AL-Azhar University
(Girls' Branch)
Faculty of commerce
Department of Statistics

A STUDY OF AN ILLNESS-DEATH MODEL THROUGH MULTI -STATE APPROACH

By

GEHAD MOHAMED ISMAIL

Department of Statistics – AL-Azhar University

Supervised by

Prof. Amina Esmail Abo-Hussien
Professor of Statistics
Faculty of Commerce
AL-Azhar University
Girl's Branch-Cairo

Dr. Hala Fawzy Mohammad
Lecturer of Statistics
Faculty of Commerce
AL-Azhar University
Girl's Branch-Cairo

1444 A.H - 2023 A.D

Explicit Mathematical Expressions for the Hazard Rates of the Illness-Death Model: The Case of Egypt 2019

Gehad M. Ismail [1], Amina E. Abu-Hussien [2] , Hala F. Mohamed [3]

[1] Master's Researcher at the Faculty of Commerce, Al-Azhar University (Girls' Branch), Cairo, Egypt.

[2] Professor of Statistics, Faculty of Commerce, Al-Azhar University (Girls' Branch), Cairo, Egypt.

[3] Lecturer of Statistics, Faculty of Commerce, Al-Azhar University (Girls' Branch), Cairo, Egypt.

Abstract

The illness-death model with three states (the health state, the illness state, and the death state) is a special case of multistate models. This model is important in several applications such as medical applications and research to analyze disease evolution or mortality. The transition probabilities are derived as functions of the hazard rates. This paper introduces a trial to obtain explicit mathematical expressions for the hazard rates of the illness-death model using data of Egypt in 2019. The hazard rates are derived using the Pearson distributions system.

Keywords

Illness-death model, multi- state models, Pearson system, transition probabilities.

1. Introduction

Event history analysis deals with data obtained by observing individuals over time, focusing on events occurring for the individuals. Thus, typical outcome data consist of times of occurrence of events and the types of events that occur. Frequently, an event may be considered a transition from one state to another therefore, multi-state models will often provide a relevant modeling framework for event history data.

According to Hougaard (1999), Multi-state models are models for a process describing a life history of an individual, which at any time occupies one of a few possible states. Commenges (1999) discussed some assumptions for multi-state models in epidemiology. He introduced assumptions for one process: non-homogenous Markov model, homogenous Markov model and semi-Markov model. Commenges *et al* (2004) presented the illness-death model for studying the incidence and the prevalence of Alzheimer's disease or dementia. Meira-Machado *et al* (2009) introduced multi state models to model the movement of patients between the various states. They reviewed several modeling approaches following the multi-state models. Hinchliffe *et al* (2013) presented the unidirectional Illness-death model with four states (alive, illness, dead without illness, dead with illness). They showed the flexible

parametric survival model for transition hazard rates. Touraine *et al* (2016) examined the illness-death model from clinical and epidemiological points of view. They focused on some quantities such as transition probabilities, cumulative probabilities, and life expectancies. Touraine *et al* (2017) presented the irreversible illness-death model with three states (initial, absorbing, and intermediate). Brinks (2018) studied the relationship between prevalence and incidence of chronic disease with the illness-death model. He described the change rate of the number of non-diseased individual and number of diseased individual by using the system of ordinary differential equation. Lauseker and Eullenburg (2019) presented the analysis of cause of death by two approaches, the first approach is progressive illness-death model with three transitions and four different states. The second approach is competing risks when ignoring the transient states. They calculated the state occupation probability for chronic phase and estimated these probabilities. Brinks and Hoyer (2020) presented simulation illness-death model with two approaches for test two type diabetes in Germany, the first method is the discrete event simulation, where relevant events, i.e., onset of disease and death, are simulated for each subject individually, the second method is the Doob-Gillespie algorithm, where for each (small) time step, the number of transitions between the states are drawn from a Poisson distribution , both algorithms are

similar in bias and variance. Chae *et al* (2021) presented the illness-death model approach to evaluate the relationship between relapse and death in patients with non-small cell lung cancer to compare with the disease-free survival model, total of 917 patients who underwent curative surgery for primary lung cancer from 2010 to 2018 were initially included in this study. Zhang and Xu (2022) presented the marginal structural models based on simple illness-death model for semi-competing data using the potential outcome framework. They presented two specific models, the usual Markova illness-death using estimating equations with inverse probability weighting and the general Markova illness-death requires a weighted EM algorithm. Using **semicmprskcoxmsm** package in **R** programming to analysis of mid-life alcohol exposure on late life cognitive impairment as well as mortality using the Honolulu-Asia Aging Study data set.

This paper is organized as follows: the illness-death mode, its hazard rates, transition probabilities and the expected duration of stay in each state are introduced in Section 2. Derivation of explicit mathematical expression of the hazard rates using Pearson approach is illustrated in Section 3. Some numerical results of the model are given in Section 4.

2. The illness –death model

Illness –death model is widely used in the medical literature to evaluate whether diseased individuals have the same risk of death as healthy individuals [Meira-Machado *et al*. (2009)]. Illness-death processes can be applied to other types of problems; for example, we may study occupations or geographic locations or, we may study any repairable disorder of a mechanical object [Chiang (1968)].

Let we have a population with two states: S_1 (health state), S_2 (illness state) and one death state S_3. Each individual in the population may travel between the states of the model in the time interval (0, t), it is assumed that the illness is so dangerous that recovery from it, is not allowed i.e., no transition from S_2 to S_1, an individual has arrived at a death state, he will remain in the state forever. No transition from S_3 to S_1 or S_2.

Transitions among the states of the model are caused by the hazard rates of morbidity v_{ij} and hazard rates of mortality μ_{ik} as follows:

$$\lim_{\Delta\to 0} \frac{1-P_{ii}(\tau,\tau+\Delta)}{\Delta} = -v_{ii}(\tau) \qquad i = 1,2 \qquad (1)$$

$$\lim_{\Delta\to 0} \frac{P_{ij}(\tau,\tau+\Delta)}{\Delta} = v_{ij}(\tau) \qquad i,j = 1,2\ ,\ i < j$$

(2)

$$\lim_{\Delta\to 0} \frac{Q_{ik}(\tau,\tau+\Delta)}{\Delta} = \mu_{ik}(\tau) \qquad i = 1,2\ ,\ k = 3$$

(3)

States of the model with the possible transitions [Commenges (1999)]

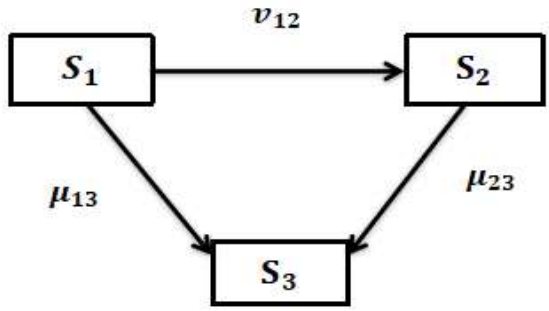

Figure 1 Simple illness-death model

During the time interval (0, t) individuals may be transient from health state to illness state through the transition probabilities (T.Ps) of illness, as follows:

P_{ij}

$= pr\{\text{an individuals in state } S_i \text{ at time 0 will be in state } S_j \text{ at time t}\}$

$$i, j = 1,2 \quad i \leq j \qquad (4)$$

If individuals transient from health or illness states to death state,S_k, the probabilities will be called T.Ps of death, as follows:

$$Q_{ik} = pr\{\text{an individual in state } S_i \text{at time 0 will be in the state } S_k \text{ at time t}\}$$

$$i=1,2 \text{ and } k=3 \qquad (5)$$

The Markov property implies important relations among the transition probabilities $P_{ij}(\tau, t)$. Let ξ be a fixed point in the interval (τ, t) so that $\tau < \xi < t$ and let $X(\tau)$, $X(\xi)$, $X(t)$ be the random variables, then :

$$Pr\{X(t) = K|X(\tau) = i, X(\xi) = j\} = Pr\{X(t) = K|X(\xi) = j\} = P_{jk}(\xi, t) \qquad (6)$$

$$Pr\{X(\xi) = j \; and \; X(t) = k|X(\tau) = i\} = P_{ij}(\tau, \xi)P_{jk}(\xi, t) \;\; i, j = 1,2, \; k = 3 \, , \; i \leq j \qquad (7)$$

Equation (7) is known as the Chapman-Kolmogorov equation.

Solution for Kolmogorov differential equations by using matrices to get the formulas of the T.Ps [chiang (1968)], is as follows:

$$P_{ij}(t) = \sum_{l=1}^{2} \frac{A_{ij}(l)}{\prod_{r=1}^{2}(\rho_l - \rho_r)} e^{\rho_l t} \qquad i,j = 1,2 \ , i \leq j$$

(8)

Where the probabilities of death are:

$$Q_{ik}(t) = \int_0^t \sum_{j=1}^{2} P_{ij}(\tau)\mu_{jk}\, dt$$

(9)

Then the transition probabilities are:

$$P_{11}(t) = e^{-[v_{12}+\mu_{13}]t}$$

(10)

$$P_{12}(t) = \frac{v_{12}[e^{-\mu_{23}t} - e^{-[v_{12}+\mu_{13}]\,t}]}{(v_{12}+\mu_{13}-\mu_{23})}$$

(11)

$$P_{22}(t) = e^{-\mu_{23}t}$$

(12)

$$Q_{13}(t) = \frac{[1-e^{-[v_{12}+\mu_{13}]t}]}{[v_{12}+\mu_{13}]}\mu_{13}$$

(13)

$$Q_{23}(t) = [1 - e^{-\mu_{23}t}]$$

(14)

The expected duration of stay in each state of the model.

Is the average duration that an individual is expected to stay in each of the statesS_1 and S_2 , within a time period of length *t*. This concept is similar to the expectation of life in life tables [Chiang (1968)].

The expected duration of stay in S_1 in the interval (*0*, *t*) for an individual whose initial state is S_1:

$$e_{11}(t) = \frac{1}{(v_{12}+\mu_{13})}\left[1 - e^{-(v_{12}+\mu_{13})t}\right]$$

(15)

The expected duration of stay in S_2 in (*0*, *t*) for an individual whose initial state is S_2: $e_{22}(t) = \frac{(1-e^{-\mu_{23}t})}{\mu_{23}}$

(16)

3. **Derivation of explicit mathematical expressions for the hazard rates in the model: the case of Egypt2019**

There is no general trend for the behavior of most of the hazard rates included in the illness-death model. They have different shapes not only among different countries, but also in the same

country over time. An approximation to these rates using the corresponding observed age-specific failure rates is introduced using published data for Egypt in 2019. The age-specific failure rates are used in drawing the observed curves from which the theoretical functions may be obtained. Pearson distributions can be used to fit the distributions for any data using the first four moments of the data, thus Pearson distributions made statistical analysis possible for any data with unknown distributions. In deriving the explicit mathematical expressions, Pearson method is used in the present study.

The Pearson method

Elderton and Johnson (1969) designed a system of curves of continuous probability distributions that describes observed data in mathematical terms for all the possible variations of the first four moments. They realized that all continuous probability distributions can be classified into a small number of family types according to the value of a criterion denoted by κ.

The coefficient κ used to determine the types of distributions, is given by

$$\kappa = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} \quad , \quad -\infty < \kappa < \infty$$

(17)

Where $\beta_1$and β_2 denote the skewness and kurtosis respectively. Several types of Pearson distribution curves depend on the values of the coefficient κ.

Using the data from the Annual Bulletin of Birth and Death Statistics 2019 and Annual Bulletin of Health Services Statistics and Treatment at State's Expense at Home and Abroad 2019 from Central Agency for Public Mobilization & Statistics in Egypt, the age intervals are adjusted, as follows:

- The age intervals of the mid-year population estimates have been modified to be: 1-, 5-, 15-, 45-, and 65-70 only.
- The age intervals of the number of deaths from infectious diseases are modified to be: 1-, 5-, 15-, 45-, and 65-70 only.
- The age intervals of the number of cases of infectious diseases are modified to be: 1-, 5-, 15-, 45-, and 65-70 only.

To derive the explicit mathematical functions for the hazard rates, the corresponding age-specific failure rates are used in drawing the observed curves from which the theoretical expressions are derived. The age-specific morbidity rate,$v_{12}(x)$, in the k-th age group may be defined as:

$$\frac{\text{The number of cases of infectious diseases in the k-th age group}}{\textit{The estimated number of population at the mid-year of the k-th age group}} \quad (18)$$

The age-specific mortality rate from other causes, $\mu_{13}(x)$, in the k-th age group may be defined as:

$$\frac{\text{The number of deaths from other causes in the k-th age group}}{\textit{The estimated number of population at the mid-year of the k-th age group}} \quad (19)$$

The age-specific mortality rate with infectious diseases, $\mu_{23}(x)$, in the k-th age group may be defined as

$$\frac{\text{The number of deaths with infectious diseases in the k-th age group}}{\textit{The estimated number of } \text{cases of infectious diseases at } \textit{the mid-year of the k-th age group}} \quad (20)$$

Derivation of explicit mathematical expressions for the morbidity

hazard rate v_{12}:

In order to specify a mathematical function that describes the morbidity hazard rate, the age-specific morbidity rates are used as follows:

1. The data required for the calculation of the age-specific morbidity rates are:
   - The distribution of population by age.
   - The distribution of cases of infectious diseases by age.

2. Calculate the age-specific morbidity rates using (18) for Egypt in 2019 as shown in Table (1).
3. The graphical presentation of the relation between age and the age-specific morbidity rates, starts with a low level, increasing in the age group (1-5), decreasing in the age group (5-45), then increasing in the age group (45-65), as shown in Figure (2).

   The investigation of the shape and using **Pearson Diagram** function in **R** programming reveals that it may be well described using Pearson type I given by:

$$v_{12}(x) = y_0(1+\frac{x}{a_1})^{m_1}(1-\frac{x}{a_2})^{m_2} \qquad ,-a_1 \le x \le a_2 \quad (21)$$

   Using the **Pearson Diagram** function to study the efficiency of the model is shown in Figure (3).

Table (1) The age-specific morbidity rates by age groups

Age groups	Population estimates at mid-year*	Number of cases of infectious diseases**	v_{12}
1-	10420291	13373	0.0012
5-	20704217	35846	0.0017
15-	37479650	23687	0.0006
45-	15270895	4993	0.0003
65-70	2188642	2152	0.0010

Sources: *Central Agency for Public Mobilization Statistics, Annual Bulletin of Birth and Death Statistics 2019 in Egypt, tables 3 and 17.

** Central Agency for Public Mobilization Statistics, Annual Bulletin of Health Services Statistics and Treatment at State's Expense at Home and Abroad 2019 in Egypt, table 23.

Plot of the age-specific morbidity rates $\boldsymbol{v_{12}}$by age groups is shown in Figure (2)

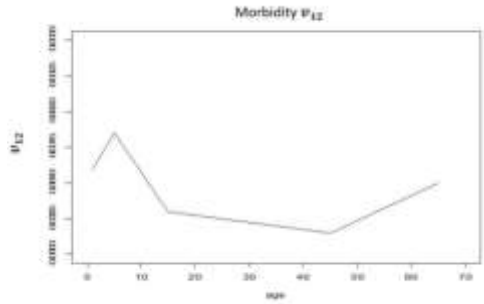

Figure 2 The age-specific morbidity rates $\boldsymbol{v_{12}}$by age groups

Pearson diagram for age-specific morbidity rates:

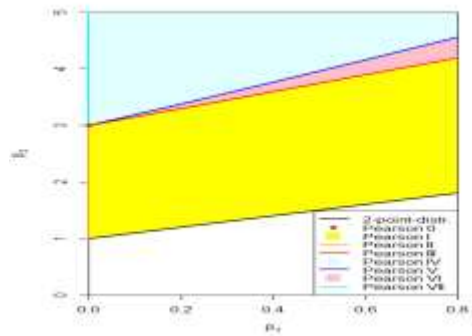

Figure 3 Pearson diagram for age-specific morbidity rates $\boldsymbol{v_{12}}$

The shaded area represent Pearson's Type I distribution. It is noticed that Pearson's Type I distribution covers a wide region in the skewness–kurtosis plane.

Derivation of explicit mathematical expressions for the mortality from other causes hazard rate μ_{13}

In order to specify a mathematical function that describes the mortality hazard rateμ_{13}, the age-specific mortality rates $\mu_{13}$are used as follows:

1. The data required for the calculation of the age-specific mortality rates $\mu_{13}$are:
   - The distribution of population by age.
   - The distribution of deaths from other causes by age.
2. Calculate the age-specific mortality rates μ_{13}using (19) for Egypt in 2019 as shown in Table (2).
3. The graphical presentation of the relation between age and the age-specific mortality rates, starts with a low level, and then the mortality rate is increasing with age increasing, as shown in Figure (4).

   The investigation of the shape and **Pearson Diagram** function in **R** programming reveals that it may be well described using Pearson type I given by:

$$\mu_{13}(x) = y_0(1+\frac{x}{a_1})^{m_1}(1-\frac{x}{a_2})^{m_2} \quad ,-a_1 \le x \le a_2$$
(22)

Using the **Pearson Diagram** function to study the efficiency of the model is shown in Figure (5).

Table (2) The age-specific mortality rates $\boldsymbol{\mu_{13}}$by age groups

Age groups	Number of cases of infectious diseases**	Deaths from other causes*	$\boldsymbol{\mu_{13}}$
1-	13373	3313	0.00032
5-	35846	2261	0.00011
15-	23687	12210	0.00032
45-	4993	51816	0.00339
65-70	2152	25188	0.01150

Sources: *Central Agency for Public Mobilization Statistics, Annual Bulletin of Birth and Death Statistics 2019 in Egypt, tables3 and 17.

** Central Agency for Public Mobilization Statistics, Annual Bulletin of Health Services Statistics and Treatment at State's Expense at Home and Abroad 2019 in Egypt, table23.

Plot of the age-specific mortality rates $\boldsymbol{\mu_{13}}$by age groups is shown in Figure (4)

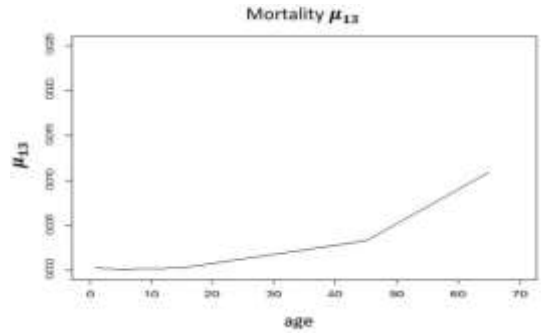

Figure 4 The age-specific mortality rates $\boldsymbol{\mu_{13}}$by age groups

Pearson diagram for age-specific mortality rates μ_{13}:

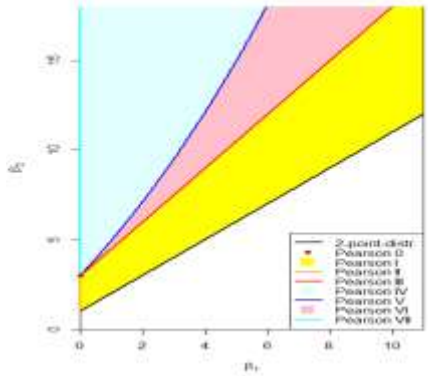

Figure 5 Pearson diagram for age-specific mortality rates $\boldsymbol{\mu_{13}}$

The shaded area represent the Pearson's Type I distribution. It is noticed that Pearson's Type I distribution cover a wide region in the skewness–kurtosis plane. **Derivation of explicit mathematical expressions for the mortality for infectious diseases hazard rate $\boldsymbol{\mu_{23}}$**

In order to specify a mathematical function that describes the mortality hazard rate μ_{23}, the age-specific mortality rates $\mu_{23}$are used as follows:

1. The data required for the calculation of the age-specific motality rates $\mu_{23}$are:
   - The distribution of cases of infectious diseases individuals by age.
   - The distribution of deaths with infectious diseases by age.
2. Calculate the age-specific mortality rates μ_{23}using (20) for Egypt in 2019 as shown in Table (3).
3. The graphical presentation of the relation between age and the age-specific mortality rates, starts with a high level, decreases in the age group (1-5), and then increases as age increases, as shown in Figure (6).

   The investigation of the shape and **Pearson Diagram** function in **R** programming reveals that it may be well described using Pearson type IV given by:

$$\mu_{23}(x) = (1 + \frac{x^2}{a^2})^{-m} e^{-v\, arctan(x/a)} \quad , \qquad -\infty \leq x \leq \infty \qquad (23)$$

   Using the **Pearson Diagram** function to study the efficiency of the model is shown in Figure (7).

Table (3) The age-specific mortality rates $\boldsymbol{\mu_{23}}$ by age groups

Age groups	Number of cases of infectious diseases**	Deaths with infectious diseases*	μ_{23}
1-	13373	697	0.05
5-	35846	377	0.01
15-	23687	1328	0.06
45-	4993	4652	0.93
65-70	2152	2147	0.99

Sources: *Central Agency for Public Mobilization Statistics, Annual Bulletin of Birth and Death Statistics 2019 in Egypt, tables3 and 17.

** Central Agency for Public Mobilization Statistics, Annual Bulletin of Health Services Statistics and Treatment at State's Expense at Home and Abroad 2019 in Egypt, table23 .

Plot of the age-specific mortality rates $\mu_{23}$by age groups is shown in Figure (6)

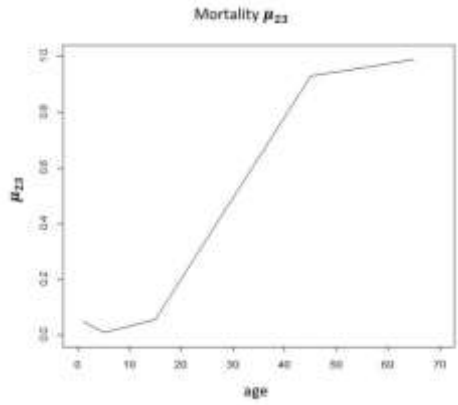

Figure 6 The age-specific mortality rates $\mu_{23}$by age groups

Pearson diagram for age-specific mortality rates $\boldsymbol{\mu_{23}}$:

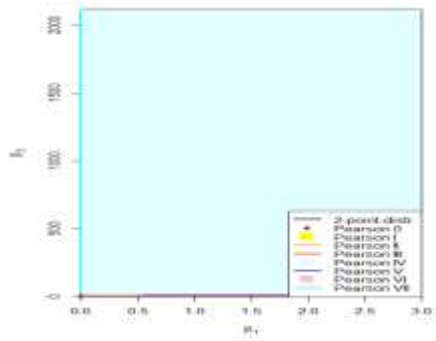

Figure 7 Pearson diagram for age-specific mortality rates $\boldsymbol{\mu_{23}}$

The shaded area represents the Pearson's Type IV distribution. It is noticed that Pearson's Type IV distribution covers a wide region in the skewness–kurtosis plane

4. Some numerical results of the model

This section presented some numerical results for the transition probabilities and expected duration of stay in each state.

The T.Ps:

The T.Ps by age groups are obtained using equations (10), (11), (12), (13) and (14). Table (4) showed the values of the T.Ps by age groups at time t=5 years.

It is noticed that:

- The T.P P_{11} starts with a low level, then increasing with age up to age 15, then decrease at the age interval (15-45), then begins to increase again with increasing age.
- The T.P P_{12} starts with a low level, then increasing with age up to age 15, then decrease at the age interval (15-65), then increase again.
- The T.P P_{22} starts with a high level, then decrease at the age interval (5-15), then begins to increase again with increasing age.
- The T.P Q_{13} starts with a high level, then decrease at the age interval (5-15), then begins to increase again with increasing age, then decrease at the age interval (65-70).
- The T.P Q_{23} starts with a low level, then increasing with age up to age 15, then begins to decrease again with increasing age.

Table 4: Transition probabilities by age groups at time t=5 years

T.ps / Age	P_{11}	P_{12}	P_{22}	Q_{13}	Q_{23}
1-	0.007٦	0.00052	0.22٠	0.208٠	0.778٠
5-	0.0090	0.00809	0.048	0.0602	0.951٠
15-	0.0045	0.00258	0.259	0.3462	0.7408
45-	0.0182	0.00031	0.990	0.9019	0.009٦
65-70	0.0605	0.00095	0.992	0.8642	0.007١

The expected duration of stay in each state of the model

Using equations (15) and (16) the expected duration of stay in each state for age groups in a time interval t=5 years is shown in Table (5) and in a time interval t=10 years is shown in Table (6)

Table 5 : The expected duration of stay in each state by age groups in a time interval t=5 years

Expected duration / Age	e_{11}	e_{22}
1-	4.98	4.42
5-	4.97	4.87
15-	4.98	4.32
45-	4.95	1.06
65-70	4.84	1.00

Table 6 : The expected duration of stay in each state by age groups in a time interval t=10 years

Expected duration / Age	e_{11}	e_{22}
1-	9.92	7.86
5-	9.91	9.51
15-	9.95	7.51
45-	9.80	1.07
65-70	9.40	1.00

It is noticed form tables 5 and 6 that:

- The expected duration of stay in the health state for all age groups take values larger than 4.8 years and 9.4 years when the observation time are 5 years and 10 years respectively, which means that the individuals in the health state are expected to live most of the interval length.
- The expected duration of stay in the illness state for all age groups and for different times of observation take the same type of the expectancy in life tables. It is noticed that it begins with same values at the age groups (1-5) then increase in the age groups (5-15), then begins to decrease as age increases.
- The expected duration of stay in the health state is larger than the expected duration of stay in the illness state for all age groups.

References

Brinks, R. (2018). Illness-Death model in chronic disease epidemiology: characteristics of a related, differential equation and an inverse problem. *Computational and Mathematical Methods in Medicine, 2018*.

Brinks, R., & Hoyer, A. (2020). Simulation of trajectories in the illness-death model for chronic diseases: discrete event simulation, Doob-Gillespie algorithm and coverage of Wald confidence intervals. *medRxiv*, 2020-04.

Chae, K. J., Choi, H., Jeong, W. G., & Kim, J. (2021). The Value of the Illness-Death Model for Predicting Outcomes in Patients with Non–Small Cell Lung Cancer. *Cancer Research and Treatment*, *54*(4), 996-1004.

Chiang, C.L. (1968). *Introduction to stochastic processes in biometrics. John wiley and sons*, Newyork.

Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime data analysis*, *5*, 315-327.

Commenges, D., Joly, P., Letenneur, L., & Dartigues, J. F. (2004). Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in medicine*, *23*(2), 199-210.

Elderton, W. and Johnson, N. (1969). Systems of frequency curves. *Cambridge.*

Hinchliffe, S. R., Scott, D. A., & Lambert, P. C. (2013). Flexible parametric illness-death models. *The Stata Journal*, *13*(4), 759-775.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime data analysis*, *5*, 239-264.

Lauseker, M., & Zu Eulenburg,C. (2019). Analysis of cause of death: Competing risks or progressive illness-death model. *Biometrical Journal*, *61*(2), 264-274.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, *18*(2), 195-222.

Touraine, C., Helmer, C., & Joly, P. (2016). Predictions in an illness-death model. *Statistical methods in medical research*, *25*(4), 1452-1470. *University Press*.

Touraine, C., Gerds, T. A., & Joly, P. (2017). Smooth Hazard: An R package for fitting regression models to interval-censored observations of illness-death models. *Journal of Statistical Software*, *79*, 1-22.

Zhang, Y., & Xu, R. (2022). Marginal Structural Illness-Death Models for Semi-Competing Risks Data. *arXiv preprint arXiv:2204.10426*.