Cortés, J.; Navarro-Quiles, A.; Romero, J.; Roselló, M. (2017). Randomizing the parameters of a Markov chain to model the stroke disease: A technical generalization of established computational methodologies towards improving real applications. Journal of Computational and Applied Mathematics. 324:225-240. doi:10.1016/j.cam.2017.04.040

Randomizing the parameters of a Markov chain to model the stroke disease: A technical generalization of established computational methodologies towards improving real applications

J.-C. Cortés[a], A. Navarro-Quiles[a], J.-V. Romero[a,*], M.-D. Roselló[a]

[a]*Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*

Abstract

Classical Markov models are defined through a stochastic transition matrix, i.e., a matrix whose columns (or rows) are deterministic values representing transition probabilities. However, in practice these quantities could often not be known in a deterministic manner, therefore, it is more realistic to consider them as random variables. Following this approach, this paper is aimed to give a technical generalization of classical Markov methodology in order to improve modelling of stroke disease when dealing with real data. With this goal, we randomize the entries of the transition matrix of a Markov chain with three states (susceptible, reliant and deceased) that has been previously proposed to model the stroke disease. This randomization of the classical Markov model permits the computation of the first probability density function of the solution stochastic process taking advantage of the so-called Random Variable Transformation technique. Afterwards, punctual and probabilistic predictions are computed from the first probability density function. In addition, the probability density functions of the time instants until a certain proportion of the total population remains susceptible, reliant and deceased are also computed. The study is completed showing the usefulness of our computational approach to determine, from a probabilistic point of view, key quantities in medical decision making, such as the cost-effectiveness ratio.

Keywords: Markov process, disease modelling, random variable transformation technique, computing the first probability density function

1. Introduction

Discrete Markov stochastic processes (s.p.'s) or discrete Markov chains are often applied to model the dynamics of medical events over evenly spaced times, $n = 0, 1, 2, \ldots$, usually referred to as periods or cycles. In particular, these kind of s.p.'s have been considered for different purposes. For example, to built and simulate models for chronic illnesses [1, 2], to analyse

*Corresponding author

Email addresses: jccortes@imm.upv.es (J.-C. Cortés), annaqui@doctor.upv.es (A. Navarro-Quiles), jvromero@imm.upv.es (J.-V. Romero), drosello@imm.upv.es (M.-D. Roselló)

data of hospital infection [3], to provide predictions based on random-effects Markov models applied to multiple sclerosis progression [4], to calculate the prevalence of certain diseases and to perform budget impact analysis [5], to model human papilloma virus [6], etc. In these models individuals are classified in several disjoint classes or states. The evolution of the percentage (or number of individuals) in each cycle n is determined by the initial distribution of the individuals and a stochastic matrix, usually termed transition matrix. An important assumption of standard Markov models is that all the states make up a closed system. This means that any individuals can neither leave nor join the system, hence having a constant population size over the time. This hypothesis holds in the clinical context where doctors and public health authorities are often interested in the evolution of patient groups in controlled studies over the time. In the case of discrete homogeneous Markov s.p.'s, the transition matrix is a constant matrix whose entries represent the probabilities to change either from one state to another or to remain in the same state between two consecutive cycles. When these probabilities depend upon time, the s.p. is termed non-homogeneous time discrete Markov chains. In both cases, the entries of the transition matrix are assumed to be deterministic quantities (numbers or functions, respectively). In this paper, we generalize this feature for discrete homogeneous Markov chains by considering that the entries in the transition matrix could be random variables (r.v.'s) rather than deterministic constants. Naturally, the r.v.'s are assumed to take values in the interval $[0, 1]$, thus representing probabilities for every realization of such r.v.'s. In this manner, we allow for more flexibility when probabilities are assigned. Throughout this paper, we will consider this approach to generalize the stroke disease model proposed in [7]. It is important to point out that the application of our approach is not limited to the stroke disease model presented later but is also valid for modelling any disease via time discrete homogeneous Markov chains.

The manuscript is organized as follows. In Section 2 we introduce the mathematical stroke model that will be considered throughout this paper. Section 3 is addressed to give the mathematical tools that will be required to study the stroke model. The reader more interested in medical results, can skip this section in a first reading. Section 4 is devoted to provide a probabilistic solution of the randomized Markov model (1)–(2) by means of the first probability density function for each subpopulation, susceptible, reliant and deceased. For the sake of clarity we have divided this section into two subsections. First, in Subsection 4.1 the main statistical properties, such as, the mean, the variance and confidence intervals, are computed. Secondly, Subsection 4.2 is addressed to obtain the probabilistic distribution until a given proportion of the population remains susceptible. In Section 5 we perform a cost-effectiveness analysis taking advantage of the mathematical tools presented in Section 3. In Section 6, all the theoretical results developed throughout Sections 4 and 5 are applied to simulate the stroke disease taking particular distributions for the random input data that are in agreement with the extant literature. In Section 7, we discuss our main findings.

2. Motivating problem: the stroke disease

Markov/semi-Markov models have demonstrated to be useful mathematical representations to model diseases. In particular, this approach has been successfully applied to study the stroke disease using different statistical techniques [7, 8, 9]. According to [7], the stroke disease can be modelled via a Markov chain considering the three following states, Susceptible (S), Reliant (R) and Deceased (D). In Figure 1 we show the influence or flow diagram associated to the Markov model. In this graphical representation, transitions among states have been included. We observe that, apart from remaining in each state, the possible transitions between states are

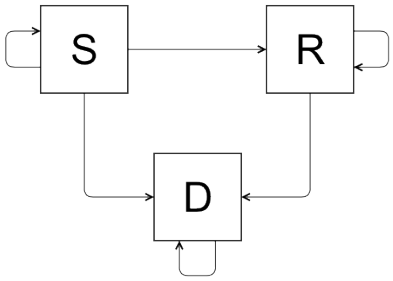

Figure 1: Flow diagram for the Markov model (1)–(2). S, R and D stand for Susceptible (S), Reliant (R) and Deceased (D), respectively.

$S \to R$, $S \to D$ and $R \to D$. Thus, the reliant population cannot recover from the disease. Obviously, the state D is an absorbing state. In this study the susceptible individuals make up a population at risk, i.e., they have certain pathologies (hypertension, cholesterol, etc.) that may conduct to suffer a stroke. Therefore, as we shall see below, the model involves a relative risk.

In [7], the Markov model is formulated as follows

$$\begin{pmatrix} S_{n+1} \\ R_{n+1} \\ D_{n+1} \end{pmatrix} = T \begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix}, \quad (S_0, R_0, D_0)^\top = (s_0, r_0, d_0)^\top, \quad n = 0, 1, 2, \ldots,$$

where S_n, R_n and D_n are the proportion of susceptible, reliant and deceased subpopulations in cycle n, respectively. As a matter of fact in dealing with markovian models, we will assume that $S_n + R_n + D_n = 1$ for each n. As it is plausible from a practical standpoint, we assume that initially there are no deaths, hence the initial cohort corresponds with the deterministic vector $(s_0, r_0, 0)^\top$, $s_0 + r_0 = 1$. Otherwise, the subsequent analysis follows analogously. Moreover, according to results given in [7], we shall assume that the transition matrix T is given by

$$T = \begin{pmatrix} \mathrm{e}^{-t_1 rr} + \mathrm{e}^{-(t_2 + t_3(rr-1))} - 1 & 0 & 0 \\ 1 - \mathrm{e}^{-t_1 rr} & 1 - p & 0 \\ 1 - \mathrm{e}^{-(t_2 + t_3(rr-1))} & p & 1 \end{pmatrix},$$

being

- rr is the relative risk of suffering a stroke,
- t_1 is the non-mortal stroke rate,
- t_2 is the deceased rate due to any cause,
- t_3 is the stroke death rate and,
- p is the probability of the transition $R \to D$,

where the rates t_1, t_2 and t_3 are given for a general population. For sake of clarity, we now explain the construction of the transition matrix, T, in connection with [7, Table 3] and the meaning of the parameters previously introduced. The element $(2, 1)$ of matrix T, T_{21}, represents the probability

of suffering a non-mortal stroke in the cycle $n+1$ given that the individual was susceptible in the cycle n ($S \to R$). The probability of having a stroke is given by 1 minus the probability of does not have it, being these kind of probabilities usually modelled by an exponential decay. In Table 3 of [7] this probability is given by $1 - \mathrm{e}^{-(\text{"non-mortal stroke rate"})}$. The "non-mortal stroke rate" is given by t_1 and taking into account that we are dealing with a population under risk, this leads to the term $t_1 rr$. T_{31} denotes the probability of the transition $S \to D$. In Table 3 of [7], this probability is given by $1 - \mathrm{e}^{-(\text{"death rate"})}$. Observe that the "death rate" involved in T_{31} is given by the $(t_2 + t_3(rr-1)) = t_2 - t_3 + t_3\, rr$, that is, the non-stroke death rate for a general population, $t_2 - t_3$, adding the term corresponding to the stroke death rate for the population under risk given by $t_3\, rr$.

At this point it is important to remark that the parameter rr though is termed *relative* risk, in the context of medicine is a positive number [10, 11].

Remark 1. From a mathematical standpoint the parameters t_1, t_2, t_3 and rr must satisfy the condition $0 < T_{21} + T_{31} < 1$. This guarantees that $T_{11} \in]0,1[$. As in practice the rates t_1, t_2 and t_3 are small, former condition holds.

As it has been pointed out previously, a major difference with respect to contribution [7] is that we will assume that some model parameters, namely, t_2, rr and p, involved in transition matrix T are absolutely continuous r.v.'s rather than deterministic constants. Hereinafter, as usual in Probability Theory, capital letters will be used to highlight this difference. Hence, the following identifications, $t_2 \Rightarrow T_2$, $rr \Rightarrow RR$ and $p \Rightarrow P$, will be used (see expression (2)). This decision is motivated inasmuch as, in practice, the death rate due to any cause, T_2, is not known in a deterministic way and the relative risk of suffering a stroke, RR, varies among physical characteristic of individuals being this variation non-deterministic. Regarding parameter P, which represents a probability, we assume that it can be described by a r.v. whose domain is contained in the interval $]0,1[$, allowing for more flexibility throughout the study. In the following, the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ will denote the common complete probability space where r.v.'s T_2, RR and P are defined.

Summarizing the model that we are going to study is

$$\begin{pmatrix} S_{n+1} \\ R_{n+1} \\ D_{n+1} \end{pmatrix} = T \begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix}, \quad (S_0, R_0, D_0)^\top = (s_0, r_0, 0)^\top, \quad n = 0, 1, 2, \ldots, \tag{1}$$

where the transition matrix is given by

$$T = \begin{pmatrix} \mathrm{e}^{-t_1 RR} + \mathrm{e}^{-(T_2 + t_3(RR-1))} - 1 & 0 & 0 \\ 1 - \mathrm{e}^{-t_1 RR} & 1 - P & 0 \\ 1 - \mathrm{e}^{-(T_2 + t_3(RR-1))} & P & 1 \end{pmatrix}. \tag{2}$$

In connection with Remark 1 and, in the random context, to guarantee the positiveness of the entry $T_{11}(\omega)$ of random matrix (2), it must be imposed that r.v.'s RR and T_2 satisfy the following condition

$$\mathbb{P}\left[0 < T_{21}(\omega) + T_{31}(\omega) = 2 - \mathrm{e}^{-t_1 RR(\omega)} - \mathrm{e}^{-(T_2(\omega) + t_3(RR(\omega)-1))} < 1\right] = 1, \quad \forall \omega \in \Omega. \tag{3}$$

To conduct our study, the so-called Random Variable Transformation (RVT) method will be used [12]. This technique has been successfully applied in previous contributions related to epidemiological models, some examples include [13, 14, 15, 16]. RVT method allows us to obtain

the first probability density function (1-p.d.f.) of the solution s.p.'s, S_n, R_n, D_n, to model (1)–(2). Additionally, we will compute the p.d.f.'s of times until a given proportion of the population remains susceptible, reliant and deceased, respectively. Finally, the p.d.f. of the cost-effectiveness ratio will be also computed taking advantage of RVT technique. This is a key quantity in medical decision making.

3. Mathematical tools

We start stating the key mathematical tool, usually referred to as Random Variable Transformation (RVT) method, that we will used in Sections 4 and 5 to give a full probabilistic solution of the disease stroke markovian model formulated in (1)–(2).

Theorem 1. *(Multidimensional version, [12, pp. 24–25]). Let* $\mathbf{U} = (U_1, \ldots, U_n)^\top$ *and* $\mathbf{V} = (V_1, \ldots, V_n)^\top$ *be two n-dimensional absolutely continuous random vectors. Let* $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ *be a one-to-one deterministic transformation of* $\mathbf{U}$ *into* $\mathbf{V}$*, i.e.,* $\mathbf{V} = \mathbf{g}(\mathbf{U})$*. Assume that* $\mathbf{g}$ *is continuous in* $\mathbf{U}$ *and has continuous partial derivatives with respect to* $\mathbf{U}$*. Then, if* $f_{\mathbf{U}}(\mathbf{u})$ *denotes the joint probability density function of vector* $\mathbf{U}$*, and* $\mathbf{h} = \mathbf{g}^{-1} = (h_1(v_1, \ldots, v_n), \ldots, h_n(v_1, \ldots, v_n))^\top$ *represents the inverse mapping of* $\mathbf{g} = (g_1(u_1, \ldots, u_n), \ldots, g_n(u_1, \ldots, u_n))^\top$*, the joint probability density function of vector* $\mathbf{V}$ *is given by*

$$f_{\mathbf{V}}(\mathbf{v}) = f_{\mathbf{U}}\left(\mathbf{h}(\mathbf{v})\right) |J|\,, \tag{4}$$

where $|J|$ *is the absolute value of the Jacobian, which is defined by*

$$J = \det\left(\frac{\partial \mathbf{h}^\top}{\partial \mathbf{v}}\right) = \det\begin{pmatrix} \frac{\partial h_1(v_1,\ldots,v_n)}{\partial v_1} & \cdots & \frac{\partial h_n(v_1,\ldots,v_n)}{\partial v_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1(v_1,\ldots,v_n)}{\partial v_n} & \cdots & \frac{\partial h_n(v_1,\ldots,v_n)}{\partial v_n} \end{pmatrix}. \tag{5}$$

In the subsequent subsection the RVT method is applied to determine the 1-p.d.f. of the solution s.p. to stroke model (1)–(2). This function will permit later to compute important statistical properties of the solution to the stroke model, namely, the mean and standard deviation functions of susceptible, reliant and deceased subpopulations. In addition, RVT technique will play a key role to compute the p.d.f. of the time until a given proportion of the population remains in some of the three states (susceptible, reliant and deceased). This will be illustrated later.

3.1. First probability density function

As it has been said previously, the goal of this subsection is to obtain the 1-p.d.f. of the number of susceptibles, reliants and deceaseds, which are the components of the solution s.p. of the random initial value problem (1)–(2). This will be done in terms of the random input data. For the sake of generality as it has been indicated previously, throughout this subsection RR, T_2 and P are assumed to be absolutely continuous dependent r.v.'s, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with joint p.d.f. $f_{RR,T_2,P}(rr, t_2, p)$ defined on a domain, say $\mathcal{D}_{RR,T_2,P}$. It generalizes the case where RR, T_2 and P are assumed to be independent r.v.'s with p.d.f.'s $f_{RR}(rr)$, $f_{T_2}(t_2)$ and $f_P(p)$, since in that case $f_{RR,T_2,P}(rr, t_2, p) = f_{RR}(rr) f_{T_2}(t_2) f_P(p)$. Although less general, independence is a hypothesis usually embraced in probabilistic applications.

As it is well known, the solution of (1)– (2) is

$$\begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix} = T^n \begin{pmatrix} s_0 \\ r_0 \\ 0 \end{pmatrix}. \tag{6}$$

To conduct our study it is convenient to recast the entries of the transition matrix T as follows

$$T = \begin{pmatrix} 1-K-Q & 0 & 0 \\ K & 1-P & 0 \\ Q & P & 1 \end{pmatrix}, \tag{7}$$

where $K = 1 - \mathrm{e}^{-t_1 RR}$ and $Q = 1 - \mathrm{e}^{-(T_2 + t_3(RR-1))}$. Then, developing the right-hand side of (6) one gets

$$\begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix} = \begin{pmatrix} (1-K-Q)^n s_0 \\ \dfrac{(1-K-Q)^n K s_0 - (1-P)^n (r_0(-P+Q+K) + K s_0)}{P-Q-K} \\ r_0 - r_0(1-P)^n + \dfrac{(P-Q+(-P+Q)(1-Q-K)^n + (-1+(1-P)^n)K)s_0}{P-Q-K} \end{pmatrix}. \tag{8}$$

Notice that as P, Q and K are absolutely continuous r.v.'s, the denominator of second and third components of expression (8) are non-zero with probability 1. Taking into account $S_n + R_n + D_n = 1$ for each n, it is enough to determine the 1-p.d.f. of susceptible and reliant subpopulations, since from them, it is straightforward to obtain the 1-p.d.f. of deceased subpopulation.

This goal will be achieved by applying RVT method twice. First, we will compute the joint p.d.f., $f_{S_n,R_n,P}(s,r,p)$, of random vector (S_n, R_n, P) from the joint p.d.f., $f_{K,Q,P}(k,q,p)$, of (K,Q,P), and secondly, we will compute the joint p.d.f., $f_{K,Q,P}(k,q,p)$, of (K,Q,P) from the joint p.d.f., $f_{RR,T_2,P}(rr,t_2,p)$, of random input data (RR, T_2, P).

Now, we fix the cycle n and use the RVT method. Then, we apply Theorem 1 with the following identifications

$$\mathbf{U} = (K,Q,P)^\top, \quad \mathbf{V} = (V_1,V_2,V_3), \quad \mathbf{V} = \mathbf{g}(\mathbf{U})^\top,$$

$$\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{g}(k,q,p) = (g_1(k,q,p), g_2(k,q,p), g_3(k,q,p))^\top = (v_1,v_2,v_3)^\top,$$

being

$$v_1 = (1-k-q)^n s_0, \quad v_2 = \frac{(1-k-q)^n k s_0 - (1-p)^n (r_0(-p+q+k) + k s_0)}{p-q-k}, \quad v_3 = p.$$

Isolating k, q and p one gets

$$\begin{aligned} k &= \frac{\left(-1+v_3+\left(\frac{v_1}{s_0}\right)^{\frac{1}{n}}\right)(r_0(1-v_3)^n - v_2)}{(1-v_3)^n s_0 - v_1}, \\ q &= \frac{v_1 - v_1\left(\frac{v_1}{s_0}\right)^{\frac{1}{n}} + (-s_0 + (r_0+s_0)\left(\frac{v_1}{s_0}\right)^{\frac{1}{n}} + r_0(-1+v_3))(1-v_3)^n - v_2\left(-1+\left(\frac{v_1}{s_0}\right)^{\frac{1}{n}} + v_3\right)}{v_1 - s_0(1-v_3)^n}, \\ p &= v_3. \end{aligned} \tag{9}$$

For the sake of clarity, hereinafter we will consider $m_1 := k$ and $m_2 := q$.

Notice that the Jacobian of mapping $\mathbf{g}^{-1}$ is

$$J = \frac{\left(\frac{v_1}{s_0}\right)^{1/n}\left(-1+\left(\frac{v_1}{s_0}\right)^{1/n}+v_3\right)}{nv_1(v_1-s_0(1-v_3)^n)}.$$

Then, taking into account (4)–(5), the joint p.d.f. of random vector $(V_1, V_2, V_3) = (S_n, R_n, P)$ is given by

$$f_{S_n,R_n,P}(v_1,v_2,v_3) = f_{K,Q,P}(m_1,m_2,v_3)\left|\frac{\left(\frac{v_1}{s_0}\right)^{1/n}\left(-1+\left(\frac{v_1}{s_0}\right)^{1/n}+v_3\right)}{nv_1(v_1-s_0(1-v_3)^n)}\right|. \tag{10}$$

Using again Theorem 1 with the following identifications

$$\mathbf{U} = \hat{\mathbf{U}} = (RR, T_2, P)^\top, \quad \mathbf{V} = \hat{\mathbf{V}} = (\hat{V}_1, \hat{V}_2, \hat{V}_3), \quad \hat{\mathbf{V}} = \mathbf{g}(\hat{\mathbf{U}})^\top,$$

$$\mathbf{g}: \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{g}(rr, t_2, p) = (g_1(rr, t_2, p), g_2(rr, t_2, p), g_3(rr, t_2, p))^\top = (\hat{v}_1, \hat{v}_2, \hat{v}_3)^\top,$$

being

$$\hat{v}_1 = 1 - \mathrm{e}^{-t_1 rr}, \quad \hat{v}_2 = 1 - \mathrm{e}^{-(t_2 + t_3(rr-1))}, \quad \hat{v}_3 = p,$$

and isolating rr, t_2 and p, one gets

$$rr = \frac{-\log(1-\hat{v}_1)}{t_1}, \quad t_2 = t_3 + \frac{t_3\log(1-\hat{v}_1)}{t_1} - \log(1-\hat{v}_2), \quad p = \hat{v}_3.$$

Moreover, the Jacobian of the mapping $\mathbf{g}^{-1}$ is

$$J = \frac{1}{t_1(1-\hat{v}_1)(1-\hat{v}_2)}.$$

Then, taking into account (4)–(5), the joint p.d.f. of random vector $(\hat{V}_1, \hat{V}_2, \hat{V}_3) = (K, Q, P)$ is given by

$$\begin{aligned} f_{K,Q,P}(\hat{v}_1,\hat{v}_2,\hat{v}_3) = \quad & f_{RR,T_2,P}\left(\frac{-\log(1-\hat{v}_1)}{t_1}, t_3 + \frac{t_3\log(1-\hat{v}_1)}{t_1} - \log(1-\hat{v}_2), \hat{v}_3\right) \\ & \times\left|\frac{1}{t_1(1-\hat{v}_1)(1-\hat{v}_2)}\right|. \end{aligned} \tag{11}$$

Compounding both (10) and (11), we determine the joint p.d.f. of (S_n, R_n, P) using the p.d.f. of the random vector (RR, T_2, P)

$$\begin{aligned} f_{S_n,R_n,P}(s,r,p) \quad = \quad & f_{RR,T_2,P}\left(\frac{-\log(1-m_1)}{t_1}, t_3 + \frac{t_3\log(1-m_1)}{t_1} - \log(1-m_2), p\right) \\ & \times\left|\frac{1}{t_1(1-m_1)(1-m_2)}\right|\left|\frac{\left(\frac{s}{s_0}\right)^{1/n}\left(-1+\left(\frac{s}{s_0}\right)^{1/n}+p\right)}{ns(s-s_0(1-p)^n)}\right|, \end{aligned} \tag{12}$$

where m_1 and m_2 are the expressions introduced in (9) and below using the identifications $v_1 \Rightarrow s$, $v_2 \Rightarrow r$ and $v_3 \Rightarrow p$.

Finally, considering n arbitrary and marginalizing (12), we obtain the 1-p.d.f.'s of the subpopulation of susceptibles, $f_1(s,n)$ and reliants, $f_1(r,n)$,

$$\begin{aligned} f_1(s,n) &= \int_{\mathcal{D}_{R_n,P}} f_{RR,T_2,P}\left(-\frac{\log(1-m_1)}{t_1}, t_3 + \frac{t_3\log(1-m_1)}{t_1} - \log(1-m_2), p\right) \\ &\quad \times \left|\frac{1}{t_1(1-m_1)(1-m_2)}\right| \left|\frac{\left(\frac{s}{s_0}\right)^{1/n}\left(-1+\left(\frac{s}{s_0}\right)^{1/n}+p\right)}{ns\left(s-s_0\left(1-p\right)^n\right)}\right| \mathrm{d}p\,\mathrm{d}r, \end{aligned} \tag{13}$$

$$\begin{aligned} f_1(r,n) &= \int_{\mathcal{D}_{S_n,P}} f_{RR,T_2,P}\left(-\frac{\log(1-m_1)}{t_1}, t_3 + \frac{t_3\log(1-m_1)}{t_1} - \log(1-m_2), p\right) \\ &\quad \times \left|\frac{1}{t_1(1-m_1)(1-m_2)}\right| \left|\frac{\left(\frac{s}{s_0}\right)^{1/n}\left(-1+\left(\frac{s}{s_0}\right)^{1/n}+p\right)}{ns\left(s-s_0\left(1-p\right)^n\right)}\right| \mathrm{d}p\,\mathrm{d}s. \end{aligned} \tag{14}$$

Now, we will provide the 1-p.d.f. of the deceased subpopulation, using that $D_n = 1 - S_n - R_n$, for each cycle n. To this end, first we apply Theorem 1 again considering the following identifications

$$\mathbf{U} = (S_n, R_n)^\top, \quad \mathbf{V} = (V_1, V_2), \quad \mathbf{V} = \mathbf{g}(\mathbf{U})^\top,$$

$$\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathbf{g}(s,r) = (g_1(s,r), g_2(s,r))^\top = (v_1, v_2)^\top,$$

being

$$v_1 = 1 - s - r, \quad v_2 = r.$$

Isolating s and r one gets

$$s = 1 - v_1 - v_2, \quad r = v_2.$$

Then, taking into account (4)–(5) and that the Jacobian takes the value -1, we obtain the joint p.d.f. of random vector $(V_1, V_2) = (D_n, R_n)$. Finally, marginalizing this latter joint p.d.f. and considering n arbitrary, it can be checked that the 1-p.d.f. of the deceased subpopulation is given by

$$\begin{aligned} f_1(d,n) &= \int_{\mathcal{D}_{R_n,P}} f_{RR,T_2,P}\left(-\frac{\log(1-m_1)}{t_1}, t_3 + \frac{t_3\log(1-m_1)}{t_1} - \log(1-m_2), p\right) \\ &\quad \times \left|\frac{1}{t_1(1-m_1)(1-m_2)}\right| \left|\frac{\left(\frac{1-d-r}{s_0}\right)^{1/n}\left(-1+\left(\frac{1-d-r}{s_0}\right)^{1/n}+p\right)}{n(1-d-r)\left((1-d-r)-s_0\left(1-p\right)^n\right)}\right| \mathrm{d}p\,\mathrm{d}r. \end{aligned} \tag{15}$$

4. A full probabilistic solution of the stroke markovian model

The 1-p.d.f. gives a full probabilistic description in each period n of the solution s.p. of a markovian model. In particular, the 1-p.d.f. of susceptibles, reliants and deceaseds to the stroke markovian model (1)–(2) are given by (13), (14) and (15), respectively. It is important to point out that all these expressions are given by closed-form formulas.

Moreover, from the 1-p.d.f., both the mean and variance functions can be straightforwardly computed for every state of the model. This information is crucial in order to provide punctual and probabilistic predictions. In medical practice it is also important to know when the percentage of susceptibles, reliants and deceaseds in the population will achieve a certain level. This information can be determined by means of the 1-p.d.f. as well. These ideas motivate the following two subsections.

4.1. Mean and variance functions. Confidence intervals

Hereinafter, we will focus on susceptible subpopulation, S_n, whose 1-p.d.f. is given by (13), although the following development can be extrapolated to reliant and deceased subpopulations, using (14) and (15), respectively. The expressions for the mean and the variance functions are

$$\mu_{S_n} = \mathbb{E}[S_n] = \int_{\mathcal{D}_{S_n}} s f_1(s,n)\,\mathrm{d}s, \qquad \sigma^2_{S_n} = \mathbb{V}[S_n] = \int_{\mathcal{D}_{S_n}} s^2 f_1(s,n)\,\mathrm{d}s - (\mu_{S_n})^2, \tag{16}$$

respectively.

Furthermore, the 1-p.d.f. is useful to construct confidence intervals. Let $\alpha \in (0,1)$ and $\hat{n}$ fixed, one can determine $s_1 = s_1(\hat{n})$ and $s_2 = s_2(\hat{n})$ such that

$$\int_0^{s_1} f_1(s,\hat{n})\,\mathrm{d}s = \frac{\alpha}{2} = \int_{s_2}^1 f_1(s,\hat{n})\,\mathrm{d}s. \tag{17}$$

Then, $(1-\alpha)\times 100\%$-confidence interval is specified by

$$1-\alpha = \mathbb{P}(\{\omega \in \Omega : S(\hat{n};\omega) \in [s_1, s_2]\}) = \int_{s_1}^{s_2} f_1(s,\hat{n})\,\mathrm{d}s. \tag{18}$$

In addition, it is of interest for doctors knowing the probability, for example, that the proportion of susceptible subpopulation lies between a and b at a specific time period, say $\hat{n}$,

$$\mathbb{P}[a \le S_{\hat{n}} \le b] = \int_a^b f_1(s,\hat{n})\,\mathrm{d}s. \tag{19}$$

4.2. Distribution of time until a given proportion of the population remains susceptible, reliant or deceased

In practice, it is useful to know when the percentage of susceptibles, reliants and deceaseds in the population will attain a certain level. This motivates the computation, in a first step, of the distribution, N_S, of the time until a given proportion of the population, ρ_S, remains susceptible. The same can be said for reliant and deceased subpopulations.

In order to compute the p.d.f. of N_s for a fixed proportion of susceptibles, ρ_S, we first isolate $n = N_S$ from the first component of the exact solution, given by (8), of the initial value problem (1)–(2)

$$N_S = \frac{\log\left(\frac{\rho_S}{s_0}\right)}{\log\left(\mathrm{e}^{-t_1 RR} + \mathrm{e}^{-(T_2+t_3(RR-1))} - 1\right)}. \tag{20}$$

Notice that expression (20) depends only on r.v.'s RR and T_2. Hence we apply RVT technique, i.e., Theorem 1 to

$$\mathbf{U} = (RR, T_2)^\top, \quad \mathbf{V} = (V_1, V_2), \quad \mathbf{V} = \mathbf{g}(\mathbf{U})^\top,$$
$$\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathbf{g}(rr, t_2) = (g_1(rr, t_2), g_2(rr, t_2))^\top = (v_1, v_2)^\top,$$

being

$$v_1 = rr, \quad v_2 = \frac{\log\left(\frac{\rho_S}{s_0}\right)}{\log\left(\mathrm{e}^{-t_1 rr} + \mathrm{e}^{-(t_2 + t_3(rr-1))} - 1\right)}.$$

Isolating rr and t_2, one gets

$$rr = v_1, \quad t_2 = t_3(1 - v_1) - \log\left[\left(\frac{\rho_S}{s_0}\right)^{1/v_2} + 1 - \mathrm{e}^{-t_1 v_1}\right].$$

The Jacobian of the inverse of the mapping $\mathbf{g}(\mathbf{U})^\top$, is given by

$$J = \frac{\left(\frac{\rho_S}{s_0}\right)^{1/v_2} \log\left(\frac{\rho_S}{s_0}\right)}{v_2^2\left(1 - \mathrm{e}^{-t_1 v_1} + \left(\frac{\rho_S}{s_0}\right)^{1/v_2}\right)}. \tag{21}$$

Then, taking into account (4)–(5) we obtain the joint p.d.f. of random vector $(V_1, V_2) = (RR, N_s)$. Finally, marginalizing with respect to r.v. RR, the expression of the p.d.f. of N_S, for each ρ_S fixed, is

$$\begin{aligned} f_1(n, \rho_S) \quad = \quad & \int_{\mathcal{D}_{RR}} f_{RR,T_2}\left(rr, t_3(1 - rr) - \log\left[\left(\frac{\rho_S}{s_0}\right)^{1/n} + 1 - \mathrm{e}^{-t_1 rr}\right]\right) \\ & \times \left|\frac{\left(\frac{\rho_S}{s_0}\right)^{1/n} \log\left(\frac{\rho_S}{s_0}\right)}{n^2\left(1 - \mathrm{e}^{-t_1 rr} + \left(\frac{\rho_S}{s_0}\right)^{1/n}\right)}\right| \mathrm{d}rr. \end{aligned} \tag{22}$$

5. Probabilistic cost-effectiveness analysis

The cost-effectiveness analysis is useful to perform an economic evaluation of sanitary interventions. Incremental cost-effectiveness ratio, CE, can be used in order to prioritise sanitary interventions and then maximizing benefits taking into account available budgets [17, 18]. CE is a ratio defined from costs and effectivenesses of two alternatives. Namely, CE is defined as

$$CE = \frac{C_2 - C_1}{E_2 - E_1}, \tag{23}$$

where C_i, and E_i, $i = 1, 2$, are the cost and the effectiveness of the alternative i, respectively. Hereinafter, we assume that C_i and E_i, $i = 1, 2$ are r.v.'s. In the following, we will compute the 1-p.d.f. of the CE to compare two treatments for the stroke disease, being the second more effective than the first, but the first cheaper. Then, differences between both are the transition matrix T (particularly the relative risk) and the cost. To obtain the 1-p.d.f. of the CE, first we will determine the expression of the total effectiveness given by the QALY (Quality Adjusted

Life Year), for each treatment. In the context of medical Markov models, the QALY has already been used, see for instance [19]. QALY is the sum of the effectiveness of susceptibles, reliants and deceaseds. In addition, these three effectivenesses are the sum of the effectiveness in each cycle until the value n of total years considered for the study. This effectiveness is the product of number of susceptibles, reliants or deceaseds in each cycle $1 \leq j \leq n$, the utility appropriate for each state and a certain constant, which depends on a discount rate r. These magnitudes will be detailed later. With this aim, we need to know the utility, or the value of life's quality, where 0 value corresponds to death and 1 value represents that stroke disease has not been suffered by individuals, [2, 7, 20]. Then, we will consider $U_S = 1$ and $U_D = 0$ the utilities of susceptibles and deceaseds, respectively. For reliants, we will model the utility, say U_R, through a r.v. Taking into account the extant literature, we consider $r = 0.03$ (3%) as the discount rate [7, 21, 22]. Then, the QALY is given by

$$E_i = \sum_{j=1}^{n} \frac{\mathbb{E}[S_{j,i}]}{(1+r)^{j-1}} + U_R \sum_{j=1}^{n} \frac{\mathbb{E}[R_{j,i}]}{(1+r)^{j-1}}, \quad i = 1, 2, \tag{24}$$

where $\mathbb{E}[S_{j,i}]$ and $\mathbb{E}[R_{j,i}]$ are the average number of susceptibles and reliants of alternative $i = 1, 2$, for each cycle j, $1 \leq j \leq n$, respectively. The second step is to determine the expression of the total cost of each treatment. We will follow the same structure that in the case of the QALY. On the one hand, we will consider that the cost of the each treatment for susceptible subpopulation is $CS_i = a_i W$, where a_i is the cost, in euros, of medicine per kilogram and W is a r.v. that represents the weight of the individual to be studied. On the other hand, we will consider that the dependence cost is a r.v., denoted by CR. This r.v. is assumed to be the same in both treatments. Then, the cost in each treatment is

$$C_i = CS_i \sum_{j=1}^{n} \frac{\mathbb{E}[S_{j,i}]}{(1+r)^{j-1}} + CR \sum_{j=1}^{n} \frac{\mathbb{E}[R_{j,i}]}{(1+r)^{j-1}}, \quad i = 1, 2. \tag{25}$$

Substituting expressions (24)–(25) into (23), one gets

$$CE = \frac{W\, d_1 + CR\, d_2}{d_3 + U_R\, d_2},$$

where

$$\begin{aligned} d_1 &= \sum_{j=1}^{n} \frac{a_2 \mathbb{E}\left[S_{j,2}\right] - a_1 \mathbb{E}\left[S_{j,1}\right]}{(1+r)^{j-1}}, \\ d_2 &= \sum_{j=1}^{n} \frac{\mathbb{E}\left[R_{j,2}\right] - \mathbb{E}\left[R_{j,1}\right]}{(1+r)^{j-1}}, \\ d_3 &= \sum_{j=1}^{n} \frac{\mathbb{E}\left[S_{j,2}\right] - \mathbb{E}\left[S_{j,1}\right]}{(1+r)^{j-1}}. \end{aligned}$$

Now, applying RVT technique, i.e. Theorem 1, we obtain the 1-p.d.f. of CE from the p.d.f. of random vector (W, CR, U_R), which is assumed to be known

$$f_1(ce, n) = \int_{\mathcal{D}_{CR,U_R}} f_{W,CR,U_R}\left(\frac{ce(d_3 + u_r\, d_2) - cr\, d_2}{d_1}, cr, u_r\right) \left|\frac{d_3 + u_r\, d_2}{d_1}\right| \mathrm{d}u_r \mathrm{d}cr. \tag{26}$$

6. Simulating the stroke disease using real data

In this section, we will show the results (simulations) for the Markov model (1)–(2) in order to study the stroke disease. These simulations are built using the results established in Sections 4 and 5 and considering the medical information from [7].

As it is plausible from a practical standpoint, hereinafter we will assume that, at the beginning, the whole population is susceptible, then, the initial condition is $(s_0, r_0, 0)^\top = (1, 0, 0)^\top$. Based upon [7], the following probability distributions for model inputs parameters are considered:

- The relative risk, RR, is a lognormal r.v. with parameters $(1.793; 0.143)$, i.e., $\log(RR) \sim \mathrm{N}(1.793; 0.143)$.
- The transition $R \to D$ is modelled by r.v. P, which is assumed to be a beta distribution with parameters $(80; 120)$, i.e., $P \sim \mathrm{Be}(80; 120)$.
- The deceased rate due to any cause, T_2, is assumed to be a r.v. with a uniform distribution on the interval $]0.02127, 0.02227[$, $T_2 \sim \mathrm{U}(]0.02127, 0.02227[)$.

With regard to the non-mortal stroke rate, t_1, and the stroke deceased rate, t_3, it is assumed that $t_1 = 0.00111$ and $t_3 = 0.00176$, respectively. These values have been taken from reference [7], taking into account that these rates correspond to a group of individuals with 65 years old. Notice that the previous theoretical results can be applied because the r.v.'s RR and T_2, with the distributions specified above, satisfy condition (3).

In Figure 2, the 1-p.d.f.'s of susceptible, reliant and deceased subpopulations, given by expressions (13)–(15), have been plotted. These graphical representations have been made in periods $\{1, 2, \ldots, 25\}$, assuming that r.v.'s RR, P and T_2 are independent. From Figure 2, we observe that the percentage of susceptibles decreases as time increases. Besides, the percentage of reliant increases at the beginning, specifically from $n = 1$ to $n = 6$, and afterwards this percentage tends to zero.

On the other hand, the deceased subpopulation is an absorbent state, therefore in the long-term all the population will reach this state. This behaviour is in agreement with the results shown in Figure 2. From this graphical representation it can be observed that both the percentage of dead and its variability increase over time. This same behaviour is observed to susceptible subpopulation for the periods plotted in Figure 2, although it will decrease as time goes on. Finally, the shape of the 1-p.d.f., $f_1(r, n)$, depicted in Figure 2 becomes sharp as standard deviation decreases.

In Figure 3, the mean plus/minus standard deviation functions of the three subpopulations are shown. Notice that graphical representations exhibited in Figure 2 and Figure 3 are in agreement.

We point out that the computation of the 1-p.d.f. is very useful in applications since from it, as we have seen previously in Subsection 4.1, one can compute exact confidence intervals in order to construct probabilistic predictions. In addition, it permits the computation of the probability associated to sets of interest. For instance, from expression (19) applied to the reliant subpopulation, we can obtain the probability that the proportion of reliants that lies between $a = 0.010$ (1%) and $b = 0.015$ (1.5%) in the time period $\hat{n} = 5$ is, approximately 0.7,

$$\mathbb{P}[0.010 \leq R_5 \leq 0.015] = \int_{0.010}^{0.015} f_1(r, 5)\, \mathrm{d}r = 0.7006.$$

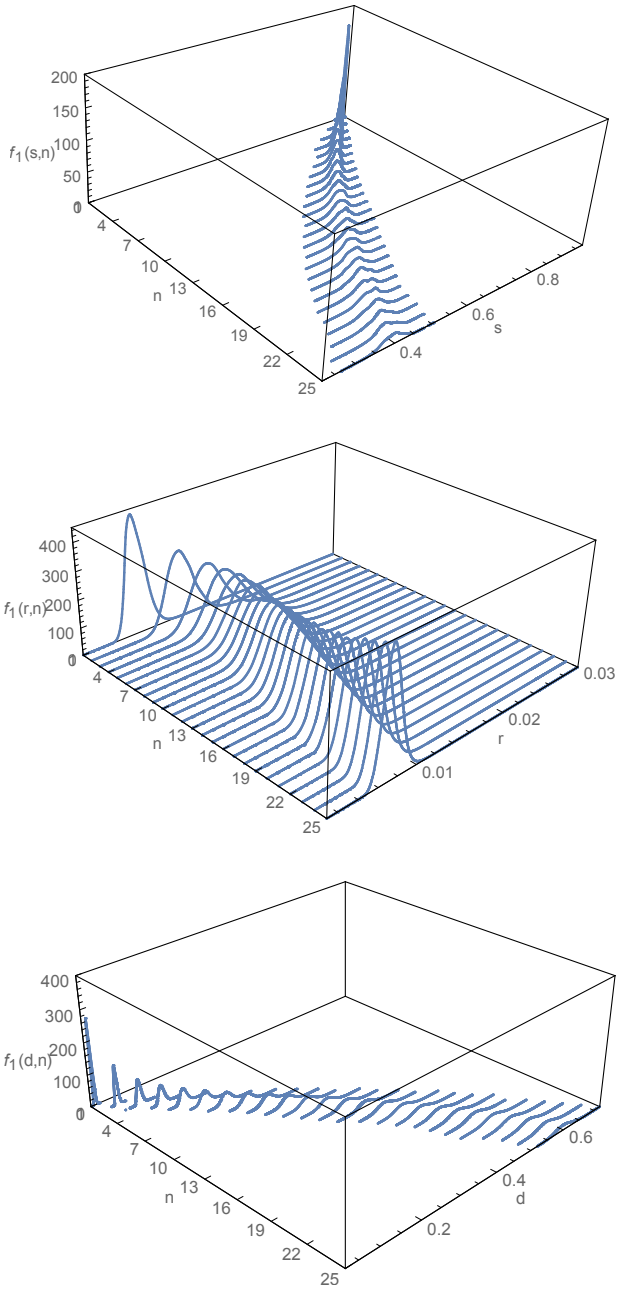

Figure 2: Plot of the 1-p.d.f.'s: $f_1(s,n)$ given by (13) (top); $f_1(r,n)$ given by (14) (middle); $f_1(d,n)$ given by (15) (bottom) at the values $n \in \{1,2,\ldots,25\}$.

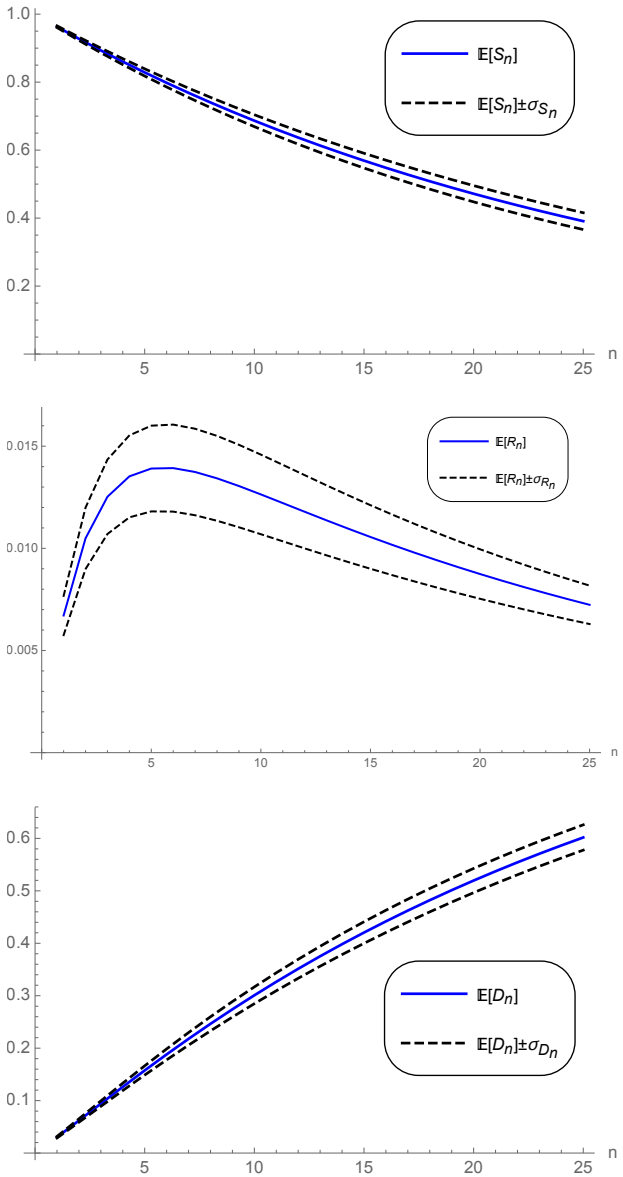

Figure 3: Plot of the expectation plus/minus standard deviation functions of susceptible (top), reliant (middle) and deceased subpopulations (bottom).

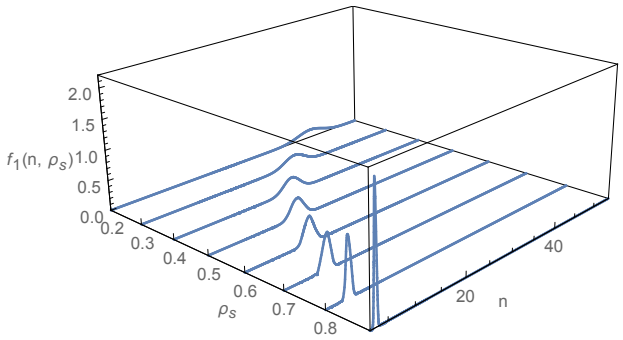

Figure 4: Plot of the p.d.f. of the time N_S until a proportion $\rho_S \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ of the population remains susceptible.

Table 1: Expectation of time N_S until a proportion, ρ_S, of the population remains susceptible for different values ρ_S.

ρ_S	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\mathbb{E}[N_S]$	42.9520	32.1312	24.4538	18.4980	13.6327	9.5189	5.9553	2.8118

Now, we will determine the p.d.f.'s of time until a given proportion of the population remains susceptible, reliant or deceased. For susceptible subpopulation this has been done using expression (22). Figure 4 shows this p.d.f. for the following values of $\rho_S \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

From the p.d.f. of N_S, we can compute the expectation of r.v. N_S for a fixed value of ρ_S, like 0.70,

$$\mathbb{E}[N_S] = \int_0^\infty n f_1(n, 0.70)\, \mathrm{d}n = 9.5190.$$

This means that, approximately, the middle of 10-th cycle (since the study starts at $n = 0$) represents the average time until 70% of the population will be susceptible. This can also be seen graphically in Figure 4. Table 1 collects the expectation, $\mathbb{E}[N_S]$, for different values of ρ_S. This is a key information for doctors when they want to study the evolution of susceptibles of stroke disease in a group of patients.

In order to obtain the p.d.f.'s of r.v.'s N_R and N_D, that denote the time until a proportion of population, ρ_R and ρ_D, remains reliant or deceased, respectively, we have applied numerical methods. We have made this decision because n cannot be isolated from the second and third components of the solution given by (8). To illustrate the process that we have followed to carry out computations, below we specify the steps for reliant subpopulation where ρ_R is assumed to be fixed:

- Step 1: To sample 500 000 values, say (rr, t_2, p), according to the specific distributions assumed for r.v.'s RR, T_2 and P.
- Step 2: For each sampled value (rr, t_2, p), to apply Newton method to calculate the value n of N_R solving the nonlinear equation defined by the second component of (8), that corresponds to the reliant subpopulation, substituting $RR \Rightarrow rr$, $T_2 \Rightarrow t_2$ and $P \Rightarrow p$. This process provides 500 000 values for n of N_R.
- Step 3: To plot the histogram of 500 000 values of n. A normalization of this histogram is an approximation of the p.d.f. of N_R.

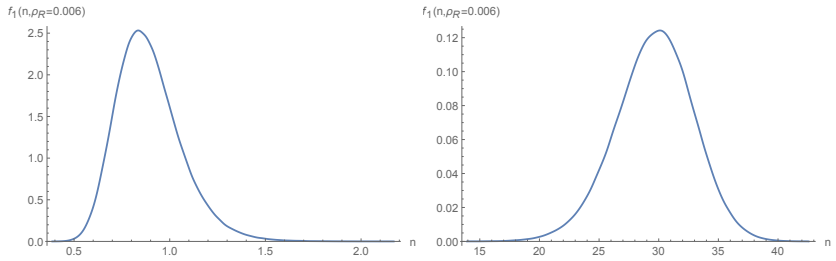

Figure 5: Plot of the p.d.f. of the time N_R until a proportion $\rho_R = 0.006$ of the population remains reliant, using as seed points $sp = 1$ (left) and $sp = 25$ (right).

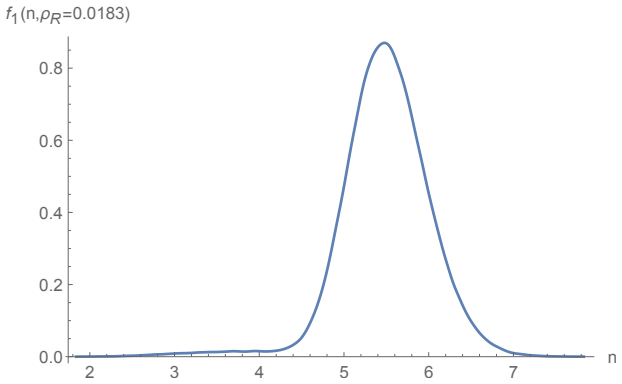

Figure 6: Plot of the p.d.f. of the time N_R until a proportion $\rho_R = 0.0183$ of the population remains reliant, using as seed point $sp = 2$.

Since the numerical convergence of Newton method heavily depends on the seed or starting point, say sp, for example, in the case that $\rho_R = 0.006$, we have obtained two graphical representations for the p.d.f. for N_R, that are shown in Figure 5. Specifically, taking the values $sp = 1$ and $sp = 25$, the p.d.f. of the time N_R has been obtained for cycles $n = 1$ and $n = 30$, respectively. This is due because the proportion of reliants reaches the value $\rho_R = 0.006$ in those two cycles. Whereas for the case $\rho_R = 0.0183$, the Newton method always converges for the cycle $n = 5$, thus defining a single p.d.f. for N_R. This p.d.f. has been plotted in Figure 6 taking as seed value $sp = 2$.

Regarding the deceased subpopulation, we have followed the same steps described previously. In Figure 7, we have plotted the p.d.f. of N_D for the following values of percentage $\rho_D \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, where $sp = 3$ has been taken as the seed point, for each value of ρ_D. In this case, to every value of ρ_D corresponds a unique value of n.

Finally, we will compute the 1-p.d.f. of CE given by (26). From it, the mean and confidence intervals will also be computed. This will be done for different ages, 30, 65 and 71. This decision has been made because the involved rates for each age are different. Computations have been carried out taking the following probability distributions for random inputs:

- The relative risk for the first alternative (the less efficient), RR_1, is a lognormal r.v. with parameters $(1.793; 0.143)$, i.e., $\log(RR_1) \sim \mathrm{N}(1.793; 0.143)$.

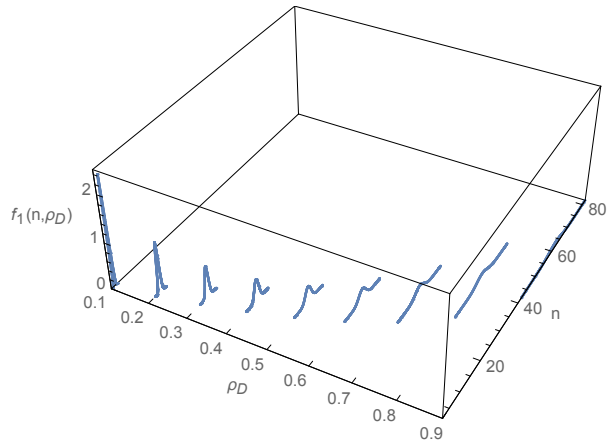

Figure 7: Plot of the p.d.f. of the time N_D until a proportion $\rho_D \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ of the population remains dead.

- The relative risk for the second alternative (the most efficient), $RR_2 = RR_1\, B$, where B is the benefit.
- B is a lognormal r.v. with parameters $(-0.964; 0.163)$, i.e., $\log(B) \sim \mathrm{N}(-0.964; 0.163)$.
- The transition $R \to D$ is modelled by r.v. P, which is assumed to be a beta distribution with parameters $(80; 120)$, i.e., $P \sim \mathrm{Be}(80; 120)$.
- The deceased rate due to any cause, T_2, is assumed to be a r.v. with a uniform distribution on the interval $]t_2 - 0.0001, t_2 + 0.0001[$, i.e., $T_2 \sim \mathrm{U}(]t_2 - 0.0001, t_2 + 0.0001[)$, where t_2 is a fixed value, which depends on age, and that will be specified below.
- The dependence cost, CR, is a lognormal r.v. with parameters $(6.936; 0.643)$, i.e., $\log(CR) \sim \mathrm{N}(6.936; 0.643)$.
- The weight, W, is a normal r.v. with parameters $(75.900; 12.290)$, i.e., $W \sim \mathrm{N}(75.900; 12.290)$. The prices of both treatments are $a_1 = 6.5$€/kg and $a_2 = 65$€/kg, [7].
- The utility, U_R, is a normal r.v. with parameters $(0.701; 0.347)$ i.e., $U_R \sim \mathrm{N}(0.701; 0.347)$.

Above, t_1, t_2 and t_3 are rates, which depend on age of the population under study. In Figure 8, we show the expectation of susceptibles, reliants and and deceaseds for the two alternatives in the three ages. For each one of them, we have considered the following rates, t_1, t_2 and t_3, which are based on [7], and end cycles values, $n_{\max}$,

- For age of 30 years: $t_1 = 0.0000298$, $t_2 = 0.00169$, $t_3 = 0.00004$ and $n_{\max} = 69$.
- For age of 65 years: $t_1 = 0.0011135$, $t_2 = 0.02177$, $t_3 = 0.00176$ and $n_{\max} = 34$.
- For age of 71 years: $t_1 = 0.0031780$, $t_2 = 0.03616$ and $t_3 = 0.00373$ and $n_{\max} = 28$.

Notice that this study is until 99 years, but we could choose another age limit. From Figure 8, we can observe, in all the ages, that the mean of susceptibles with the second treatment is greater

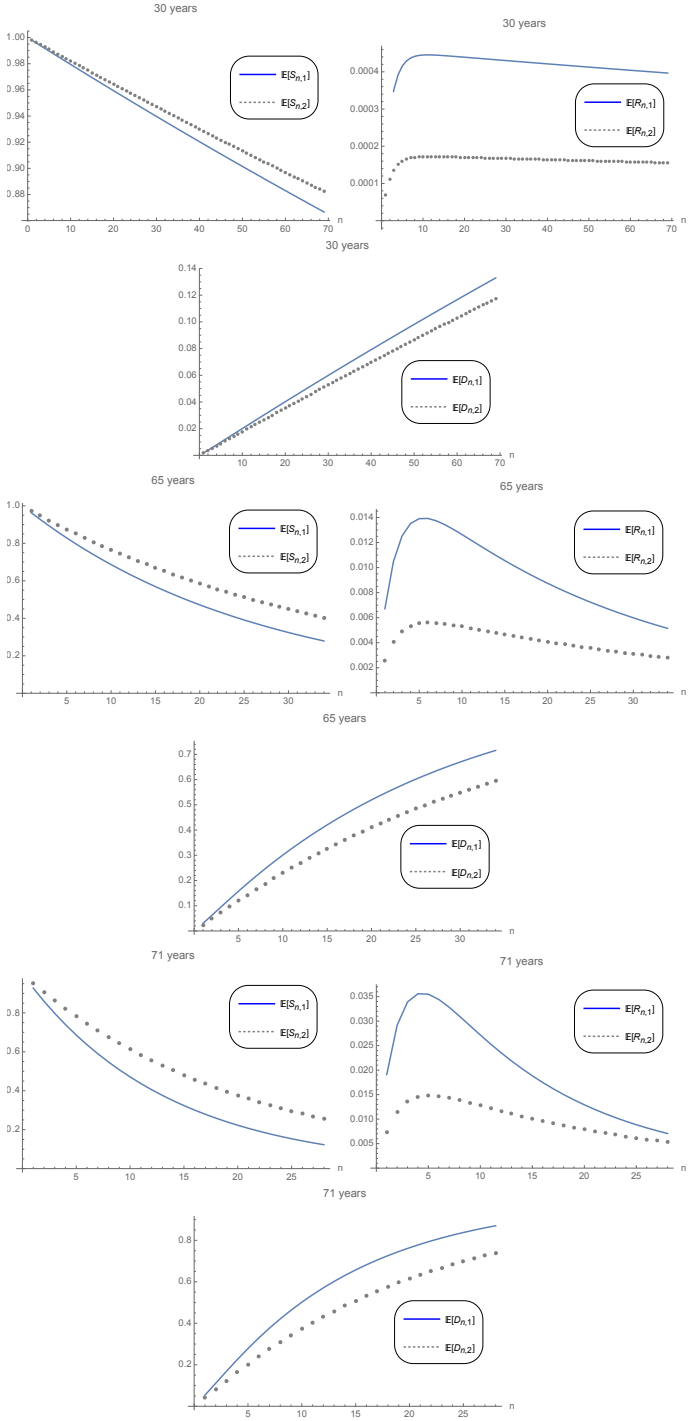

Figure 8: Plots of expectation of susceptibles, reliants and deceaseds for the two alternatives 1 (the cheapest) and 2 (the most expensive) in the three ages: 30 years (first row), 65 years (second row) and 71 years (third row).

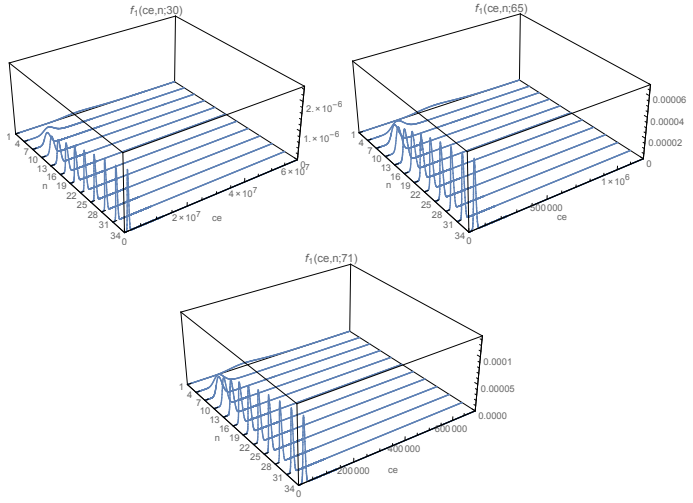

Figure 9: P.d.f.'s de CE given by (26) considering both alternatives in the three ages: 30 years (left), 65 years (center) and 71 years (right).

than considering the first, and the reverse for the expectation of reliants and deceaseds. This is consistent with the fact that the second alternative is better than the first.

In Figure 9 we have plotted the 1-p.d.f.'s of the CE, given by (26), for each age from cycle 1 to 34. Notice that graphical representations shown in Figure 9 are in agreement with Figure 10, where expectation plus/minus standard deviation functions of CE for each age have been plotted. To facilitate comparison between both alternatives, the value 30 000€/QALY (red straight line) has also been plotted as a threshold. This benchmark value has been chosen because, according to [23], is a standard value in the literature. From Figure 10, we can observe that for people aged 71 years old the second alternative (the most expensive), is more effective than the first alternative. Naturally, for people aged 30 years old, the best alternative is the first one because they have longer lifetime. For people aged 65 years old, it might be controversy because their CE is very close to the threshold.

7. Discussion

Although Markov models have been used extensively for modelling the dynamics of numerous diseases, to the best of our knowledge, few attempts have been made regarding the stroke disease. The markovian approach is useful to perform the clinical control of patients that suffer this disease. Indeed, Markov models allow us to forecast not only the number of patients belonging to each subpopulation (susceptibles, reliants and deceaseds) at every cycle but also to account for significant medical information. In this regard, the time until a given proportion of patients remain susceptible, reliant or deceased are, for example, important information in the medical treatment of the stroke. This is a key information to answer crucial questions like "what is the expected time before twenty percent, for instance, of the population remains susceptible?".

In this paper, we have given a technical generalization of classical Markov methodology that enables the exact determination of the crucial medical information previously indicated. This generalization is aimed to improve the modelling of stroke disease when dealing with real data,

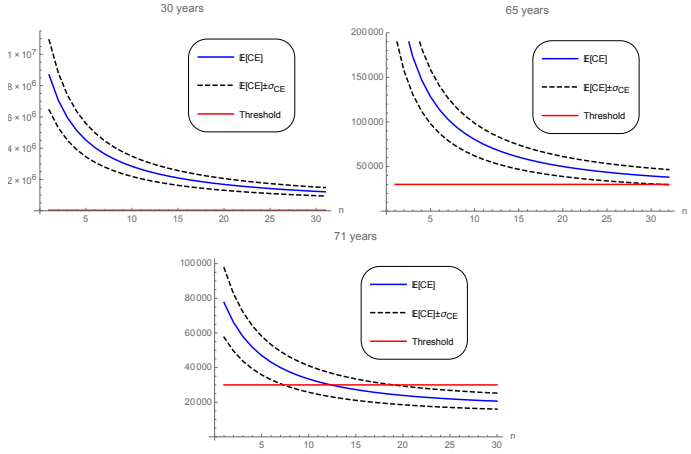

Figure 10: Plots of expectation plus/minus standard deviation functions of CE for each age considering both alternatives in the three ages: 30 years (left), 65 years (center) and 71 years (right). The red straight line represents the threshold value 30 000€/QALY usually taking as reference [23].

although an important issue is that this technique can easily be adapted to another diseases using the markovian paradigm. Our approach resorts in the so-called Random Variable Transformation method to randomize classical Markov chains. This randomization has been done through some of the entries of the transition matrix of a classical Markov chain which has been previously proposed to model the stroke disease. Our approach allows us the computation of the first probability density function of the solution stochastic process, and then obtaining punctual and probabilistic predictions as well as the important probabilistic information that we have underlined previously.

Moreover, we have conducted a probabilistic cost-effectiveness analysis, based on the application of the Random Variable Transformation technique, that to the best of our knowledge, has not been done yet. The main advantage of this computational approach is that results can be obtained in an exact manner rather than using simulations.

Acknowledgements

This work has been partially supported by the Ministerio de Economía y Competitividad grant MTM2013-41765-P. Ana Navarro Quiles acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València. Authors would like to thank Prof. Dr. Javier Mar for providing us medical data about stroke disease from his research.

Conflict of Interest Statement

The authors declare that there is no conflict of interests regarding the publication of this article.

[1] A. Briggs, M. Sculpher, An introduction to Markov modelling for economic evaluation, PharmacoEconomics 13 (4) (1998) 397–409. doi:10.2165/00019053-199813040-00003.

[2] F. Sonnenberg, R. Beck, Markov models in medical decision making: a practical guide, Medical Decision Making 13 (4) (1993) 322–338. doi:10.1177/0272989X9301300409.

[3] B. Cooper, M. Lipsitch, The analysis of hospital infection data using hidden Markov models, Biostatistics 5 (2004) 223–237.
[4] M. Mandel, R. A. Betensky, Estimating time-to-event from longitudinal ordinal data using random-effects Markov models: application to multiple sclerosis progression, Biostatistics 9 (2008) 750–764.
[5] J. Mar, M. Sainz-Ezkerra, E. Miranda-Serrano, Calculation of prevalence with Markov models: Budget impact analysis of thrombolysis for stroke, Medical Decision Making 28 (4) (2008) 481–490.
[6] I. A. Korostil, G. W. Peters, J. Cornebis, R. G. Regan, Adaptive Markov chain Monte Carlo forward projection for statistical analysis in epidemic modelling of human papillomavirus, Statistics in Medicine 32 (11) (2012) 1917–1953.
[7] J. Mar, F. Antoñanzas, R. Pradas, A. Arrospide, Los modelos de Markov probabilísticos en la evaluación económica de tecnologías sanitarias: una guía práctica, Gaceta Sanitaria 24 (3) (2010) 209–214. doi:10.1016/j.gaceta.2010.02.006.
[8] S. L. Pan, H. M. Wu, A. M. F. Yen, T. H. H. Chen, A Markov regression random-effects model for remission of functional disability in patients following a first stroke: A Bayesian approach, Statistics in Medicine 26 (29) (2007) 5335–5353.
[9] V. Kapetanakis, F. E. Matthews, A. van den Hout, A semi-Markov model for stroke with piecewise-constant hazards in the presence of left, right and interval censoring, Statistics in Medicine 32 (4) (2013) 697–713.
[10] J. Zhang, K. F. Yu, What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes, Journal of American Medical Association (JAMA) 290 (19) (1998) 1690–1691.
[11] L. A. McNutt, C. Wu, X. Xue, J. P. Hafner, Estimating the relative risk in cohort studies and clinical trials of common outcomes, American Journal of Epidemiology 157 (10) (2003) 940–943. doi:10.1093/aje/kwg074.
[12] T. T. Soong, Random Differential Equations in Science and Engineering, Academic Press, New York, 1973.
[13] B. Kegan, R. Webster West, Modeling the simple epidemic with deterministic differential equations and random initial conditions, Mathematical Biosciences 195 (5) (2005) 179–193. doi:10.1016/j.mbs.2005.02.004.
[14] M. C. Casabán, J. C. Cortés, J. V. Romero, M. D. Roselló, Probabilistic solution of random SI-type epidemiological models using the Random Variable Transformation technique, Communications in Nonlinear Science and Numerical Simulation 24 (1–3) (2015) 86–97. doi:10.1016/j.cnsns.2014.12.016.
[15] M. C. Casabán, J. C. Cortés, A. Navarro-Quiles, J. V. Romero, M. D. Roselló, R. J. Villanueva, A comprehensive probabilistic solution of random SIS-type epidemiological models using the Random Variable Transformation technique, Communications in Nonlinear Science and Numerical Simulation 32 (2016) 199–210. doi:10.1016/j.cnsns.2014.12.016.
[16] F. A. Dorini, M. S. Cecconello, L. B. Dorini, On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density, Communications in Nonlinear Science and Numerical Simulation 33 (2016) 160–173. doi:10.1016/j.cnsns.2014.12.016.
[17] L. Prieto, J. A. Sacristán, F. Antoñanzas, C. Rubio-Terrés, J. L. Pinto, J. Rovira, Análisis coste-efectividad en la evaluación económica de intervenciones sanitarias, Med. Clin. 122 (13) (2004) 505–510.
[18] G. Karlsson, M. Johannesson, The decision rules of cost-effectiveness analysis, PharmacoEconomics 9 (2) (1996) 113–120.
[19] G. Hazen, Z. Li, Cohort decomposition for Markov cost-effectiveness models, Medical Decision Making 31 (1) (2011) 19–34.
[20] J. L. Pinto-Prades, J. Puig-Junoy, V. Ortún-Rubio, Análisis coste-utilidad, Atención Primaria 27 (2001) 569–573. doi:10.1016/S0212-6567(01)78861-0.
[21] X. Badia, H. Bueno, J. R. González-Juanatey, V. Valentín, M. Rubio, Análisis de la relación coste-efectividad a corto y largo plazo de clopidogrel añadido a terapia estándar en pacientes con síndrome coronario agudo en España, Rev. Esp. Cardiol. 58 (2005) 1385–1395. doi:1016/S0300-8932(05)74068-9.
[22] F. Antoñanzas, F. Brenes, J. M. Molero, A. Fernández-Pro, A. Huerta, R. Palencia, J. Cozar, Cost-effectiveness of the combination therapy of dutasteride and tamsulosin in the treatment of benign prostatic hyperlasia in Spain, Actas Urol. Esp. 35 (2) (2011) 65–71. doi:10.1016/j.acuro.2010.11.008.
[23] J. A. Sacristán, J. Oliva, J. Del Llano, L. Prieto, J. L. Pinto, ¿Qué es una tecnología sanitaria eficiente en España?, Gaceta Sanitaria 4 (16) (2002) 334–343.