Hipotenusa : Journal of Mathematical Society
Volume 4 Issue 2 Year 2022
ISSN: 2716-3156
Website: https://hipotenusa.iainsalatiga.ac.id/index.php/hipotenusa/index

The Evolution of the Sickle Cell Anaemia in the Region of Kinshasa: a Mathematic Modelling

R. Gilles.Bokolo[1*], Musema Sinamuli[2], Dibere S. Kodondi[3], Bruce S. Mafuta[1]
[1] Department of Mathematics, Faculty of Sciences, University of Kinshasa, D.R.Congo
[2]Department of Physics, Faculty of Sciences, University of Kinshasa, D.R. Congo
[3]Institut de Recherche en Sciences de la Santé, Kinshasa I, D.R. Congo

*Corresponding Author. E-mail: gilles.bokolo@unikin.ac.cd
DOI: 10.18326/hipotenusa.v4i2.7322

Article submitted : May 23, 2022
Article revised : October 17, 2022
Article accepted : November 31, 2022

Abstract

We attempt to study the evolution of the sickle cell anaemia in the region of Kinshasa by proposing three models based on Markov processes. The first two models namely, the idealistic and the quasi- idealistic models fail to completely describe the data sampled. The third model, the realistic one, leads to results that match the data collected and meets the predictions of serious institutions such as the World Health Organization (about 36 percent of the population belongs to the electrophoresis categories AS or SS).

Keywords: *markov chain, markov processes, sickle cell anaemia, electrophoresis, heterozygous, homozygous, alpha and beta chains*

INTRODUCTION

The evolution of the human race throughout the different ages has also been characterized by the rapport of humans with diseases. Some of the diseases are hereditary or have been passed on by parents to their offspring. These diseases have also significantly affected the way people, who have been exposed to them, have organized their societies depending on them. More precisely, the way these diseases have impacted on their production, health, life expectancy and so on.

Because of that reason many societies had figured out several ways to overcome or to tackle these diseases. One way was to study their evolution over a long period of time in order to find some preventive methods to get rid of them.

The above statement also explains the fact that one of the most exciting and appealing fields that researchers have been working on these recent years is on the modelling of the evolution of some diseases. The diseases that have been mainly observed are those that have the reputation to be genetics ones, that is, those that are passed on to a generation by the precedent one. One of the most renowned genetic diseases is the sickle cell anaemia and it is driving the attention of many researchers as it rises a good number of questions, especially on its origin.

In the current project the focus will not be on the potential causes of the disease, it will rather be on its evolution in a particular region of the globe. To this end, a model of study will be attempted and will consist to look at the evolution of the disease as a process which can be approximated to a stochastic one. The different states of the population of a generic region with rapport to the disease can be seen as configurations of a chain which can be assimilated to a Markov chain. From this chain, it will be possible to figure out a matrix of the process that will allow to predict the evolution of the disease over a large interval of time (generations) under some initial hypotheses.

For the sake of contributing to the improvement of the disease prevention care rather than the therapeutic one. In section 2, we will give a brief insight on sickle cell anaemia, particularly on how it is manifest on people in term of genes. We will particularly stress in section 3 on the configuration of the disease in the region we are studying (region of Kinshasa). The section 4 is devoted to the evolution of the disease in the region of Kinshasa over a period of time based on three proposed models. A conclusion and discussions on results obtained will constitute our section 5. The last section will be a short appendix on the composition of the population of the region of Kinshasa in order to justify our starting hypotheses.

An Insight on Sickle Cell Anaemia

Sickle cell anaemia is a common genetic disease due to a haemoglobin anomaly. It is characterized by a modification of the red corpuscle shape, which normally biconcave, takes a croissant or sickle shape after which it has been named (" drepanos" in Greek). It is due to a punctual mutation of the sixth codon of the adenine and its replacement by the thymine in the globin beta chain. This desoxyribonucleic acid mutation is characterized by the substitution of a amino acid, the glutamic acid by the valine, in the sixth position of the beta chain. The anomaly is in chromosome 11 [1–3].

The genetic transmission of the disease does not depend on the sex and occurs when two chromosomes passed on by the parents bear the disease gene. In the case where only one chromosome bears the gene of the HbS (passed on by either the mother or the father), the subject is said heterozygous and healthy.

The human haemoglobin is made of two protein chains namely: the alpha and beta chains. The disease mainly concerns the beta chain of the haemoglobin. The alpha chain of the haemoglobin is not concerned (Gigoyan et al. 2000). Thereby two situations may occur: 1) the AS form (heterozygous) or minor sickle cell anaemia syndrome: asymptomatic, human reservoir responsible for the transmission of the disease; and 2) the serious form or major sickle cell anaemia syndrome: symptomatic, developing the disease with different genotypes. The SS (homozygous) is the most common form in Democratic Republic of Congo.

Sickle cell anaemia is the first genetic disease, reaching about fifty million people worldwide. It is a haemoglobin pathology which constitutes a real public health issue and is observed worldwide. The disease is very frequent amongst people from sub-Saharan Africa, India, Saudi Arabia and Mediterranean countries. The African union assembly, in its fifty ordinary sessions, has added the sickle cell anaemia on the list of public health priorities. According to the World Health Organization (WHO), every year three hundred thousand children are born with a major anomaly of the haemoglobin and over two hundred thousand cases have been observed only in Africa. For the African continent, the WHO reports a prevalence of 13 percent. In the Democratic Republic of Congo, the prevalence of the sickle cell character varies from 30 to 40 percent [4, 5] and the homozygous subjects (SS) represent 0.8 to 2 percent of the population. Numerous national as well as international associations have arisen since.

Actual Configuration of The Disease in The Region of Kinshasa

As mentioned earlier in the introduction, the current section is mainly devoted to the configuration of the dis- ease in these current days. To this end, we have been considering a representative sample, in which, are included the age, sex, district and electrophoresis of patients of the" Centre de Medecine Mixte et d'Anemie", located at Kinshasa in the Democratic Republic of Congo. In the afore- mentioned sample are fairly represented the 24 districts of Kinshasa, the patients' ages range from 1- to 75-year-old approximately and there are consistent enough numbers of male and female individuals.

The electrophoresis probability distributions are summarized in the table 1, table 2 and table 3.

Table 1. The electrophoresis of a population whose age varies from 1- to 25-year-old

1-25	AA	AS	SS
Female	442	178	450
Male	281	108	394
Total	723	286	844

Table 1: The above table provides the electrophoresis of a population whose age varies from 1- to 25-year-old. These data have been sampled within a period going from January 2015 to June 2017. It is clear to notice that the number of SS is significantly greater than the two other groups. This prevalence can be explained by the fact that most of the patients presenting symptoms of the anaemia are those who are under 25 years.

Table 2. The electrophoresis of a population whose age is located within an interval of 26- to 50-years-old

26-50	AA	AS	SS
Female	358	133	33
Male	375	163	31
Total	733	296	64

Table 2: The current table gives the electrophoresis of a population whose age is located within an interval of 26- to 50-years-old. These data have been sampled within a period going from January 2015 to June 2017.

In order to make our study possible we have divided our sample up to three age subgroups of 25 years, that is, from 1 to 25 years, 26 to 50 years, and 51 to 50 years. These age groups are regarded as three different generations (as we consider a generation to be defined by an interval of 25 years). To better understand how the prob- abilities of belonging in a given electrophoresis category vary from a generation to another; we model the evolution using Markov processes by finding probability matrices linking one probability configuration to another.

Evolution of The Disease Over a Period of Seventy-Five Years

Here, we focus on the evolution of the disease during a period of 75 years in the region of Kinshasa. Prior to proceed to the study we consider some hypotheses to enable a mathematical modelling of the problem. We suppose that the population of the region

is" homogeneous" which, means that the diverse migrations of the population do not affect by many order the probability distributions of electrophoresis.

Table 3. The electrophoresis of a population whose age ranges from 51- to 75-year-old

51-75	AA	AS	SS
Total	132	70	9

Table 3: In the table are given the electrophoresis of a population whose age ranges from 51 to 75 year old. These data have been sampled within a period going from January 2015 to June 2017.

75 years interval into three intervals of 25 years by assuming that a generation characterizes a group of individuals whose ages span over a period of 25 years.

To start our analysis, we first drop the notations of the homozygous and heterozygous AA, AS and SS to substitute them by a, b and c, respectively. From the table I we can infer that the electrophoresis distributions of patients of 1- to 25-year-old read as

$$p_{a\leq 25} = 0.39 \qquad p_{b\leq 25} = 0.15 \quad p_{c\leq 25} = 0.46 \qquad (1)$$

It is easy to notice that these probabilities do not represent whole region of Kinshasa because most of the patients in that age interval are those who present the symptoms of the disease. However, we can still assume that the relative proportions between patients of electrophoresis *a* and *b* constitute a representative sample (in term of relative proportion) of the a and b population for the whole region of Kinshasa.

The second group or generation consists of a popula- tion whose ages goes from 26 to 50 years. The probability distributions as given in table 2 are

$$p_{25<a\leq 50} = 0.67 \qquad p_{25<b\leq 50} = 0.27 \quad p_{25<c\leq 50} = 0.06 \qquad (2)$$

As we remark, there is a small portion of the population suffering from sickle cell anaemia. Indeed, this sample makes sense since most of the people developing the disease die in their young age (1- to 10-year-old). Therefore, we definitely assume that the above electrophoresis probability distribution represents not only the patients but also the whole population of the region of Kinshasa.

The last group represents the oldest population of the sample whose age varies from 51 to 75 years. The electrophoresis probability distributions for patients in this group are provided in table III as follows

$$p_{a\leq 50} = 0.63 \qquad p_{b\leq 50} = 0.33 \quad p_{c\leq 50} = 0.04 \qquad (3)$$

representative sample of the region of Kinshasa as well. From these above tables it is easy to observe the age pyramid of the population of the region of Kinshasa. These results are congruent with most of the studies done on the age pyramids of the sub-Saharan regions.

Now, let us consider an individual with a given electrophoresis state (a, b, c), the probability that he has an offspring of electrophoresis (a, b, c) after meeting a partner of electrophoresis (a, b, c) is given by the following table

.	a	b	c
a	a	$\frac{1}{2}(a+b)$	b
b	$\frac{1}{2}(a+b)$	$\frac{1}{4}(a+2b+c)$	$\frac{1}{2}(b+c)$
c	b	$\frac{1}{2}(b+c)$	c

This can be reinterpreted as follows: an individual of electrophoresis a after meeting a partner of electrophoresis. a, b, c will procreate individuals of electrophoresis (a, b) with equal probabilities. Likewise, an individual of electrophoresis b will procreate individuals of electrophoresis (a, b, c) with (a, c) equally probable and b twice their probability. Finally, an individual of electrophoresis c will procreate individuals of electrophoresis (b, c) with equal probabilities.

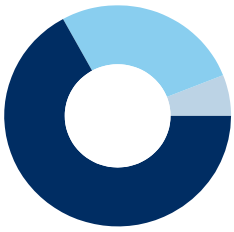

Figure 1: This Pie Chart represents the electrophoresis probability distributions for the patients whose age is between 26 to 50 years. AA are in dark blue, AS in blue and SS in light blue.

We can summarize the above statement by the equation

$$\begin{array}{lll} a & \rightarrow & \frac{1}{2}a+\frac{1}{2}b \\ b & \rightarrow & \frac{1}{4}a+\frac{1}{2}b+\frac{1}{4}c \\ c & \rightarrow & \frac{1}{2}b+\frac{1}{2}c \end{array} \qquad (4)$$

Using equation (4), we can write an" idealistic" Markov's matrix for the process, which will help us to model the evolution of the population electrophoresis from a generation to another.

When considering the electrophoresis probability distributions for the group of 26 to 50 years, introduced in (2) and represented in figure 1, we find that the next generation probability distributions look like [6–8]

$$[0.670.270.06].\begin{bmatrix} O.50 & 0.50 & 0 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0.50 & 0.50 \end{bmatrix} = [0.40 \quad 0.50 \quad 0.10] \rightarrow [0.25 \quad 0.50 \quad 0.25] \quad (5)$$

The line before the last line of (5) are the probability distributions for the next generation (1 to 25 years) and the last line represent the final probability distributions (see figure 2) that will occur around the next 10 generations. It happens that this" idealistic" model does not represent the reality since the results it provides do not match with the data collected (noting that the ratio $p_a/p_b = 4/5$ while the expected one is 39/15). In fact, the distributions (1) are not even close to that of us" idealistic" model predicts.

To tackle this problem we will consider a "quasi idealistic" model, in which, we assume that the population suffering from the sickle cell disease does not procreate.

This is justified by the fact that most of the individuals presenting the disease symptoms have died before reaching 26-year-old or do not want to take the risk of procreating.

The relative electrophoresis probability distributions (p_a, p_b) derived from the table III, II and I read as

$$p_0 = [0.66 \quad 0.34]$$
$$p_1 = [0.71 \quad 0.29] \qquad (6)$$
$$p_2 = [0.72 \quad 0.28]$$

respectively.

Thus, we will only consider the table

.	a	b
a	a	$\frac{1}{2}(a+b)$
b	$\frac{1}{2}(a+b)$	$\frac{1}{4}(a+2b+c)$

and equation (4) will be reduced to

$$\begin{aligned} a &\rightarrow \frac{3}{4}a+\frac{1}{4}b \\ b &\rightarrow \frac{3}{8}a+\frac{1}{2}b+\frac{1}{8}c \end{aligned} \qquad (7)$$

Discarding the *c* electrophoresis from the table, it results the following Markov's matrix

$$M=\begin{bmatrix} 3/4 & 1/4 \\ 3/7 & 4/7 \end{bmatrix} \qquad (8)$$

We obtain for example that from the (51 to 75 years) group to the (26 to 50 years) one

$$[0.66 \quad 0.34].\begin{bmatrix} 3/4 & 1/4 \\ 3/7 & 4/7 \end{bmatrix}=[0.64 \quad 0.36] \qquad (9)$$

It seems from the above computation that we have a better approximation here since the ratio $p_a/p_b = 16/9$ while the expected one is 71/29. This only tells us about the relative probability distributions between the electrophoresis a and b. What about the *c*?

To answer to this question, we label the electrophoresis probability distributions of the ascendants by (p_a^0 , p_b^0, p_c^0) with ($p_c^0 = 0$) and the ones of descendants by (p_a^1 , p_b^1, p_c^1).

We find after using (7) that

$$\begin{aligned} p_a^1 &= \frac{3}{4}p_a^0+\frac{3}{8}p_b^0 \\ p_b^1 &= \frac{1}{4}p_a^0+\frac{1}{2}p_b^0 \\ p_c^1 &= \frac{1}{8}p_b^0 \end{aligned} \qquad (10)$$

Hence

$$p_a^1+p_b^1+p_c^1=p_a^0+p_b^0+p_c^0 \quad (11)$$

and

$$\begin{bmatrix} p_a^1 \\ p_b^1 \\ p_c^1 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{3}{8} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{8} & 0 \end{bmatrix} \begin{bmatrix} p_a^0 \\ p_b^0 \\ p_c^0 \end{bmatrix} \qquad (12)$$

This expression helps us to recover the total electrophoresis probability distributions (pa; pb; pc) from the relative one $(p_a; p_b)$ (this applies when we have data which do not represent the whole population of the region as in table I). As an example, when starting with the electrophoresis probability distributions of patients of the group (26 to 50 years) we obtain for the group (1 to 25 years):

$$\begin{bmatrix} \frac{3}{4} & \frac{3}{8} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{8} & 0 \end{bmatrix} \begin{bmatrix} 0.71 \\ 0.29 \\ 0.00 \end{bmatrix} = \begin{bmatrix} 0.64 \\ 0.32 \\ 0.04 \end{bmatrix} \qquad (13)$$

This gives approximately the electrophoresis probability distribution of the group (1 to 25 years) who's the sample is not representative of the region population.

So far, the two models that we have used lead to results which do not match the sampled data. Although the" quasi-idealistic" model with a reduced matrix brings results that are closer to the sampled data, it still does not correctly model them.

From the relative probability distributions (6), we can infer that a" realistic" model characterized by the Markov's process who's the matrix reads as

$$M = \begin{bmatrix} 0.75 & 0.25 \\ 0.64 & 0.36 \end{bmatrix} \qquad (14)$$

provides results that match the data collected. It easy to check that the matrix M satisfies the following relations

$$p_0 = p_0 \qquad p_1 = p_0 M$$

$$p_2 = p_0 M^2 \quad \ldots \quad p_n = p_0 M^n \quad (n \geq 2) \qquad (15)$$

where

$$p_n = [0.72 \quad 0.28] \qquad (16)$$

Equations (15) and (16) predict the evolution of the disease in the region of Kinshasa insofar we consider the population thereof to be" homogeneous". The trend of the electrophoresis probability distributions stemming from this model is illustrated in figure 3.

DISCUSSIONS AND CONCLUSION

The method used in the current project has led us to some important results that are very close to the statistic predicted by many credible institutions. We have been able to propose three models namely the idealistic, quasi- idealistic and realistic models.

In the idealistic model, we assumed that all the population, including the homozygous ones, were able to procreate without problem. This has led us to predictions that failed to be conform with the data sampled as it predicts that the homozygous population will reach 25 percent of the total population within a number of generations.

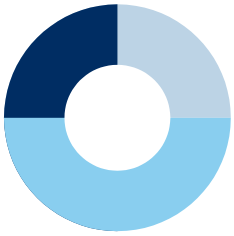

Figure 2. The above Pie Chart depicts the idealistic configuration for the electrophoresis probability distributions of the population whose age varies from 1 to 25 years. AA being in dark blue, AS in blue and SS in light blue.

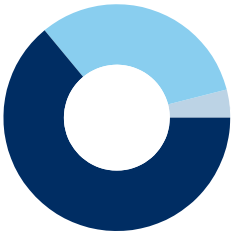

Figure 3. The current Pie Chart gives the" realistic" electrophoresis probability distributions of patients with age between 1 to 25 years. This diagram is based on a model in which the population suffering from sickle cell anaemia do not procreate. AA are represented in dark blue, AS in blue and SS in light blue

The second model, which is a quasi-idealistic one, pro- vided us with better predictions than the idealistic model but has not matched the expected result. Indeed, in

this model we supposed that the homozygous population does not procreate due to the fact the most of the individuals developing sickle cell anaemia symptoms either die when they are too young or decide not to procreate. De- spite the fact that in some regions in Europe studies have proved that the number of early deceases has considerably dropped [9] it is still fair to assume that early deceases is a reality not to be neglected in the region being studied in the current project.

The last model is the realistic one, in which we generated an approximative matrix of the process that is congruent with the data collected. In fact, that matrix has been proposed so that it leads to results which match the data sampled. This model allows us to predict the evolution of the disease over a period spanning more than one generation.

The upshot of this model is that the relative probabilities between the population of type AA and AS tend to 0.72 and 0.28, respectively. This corresponds to prob- abilities (for populations under 26 years) of 0.64, 0.32 and 0.04 for populations of type AA, AS and SS, respectively. We therefore consider that this result is reliable because it is not that far from the ones given in previous works (30 to 40 percent of the prevalence of the sickle cell character) [4, 5].

Further studies based on the third model can be undertaken on a bigger sample collected over a longer period of time.

ACKNOWLEDGEMENTS

We are very thankful to the Centre de Medicine Mixte et d'Anemie for providing us with the data.

COMPOSITION OF THE POPULATION OF THE REGION OF KINSHASA

The purpose of this appendix is to give a glimpse on the composition of the population in the region being studied. Prior to start our study numerous restrictions have been imposed on the model and one of them was on the homogeneity of the population of the region. In fact, in order to be able to conduct an investigation on the evolution of the disease in the region we considered that there were not consistent enough interactions between the populations established in the region and those coming from remote regions. As we can see in recent works [10], it has been shown that only 2 percent of the population of the region of Kinshasa are foreigners and over 75 percent of them come from neighboring countries. Since the populations of neighbor countries share

similar environmental realities, it is enough to assume that the population of the region Kinshasa is homogeneous.

REFERENCES

[1] Berchel C. Diarra, J.P. Loreth et al, "Histoire naturelle de la drepanocytose", Rev.Prat. 1992; 42: 1885-91.

[2] I. Okpala," Management of pain in sickle cell disease", Journal of the royal society of medicine 2002, 95: 456-58.

[3] F. Galacteros, "La drepanocytose: physiopathologie, physiologie et diagnostic", Rev.Prat. 1995; 45: 351-60.

[4] OMS, "Drepanocytose: Une strategie pour la region africaine de l'OMS", Organisation Mondiale de la Sante.

[5] S. Latoundji, L. Anani, E. Ablet, and I. Zohoun, "Morbidite et Mortalite drepanocytaire au Benin".

[6] P. A. Gagnius, "From Theory to Implementation and Experimetation",USA NJ: John Wiley and Son pp, 9-11. ISBN 978-1-119-38755-8 (2017).

[7] S. R. Asmussen, "Markov Chains", Applied Probabilty and Queues. Stochastic Modelling and Applied Proba- bility. 51. pp. 3-8. doi: 10. 1007/0-387-21525-5-1. ISBN 978-0-387-00211.8.

[8] B. Hayes, "First links in the Markov Chains", American Scientist. 101(2):92-96.

[9] E. Gomes, K. Castetbon, V. Goulet, "Sickle-cell related

mortality in France: Age and cause of death (1979-2010)".

[10] PNUD, "Province de Kinshasa: Pauvreté et conditions de vie des ménages", Programme des Nations Unies pour le Développement Unité de Lutte Contre la Pauvreté.