FACTORS ASSOCIATED WITH SURVIVAL TIME FOR PATIENTS WITH HIV/AIDS IN THE STATE OF MATO GROSSO DO SUL: PARAMETRIC APPROACH

Marcos Vinicius BUENO[1]
Robson Marcelo ROSSI[1]

■ ABSTRACT: The goal of this study was to use frequentist and Bayesian methodologies to adjust some probability distributions for survival time in HIV/AIDS patients in Mato Grosso do Sul, Brazil, followeds from 2009 to 2018. The influence of explanatory variables on the response variable can be calculated using regression models. The Log-Normal distribution was shown to be the most parsimonious for the data using the Akaike information criterion (AIC) values and the maximum likelihood logarithm. Two regression models were built based on the described methodologies, converging to the same interpretation of the explanatory variables: sex, race, education, and injecting drug use. The median time to death from HIV/AIDS is approximately: 2.1 higher for females, 1.8 higher for white people, 5.4 higher for individuals with more than 8 years of education, 5.5 higher for individuals who do not use injecting drugs, according to the study. Based on the interpretations of the coefficients of the model parameters, the need for prevention and early diagnosis policies focused on groups that have a shorter median survival time after notification of HIV infection can be discussed.

■ KEYWORDS: Bayesian inference; failure times; probabilistic models; survival analysis.

1 Introduction

Survival analysis consists of the study of the length of time until the observation of an event of interest, which is called failure time. Some examples can be given, such as the time until death, cure, or relapse of a disease in a patient.

According to Colosimo and Giolo (2006), the main characteristic of survival data is the presence of partial observations of responses, called censored lifetimes.

[1]Universidade Estadual de Maringá - UEM, Departamento de Estatística, CEP: 87.020-900, Maringá, PR, Brasil. E-mail: *viniciusbueno201@outlook.com, rmrossi@uem.br*

Thus, the term survival analysis refers to medical situations that involve censored data.

In a historical context, some important dates for the development of this type of statistical analysis were presented by Bastos and Rocha (2006), which are the publication of the book "Natural and Political Observations upon the Bill of Mortality", by John Graunt in 1962 and the development of the first mortality table in 1693 by Edmund Halley, similar to those currently used in demography.

Through regression models, it is possible to verify the effect of explanatory variables on the response variable, in this case, time. In the survival analysis, there are two classes of proposed models in the literature: the semiparametric Cox model and the parametric models.

To use parametric models, it is necessary to assume that the response variable follows a probability distribution. The model is composed of a random component, which describes the survival time behavior in terms of probability, and a systematic component, which describes the relationship between the parameters of the probabilistic model and the explanatory variables (CARVALHO *et al.*, 2011).

The Bayesian models have gained greater relevance over the years in literature thanks to the advancement of computational power for carrying out the analyses. In Bayesian inference, it is considered that the parameters under study are random and characterized through an a priori distribution, and from this and the sample data, posterior estimates are found using the Bayes rule (ROSSI, 2011).

Acquired immunodeficiency syndrome (AIDS) is a chronic disease where the immune system is damaged which is caused by the human immunodeficiency virus (HIV). The 2020 HIV/AIDS Epidemiological Bulletin presents data on HIV and AIDS in Brazil. According to it, 41,909 new cases of HIV and 37,308 cases of AIDS were diagnosed in 2019. From 1980 to June 2020, a total of 1,011,617 AIDS cases were identified in Brazil. Regarding mortality in the country, from 1980 to 2019, 349,784 deaths with HIV/AIDS as the underlying cause were registered, of which 57.7% belong to the South region, 17.8% to the Northeast region, 5.3% to the Center-West region, and 5.3% the North region (BRASIL, 2020).

In a discussion of clinical aspects, Neto *et al.* (2021) argue that HIV infection can be classified into three phases, named: Acute HIV Infection, which occurs between the first three weeks of infection and presents nonspecific symptoms such as fever, headache, asthenia, adenopathy, among others; Clinical latency, generally asymptomatic and lasting for years, generating a gradual drop in LT-CD4+ (defense cells), with the intermittent appearance of infections or reactivation of old infections; AIDS, at this stage, there are manifestations of advanced immunodeficiency, in which its indications are the appearance of opportunistic infections or neoplasms.

Ribeiro *et al.* (2019) present a spatial analysis of data on the incidence of AIDS by States in Brazil from 1992 to 2017, allowing to visualize the evolution of the disease over the years through graphic techniques. Among the results presented, the appearance of the first cases in the 1980s in the Southeast region is discussed along with its spread to the South and Center-West regions in the 1990s and the North and Northeast from 2000 on.

Although Colosimo and Giolo (2006) report that survival analysis has been one of the fastest-growing statistical areas in the last two decades of the last century, there is still a paucity of studies on HIV/AIDS data in the parametric context.

Ergo, this study aimed to adjust a probability distribution for the time (in weeks) from diagnosis to death of patients with HIV/AIDS using data from The State of Mato Grosso do Sul, in Brazil, followeds from 2009 to 2018, as well as verifying which explanatory variables, among some considered in the literature, are relevant to the model and indicate an association with the response variable using frequentist and Bayesian procedures in the context of survival analysis.

2 Materials and Methods

The analyzed dataset refers to an epidemiological, cross-sectional, and analytical study with data from HIV/AIDS notifications from 2009 to 2018. The data are addressed by Werle (2021), initially containing a total of 9,021 individuals from the Federative Unit of Mato Grosso do Sul, Brazil. When checking the dataset, observations were found that presented the same day for notification and death (0 days until death from HIV/AIDS), in addition to missing information for some explanatory variables. It was decided, then, to remove these observations, and thus, the dataset had 4,681 observations. The study originating from the data met the national and international standards of ethics in research with human beings and it was approved with opinion number 3,789,678. In the present study, the appreciation of the Research Ethics Committee (CEP), the National Health Council (CNS), and the National Research Ethics Committee (CONEP) is dispensed, based on Resolution 466/2012 CEP-CONEP.

In addition to the time (in weeks) and the failure or censored lifetime indicator, the dataset includes information on some explanatory variables such as age group; sex; breed; education; sexual transmission; vertical transmission; injecting drug use; blood transfusion for hemophilia; blood transfusion; accident with biological material. In Table 1, the classes of explanatory variables are presented and the absolute and relative frequencies of each one.

The main characteristic of survival data is the presence of partial observations of the response, called censored lifetimes. A censored observation indicates that the individual in question, for some reason, did not present the event of interest. Some examples may be the individual having died for another cause, moving to another city, or the study had ended before the event of interest occurred. The presence of a censored observation time indicates that the failure time is longer than that observed (COLOSIMO; GIOLO, 2006).

Table 1 - Descriptive summary for explanatory variables

Explanatory variables	Categories	Frequency	%	Censored lifetimes (%)
Age group (years)	10 - 14	9	0.19	100.00
	15 - 19	236	5.04	97.88
	20 - 34	2,160	46.14	94.35
	35 - 49	1,562	33.37	88.35
	50 - 64	613	13.10	83.69
	65 or more	101	2.16	80.20
Sex	Male	2,953	63.08	90.45
	Female	1,728	36.92	91.49
Race	White	2,068	44.18	92.12
	Non-white	2,613	55.82	89.82
Education (years)	> 8	1,986	43.43	94.31
	≤ 8	2,695	57.57	88.27
Probable mode of transmission				
Sexual	With men	2,662	56.87	93.58
	With women	1,783	38.09	86.76
	With both	236	5.04	90.68
Vertical	Yes	43	0.92	93.02
	No	4,638	99.08	90.81
Injecting drug use	Yes	153	3.27	80.33
	No	4,528	96.73	91.19
Blood transfusion for hemophilia	Yes	2	0.04	100.00
	No	4,769	99.96	90.83
Blood transfusion	Yes	30	0.64	93.33
	No	4,651	99.36	90.82
Accident with biological material	Yes	2	0.04	100.00
	No	4,769	99.96	90.83

In the context of survival analysis, the failure time is usually specified by the survival and hazard functions. The survival function is defined as the probability that the observed individual survives a specific time t, in probabilistic terms, we have $S(T) = P(T \geq t)$. Consequently, the cumulative distribution function allows finding the probability that the individual will not survive the time t, that is, $F(t) = 1 - S(t)$.

The hazard function, or failure rate function, describes how the failure rate changes over time. It is necessary to emphasize that this function provides the value of a rate and not a probability, and can assume any positive value other than 0. Mathematically, it is expressed by: $h(t) = \lim_{\Delta t \to 0} P(t \leq T < t + \Delta t | T \geq T)/\Delta t$. From this, the cumulative hazard function can be found by integrating the hazard

function into an interval $[0, t]$.

Among the characteristic functions in survival analysis, it is possible to verify some important mathematical relationships between the functions, such as the hazard function $h(t) = f(t)/S(t)$. These relationships are described and discussed in Lee and Wang (2003), showing that it is possible to obtain the others from the knowledge of one.

Although the hazard function is also of great importance in survival analysis, in this study, we chose to discuss only the survival function.

An important nonparametric technique in survival analysis is the Kaplan and Meier (1958) estimator, also known as the limit-product estimator. This estimator is an adaptation of the empirical survival function and uses concepts of conditional probability and event independence, and defined as: $\prod_{i:t_i<t}(1-(d_i/n_i))$, on what $t_1 < t_2 < ... < t_k$ is considered the k distinct and ordered failure times, d_i the number of failures at a time t_i and n_i the number of individuals at risk in t_i.

The Cox proportional hazards model, proposed by Cox (1972), allows estimating the effects of explanatory variables without the need to make assumptions regarding the probability distribution of survival time. According to Carvalho *et al.* (2011), this model is called semiparametric because it does not assume any probabilistic distribution for the baseline hazard function, $h_0(t)$. Generally speaking, considering p explanatory variables and the component vector $\boldsymbol{x} = (x_1, ..., x_p)'$, the general expression of the Cox regression model is given by: $h(t|\boldsymbol{x}) = h_0(t)g(\boldsymbol{x'\beta})$. According to Colosimo and Giolo (2006), the parametric component is often used in the following multiplicative form: $g(\boldsymbol{x'\beta}) = \exp\{\boldsymbol{x'\beta}\} = \exp\{\beta_1 x_1 + ... + \beta_p x_p\}$.

Nonparametric methods and/or Cox regression models are more common in data analysis of patients with HIV/AIDS, as seen in Colosimo and Vieira (1996), Medeiros *et al.* (2017), Müller and Borges (2020), and Melo *et al.* (2017). Overall, the purpose of these analyzes is to look for associations between explanatory variables and disease without considering that the data follow a probability distribution.

On the other hand, we have the parametric regression models, which are more efficient but less flexible than the Cox model. To adjust such models, it is necessary to assume a probability distribution for the response variable, which defines the survival function. For the analyzed data in this study, the following probability distributions were considered: Exponential, Weibull, Gamma, Generalized Gamma, Log-Normal, and Log-Logistic.

In contrast to the frequentist methodology, there are the Bayesian models, in which the parameter (considered random effect) is quantified in terms of probability, and formally it is said that it follows a prior distribution, in which, based on the sample and prior information, it is possible to model and update the estimates of the posterior parameters using the Bayes rule. Applications of Bayesian inference in the context of survival analysis can be seen in Santos and Achcar (2011), Brunello and Nakano (2015), among others.

To select the best probabilistic model, the Akaike information criteria (AIC)

will be used, in addition to the logarithm of the maximum likelihood function ($\log L(\theta)$). The probability density functions for each probabilistic model are presented in Table 2, so that their survival, hazard, and cumulative hazard functions can be obtained based on the relationships between functions presented in Lee and Wang (2003).

Table 2 - Probability density function of the considered probabilistic models

Model	Probability density function
Exponential	$f(t) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}$
Weibull	$f(t) = \frac{\gamma}{\alpha^\gamma} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}$
Gamma	$f(t) = \frac{t^{k-1}}{\Gamma(k)\alpha^k} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}$
Generalized Gamma	$f(t) = \frac{\gamma}{\Gamma(k)\alpha^{\gamma^k}} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}$
Log-Normal	$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right\}$
Log-Logistic	$f(t) = \frac{\gamma}{\alpha^\gamma} r^{\gamma-1}\left[1+\left(\frac{t}{\alpha}\right)^\gamma\right]^{-2}$

To verify the probability distribution for the response variable, as well as the parametric regression model, the statistical environment R, (R CORE TEAM, 2021), as well as the statistical packages survival (THERNEAU, 2020) and flexsurv (JACKSON, 2016) was used. To estimate the parameters of the regression models, the maximum likelihood method was used for the frequentist model and integrated nested Laplace approximation, through the INLA package Rue (RUE *et al.*, 2019), for the Bayesian model.

From the best probability model adjusted to the data, the objective is to carry out two regression models, approaching frequentist and Bayesian methodologies.

In the context of survival analysis, the likelihood function for regression models is defined by:

$$L(\theta) = \prod_{i=1}^{n} \left[f(t_i|x_i)\right]^{\delta_i} \left[S(t_i|x_i)\right]^{1-\delta_i}, \tag{1}$$

where δ_i is the failure or censored lifetime indicator.

The interpretation of the estimated coefficients for the Log-Normal and Weibull models is discussed in Colosimo and Giolo (2006) so that for a binary variable, it is necessary to use the ratio of median times:

$$\frac{t_{0,5}(x=1,\hat{\beta})}{t_{0,5}(x=0,\hat{\beta})} = e^{\hat{\beta}} \tag{2}$$

This interpretation can also be extended to categorical and continuous covariates. A rigorous discussion of the interpretation of estimated results can be seen in Hosmer and Lemeshow (1999).

Residual analysis is an essential step in constructing a regression model since evaluating the adequacy of the adjusted model is a fundamental part of data analysis. To verify the fits of the models, the Cox-Snell residuals proposed by Cox and Snell (1968) were used.

The verification technique consists of calculating the residuals based on the cumulative hazard function, obtained through the estimated model ($\hat{e}_i = \hat{\Lambda}(t_i|x_i)$), and observing if the residuals follow a standardized Exponential distribution (LAWLESS, 2003).

3 Results and Discussion

Initial data analysis consisted of verifying the survival curve estimated by Kaplan-Meier, presented in Figure 1. A sharp drop in survival due to HIV/AIDS in the first weeks can be observed.

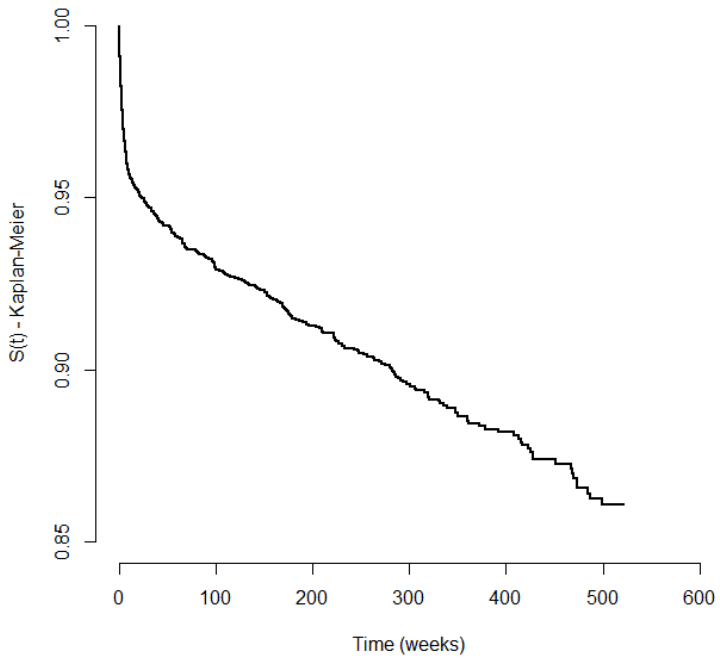

Figure 1 - Empirical HIV/AIDS survival curve.

After frequentist adjustments, considering the distributions shown in Table 2, the results (Table 3) point to a better adequacy of the Log-Normal distribution for the response variable, as this model presented the smallest AIC (6,879.91) and the largest $\log L(\theta)$ (-3,437.95). Based on this, it was then considered that the

response variable follows a Log-Normal probability distribution for the next step of the analysis.

Table 3 - Results for the AIC and $\log L(\theta)$ for the considered probabilistic models

Model	AIC	$\log L(\theta)$
Exponential	7,538.97	-3,768.48
Weibull	6,902.78	-3,449.39
Gamma	6,904.64	-3,450.32
Generalized Gamma	6,888.81	-3,441.40
Log-Normal	**6,879.91**	**-3.437.95**
Log-Logistic	6,901.08	-3,448.54

In particular, the Log-Normal regression model is given by:

$$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(t) - \mu(\boldsymbol{x'\beta})}{\sigma}\right)^2\right\}, \quad t > 0, \tag{3}$$

which σ is the dispersion parameter and μ is the location parameter, such that $\mu(\boldsymbol{x'\beta}) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$, through the identity link function.

Santos and Nakano (2015) use the Log-Normal regression model to compare with the Cox model, working with data on the length of stay of workers in the labor market in the Federal District, Brazil.

Another application of the Log-Normal regression model can be seen in Silva *et al.* (2018). The authors, in the context of survival analysis, work with data on the dropout rate of students in the Statistics course at the Federal University of Paraíba.

In a Bayesian context, for the Log-Normal model, the following prior distributions are considered default according to the INLA parameterization in R:

$\tau \sim$ Log-Gamma$(1; 0,00005)$;

$\beta_0 \sim N(0; 0,001)$;

$\beta_j \sim N(0; 0,001), \quad j = 1, 2, ..., p,$

which $\sigma = 1/\sqrt{\tau}$, β_0 is the intercept and β_j are the parameters associated with each explanatory variable.

Werle (2021), through a Cox regression model on the data, performs the verification of the significance of the parameters associated with the explanatory variables at the level of 5% of significance. Based on the results, a discussion about the risk of death for individuals from certain groups present in the study is carried out. Significant explanatory variables for the final adjusted model are pointed out: sex, sexual transmission, race (in another codification), education, and injecting drug use.

Based on the results presented by Werle (2021), different models were tested, in which it was verified that the presence of the explanatory variable sexual transmission caused the explanatory variable sex to have an opposite effect to that

observed in the empirical survival curves. Thus, the final models had the following explanatory variables: gender, race, education injecting drug use.

Parametric regression models were built for the data in question, using the Log-Normal distribution, considering the frequentist and Bayesian approaches. The parameter estimates for both models are shown in Table 4. The results show the similarity between both adjustments.

The Cox-Snell residuals estimated by the Kaplan-Meier method and the standardized Exponential model, as well as the estimated survival curve for the Bayesian model are shown in Figure 2. It was decided not to present Cox-Snell residual plots for the frequentist model, as they present behavior analogous to Bayesian.

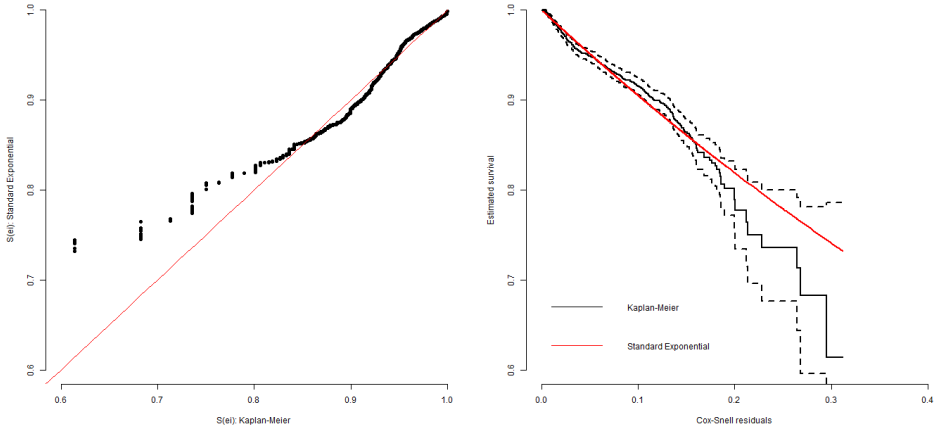

Figure 2 - Cox-Snell residuals of the Bayesian regression model estimated by the Kaplan-Meier method and by the Standardized Exponential (left plot) and respective estimated survival curves (right plot).

The residual analysis was verified according to the graphic technique described in the previous section. It should be noted that both models presented similar results, in addition to not presenting problems of lack of adjustment, since the curvesince do not deviate from the Kaplan-Meier survival curve for the residuals.

Rev. Bras. Biom., Lavras, v.40, n.3, p.327-341, 2022 - doi: 10.28951/bjb.v40i3.574

Table 4 - Estimates of the parameters of frequentist (classic) and Bayesian regression models

Parameter	Categories	M.L.E. (S.E.)	C.I. 95%	$e^{\hat{\beta}}_F$	M_P (S.E.)	C.Ir. 95%	$e^{\hat{\beta}}_B$
β_0	-	13.527 (0.467)	(12.611 ; 14.443)	-	13.596 (0.469)	(12.714 ; 14.547)	-
σ	-	5.345 (0.218)	(4.934 ; 5.790)	-	5.419 (0.078)	(4.155 ; 6.962)	-
β_1: Sex	Male	-	-	-	-	-	-
	Female	0.757 (0.290)	(0.189 ; 1.326)	2.133	0.762 (0.293)	(0.192 ; 1.340)	1.807
β_2: Race	White	-	-	-	-	-	-
	Non-white	-0.589 (0.281)	(-1.141 ; -0.038)	0.554	-0.592 (0.284)	(-1.153 ; -0.038)	0.553
β_3: Education (years)	> 8	-	-	-	-	-	-
	≤ 8	-1.677 (0.307)	(-2.280 ; -1.074)	0.187	-1.691 (0.310)	(-2.312 ; -1.094)	0.184
β_4: Injectable drugs	No	-	-	-	-	-	-
	Yes	-1.701 (0.629)	(-2.932 ; -0.470)	0.182	-1.716 (0.633)	(-2.954 ; -0.466)	0.180

M.L.E.: Maximum Likelihood Estimates; S.E.: Standard error; C.I: Confidence interval; $e^{\hat{\beta}}_F$: Ratio of median times frequentist; M_P:Posterior mean; C.Ir.: Bayesian credibility interval; $e^{\hat{\beta}}_B$: Ratio of median times Bayesian.

According to Equation 2, one can interpret the ratio of median survival times for the Log-Normal model. Both models have remarkably similar interpretations, so that, according to the Bayesian model, the median time to death from HIV/AIDS is approximate:

- 2.1 times higher for females;
- 1.8 times higher for white individuals;
- 5.4 times higher for individuals with more than 8 years of schooling;
- 5.5 times higher for individuals who do not use injecting drugs.

According to the present results, lower education and injecting drug use were associated with lower survival, akin to the results presented by the Cox model in Werle (2021).

The explanatory variable sex is common in Cox model with data from a hospital in the city of João Pessoa (PB), Brazil, discussed by Melo *et al.* (2017), where female sex was related to more prolonged survival, akin with the results presented in this study.

Similarly, compared to the results of the Cox model presented by Müller and Borges (2020), for data from the Region of Campos Gerais (PR), Brazil, the authors argue that females, whites, and higher education levels, were associated with more prolonged survival, aligned with the results found in this study for the parametric models.

Based on the estimated models, it is possible to calculate estimates for the risk and survival of patients notified with HIV in the State of Mato Grosso do Sul according to the explanatory variables.

The development of new probability distributions is of paramount importance in Statistics, as they can promote greater flexibility in applications. Based on the Kumaraswamy distribution presented by Jones (2009), Pascoa *et al.* (2011) introduce and discuss the properties of the Kumaraswamy Generalized Gama distribution. The importance of this probabilistic model, according to the authors, is related to its ability to model monotonous and non-monotonous failure rate functions. Its probability density function can adjust to distorted data that other known distributions cannot adjust due to the flexibility of their parameters, thus enabling their use in various areas (PASCOA *et al.*, 2011). Using a dataset with information on 143 children born at the Hospital das Clínicas of the Faculty of Medicine of Ribeirão Preto, to 1995 from 2001, exposed to HIV through vertical transmission, the researchers use parametric techniques in the frequentist and Bayesian context to demonstrate the efficiency of the model.

Moreover, in Figure 1, it is observed that the behavior of empirical survival estimates indicates the presence of a fraction of long-term survivors (possibly immune to the event of interest), indicating that the inclusion of a cure fraction parameter could improve the quality of the adjustments.

Therefore, we have enough motivation to investigate tsuch a distribution, among others, with a parameter to model the fraction of those cured in further research, considering the HIV/AIDS data presented here.

Conclusions

The Log-Normal regression model was the most parsimonious to adjust the survival data of patients with HIV/AIDS in the State of Mato Grosso do Sul, Brazil, considering frequentist and Bayesian methods, whose parameter estimates were similar.

Based on the interpretations of the coefficients of the parameters of the predictive variables, the need for prevention and early diagnosis policies focused on groups that have a shorter median survival time after notification of HIV infection can be discussed.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

BUENO, M. V.; ROSSI, R. M. Fatores associados ao tempo de sobrevida para pacientes com HIV/AIDS no Estado do Mato Grosso do Sul: abordagem paramétrica. *Braz. J. Biom.,* Lavras, v.40, n.3, p.327-341, 2022.

- *RESUMO: Neste trabalho, objetivou-se ajustar algumas distribuições de probabilidade para o tempo de sobrevida em pacientes com HIV/AIDS no Mato Grosso do Sul, Brasil, acompanhados entre 2009 e 2018, por meio de métodos frequentistas e Bayesianos. A partir de modelos de regressão, pode-se estimar o efeito de variáveis explicativas sobre a variável resposta. Com base nos valores do critério de informação de Akaike (AIC) e o no logaritmo de máxima verossimilhança, foi verificado que a distribuição Log-Normal foi a mais parcimoniosa para os dados. Foram construídos dois modelos de regressão com base nas metodologias descritas, de modo que ambos convergiram para uma mesma interpretação das variáveis explicativas: sexo, raça, escolaridade, e uso de drogas injetáveis. Verificou-se que o tempo mediano até a morte por HIV/AIDS é aproximadamente: 2,1 vezes maior para indivíduos do sexo feminino, 1,8 vezes maior para indivíduos brancos, 5,4 vezes maior para indivíduos com escolaridade superior a 8 anos e 5,5 vezes maior para indivíduos que não fazem uso de drogas injetáveis. A partir das interpretações dos coeficientes dos parâmetros do modelo, pode-se discutir a necessidade de políticas de prevenção e diagnóstico precoce focadas em grupos que apresentam um menor tempo mediano de sobrevida após a notificação da infecção pelo vírus do HIV.*

- *PALAVRAS-CHAVE: Análise de sobrevivência; inferência Bayesiana; modelos probabilísticos; tempos de falha.*

References

BASTOS, J.; ROCHA, C. Análise de sobrevivência: conceitos básicos. *Arquivos de Medicina*, ArquiMed-Departamento de Edições Científicas da AEFMUP, v.20, n.5-6, p.185–187, 2006.

BOX, G. E.; TIAO, G. C. *Bayesian inference in statistical analysis*. [S.l.]: John Wiley, 1973.

BRASIL. *Boletim epidemiológico*: HIV/AIDS. Brasília: Ministério da Saúde, 2020.

BRUNELLO, G. H. V.; NAKANO, E. Y. Inferência bayesiana no modelo weibull discreto em dados com presença de censura. *TEMA*, SciELO Brasil, v.16, n.2, p.97–110, 2015.

CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; BARBOSA, M. T. S.; SHIMAKURA, S. E. *Análise de sobrevivência*: teoria e aplicações em saúde. 2. ed. Rio de Janeiro: Fiocruz, 2011.

COLOSIMO, E.; VIEIRA, A. O modelo de regressao de Cox com covariável dependente o tempo: Uma aplicaçao envolvendo pacientes infectados pelo HIV. *Revista Brasileira de Estatística*, v.54, n.57, p.139–152, 1996.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Blucher, 2006.

COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v.34, n.2, p.187–202, 1972.

COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v.30, n.2, p.248–265, 1968.

HOSMER, D. W.; LEMESHOW, S. *Applied survival analysis*. New York: John Wiley and Sons, 1999.

JACKSON, C. flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, v.70, n.8, p.1–33, 2016.

JONES, M. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, Elsevier, v.6, n.1, p.70–81, 2009.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Taylor & Francis, v.53, n.282, p.457–481, 1958.

LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2003.

LEE, E. T.; WANG, J. *Statistical methods for survival data analysis*. 3. ed. New Jersey: John Wiley & Sons, 2003.

MEDEIROS, A. R. C.; LIMA, R. L. F. C.; MEDEIROS; L. B. D., MORAES; R. M. D.; VIANNA, R. P. D. T. Análise de sobrevida de pessoas vivendo com HIV/AIDS. *Revista de Enfermagem UFPE Online*, v.11, n.1, p.47–56, 2017.

MELO, M. C.; DONALISIO, M. R.; CORDEIRO, R. C. Sobrevida de pacientes com AIDS e coinfecção pelo bacilo da tuberculose nas regiões Sul e Sudeste do Brasil. *Ciência & Saúde Coletiva*, SciELO Public Health, v.22, p.3781–3792, 2017.

MÜLLER, E. V.; BORGES, P. K. O. Sobrevida de pacientes HIV/AIDS em tratamento antirretroviral e fatores associados na Região dos Campos Gerais, Paraná: 2002-2014. *Brazilian Journal of Development*, v.6, n.5, p.28523–28542, 2020.

NETO, L. F. S. P.; PERINI, F. D. B.; ARAGÓN, M. G.; FREITAS, M. A.; MIRANDA, A. E. Protocolo brasileiro para infecções sexualmente transmissíveis 2020: infecção pelo HIV em adolescentes e adultos. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v.30, p.e2020588, 2021.

PASCOA, M. A. D.; ORTEGA, E. M.; CORDEIRO, G. M. The kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, Elsevier, v.8, n.5, p.411–433, 2011.

R CORE TEAM. *R*: a language and environment for statistical computing. Vienna, 2021. Accessible at: http://www.R-project.org¿. Accessed on: Jul. 2021.

RIBEIRO, R. A.; FONSECA, F. F.; PEREIRA, G. F. M. Evolução da AIDS no Brasil: Uma análise espacial. *Revista do Seminário Internacional de Estatística com R*, v.4, n.2, 2019.

ROSSI, R. M. *Introdução aos métodos Bayesianos na análise de dados zootécnicos com uso do WinBUGS e R*. Maringá: Eduem, 2011.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, v.71, p.319–392, 2009.

SANTOS, C. A.; ACHCAR, J. A. A Bayesian analysis in the presence of covariates for multivariate survival data: an example of application. *Revista Colombiana de Estadística*, Universidad Nacional de Colombia., v.34, n.1, p.111–131, 2011.

SANTOS, R. O.; NAKANO, E. Y. Análise do tempo de permanência de trabalhadores no mercado de trabalho do Distrito Federal via modelo de riscos proporcionais de Cox e log-normal. *Revista Brasileira de Biometria*, v.33, n.4, p.570–584, 2015.

SILVA, A. O.; FILHO, A. G. C. G.; SILVA, C. R.; LEITE, D. R. A.; SILVA, L. C. M.; FREITAS, W. W. L. Modelos de sobrevivência aplicados à evasão dos alunos de Estatística da UFPB. *Revista InterScientia*, v.6, n. 2, p.134–145, 2018.

THERNEAU, T. M. *A Package for Survival Analysis in R*. [S.l.], 2020. R package version 3.2-3. Available at: https://CRAN.R-project.org/package=survival. Accessed on: Jul. 2021.

WERLE, J. E. *HIV/AIDS em Mato Grosso do Sul*: análise de tendência, distribuição espacial e sobrevida dos casos, 2021. 75p. Dissertation (Mestrado em Enfermagem) — Universidade Federal do Mato Grosso do Sul, Campo Grande, 2021.

Received on 20.07.2021.

Approved after revised on 22.11.2021.