and hepatitis C virus transmission by blood transfusion in South Africa. *Transfusion*. 2009;49(6):1115-1125. doi:10.1111/j.1537-2995.2009.02110.x

8. Masser BM, Davison TE, Chapman CM. How can we encourage our voluntary non-remunerated donors to donate more frequently? *VOXS*. 2017;12:112-118. doi:10.1111/voxs.12312

9. Neto CA. Retention of blood donors: strategies to fulfill the requirements of blood centers. *Rev Bras Hematol Hemoter*. 2011;33:174-175. doi:10.5581/1516-8484.20110046.

10. Gillespie TW, Hillyer CD. Blood donors and factors impacting the blood donation decision. *Transfus Med Revl*. 2002;16(2):115-130. doi:10.1053/tmrv.2002.31461

11. Dyantikapm. *Analysis of Blood Donor Behaviour using Hidden Markov Model*. https://medium.com/nerd-for-tech/analysis-of-blood-donor-behavior-using-hidden-markov-model-30ba672ec320

12. Abubakar UY, Hakimi D, Ndanusa A, Jevwia VA. Analysis of blood transfused in a city hospital with the principle of Markov-dependence. *Appl Math Sci*. 2014;8(173):8611-8621. doi:10.12988/ams.2014.410801

13. Bar-Lev SK, Boxma O, Mathijsen B, David P. A blood bank model with perishable blood and demand impatience. *Stochast Syst*. 2017;7(2):237-263. doi:10.1287/stsy.2017.0001

14. Hosseinifard Z, Abbasi B, Fadaki M, Clay NM. Postdisaster volatility of blood donations in an unsteady blood supply chain. *Decision Sci*. 2020;51:255-281. doi:10.1111/deci.12381

15. World Health Organisation. *Blood Donor Selection: Guidelines on Assessing Donor Suitability for Blood Donation*. World Health Organization; 2012. http://apps.who.int/iris/handle/10665/76724/

16. Jackson CH. Multi-state models for panel data: the msm package for R. *J Stat Softw*. 2011;38(8):1-29. http://www.jstatsoft.org/v38/i08/

17. Kalbfleisch J, Lawless J. The analysis of panel data under a Markov assumption. *J Am Stat Assoc*. 1985;80(392):863-871. doi:10.2307/2288545

18. Mohammed S, Essel B. Motivational factors for blood donation, potential barriers, and knowledge about blood donation in first-time and repeat blood donors. *BMC Hematol*. 2018;18:36. doi:10.1186/s12878-018-0130-3

19. Kasraian L. Causes of discontinuity of blood donation among donors in Shiraz, Iran: cross-sectional study. *Sao Paulo Med J*. 2010;128(5):272-275. doi:10.1590/S1516-31802010000500006

20. Duboz P, Cunéo B. Impact of socioeconomic status on blood donation. *Transfus Clin Biol*. 2009;16(4):371-378. doi:10.1016/j.tracli.2009.02.003

21. Yu PL, Chung KH, Lin CK, Chan JS, Lee CK. Predicting potential drop-out and future commitment for first-time donors based on first 1.5-year donation patterns: the case in Hong Kong Chinese donors. *Vox Sang*. 2007;93(1):57-63. doi:10.1111/j.1423-0410.2007.00905.x

22. Gemelli C, Hayman J. Frequent whole blood donors: understanding this population and predictors of lapse: profiling frequent blood donors. *Transfusion*. 2016;57:108-114. doi:10.1111/trf.13874.

23. Ou Y, Yau KK, Poon CM, Hui YV, Lee SS, Lee CK. Donation frequency and its association with demographic characteristics a 1 year observational study. *Transfus Med*. 2015;25(6):366-373.

24. Veldhuizen IJ, Doggen CJ, Atsma F, De Kort WL. Donor profiles: demographic factors and their influence on the donor career. *Vox Sang*. 2009;97(2):129-138.

25. Burgdorf KS, Simonsen J, Sundby A, et al. Socio-demographic characteristics of Danish blood donors. *PLoS One*. 2017;12(2):e0169112. doi:10.1371/journal.pone.0169112

26. Glynn SA, Kleinman SH, Schreiber GB, et al. Motivations to donate blood: demographic comparisons. *Transfusion*. 2002;42:216-225. doi:10.1046/j.1537-2995.2002.00008.x