SURVIVAL ANALYSIS APPLIED TO MEMBER TIME OF RETENTION TO THE HEALTH INSURANCE PROVIDER

Caroline da Silva LIMA[1]
Luiz Carlos SANTOS JÚNIOR[1]
Marcelo Coelho de SÁ[2]

▪ ABSTRACT: Despite the profits earned in the supplementary health segment, the number of policyholders linked to it fluctuated between 2015 and 2018, in contrast to the behavior presented between 2000 and 2014, when it only grew. In order to understand part of this flow, that is, the departure of policyholders from this market, the objective is to analyze member time of retention to the health insurance provider, based on data composed of 122,381 policyholders (and former policyholders) monitored between the years 1984 and 2018. Using the traditional survival analysis, the Kaplan-Meier estimator and parametric and semi-parametric models, the following results stand out: a) the median time of retention in the plan is 4.62 years; b) the mass of insurance provider is composed (over the years observed) predominantly of women, single, young, holders and adherents to the individual / family contract; c) according to the selected Cox model, being a man (in relation to the woman), being young (in relation to the adult), being dependent (in relation to the holder) and being married (in relation to the partner) increase the risk of leaving company analyzed. These results are expected to assist this one in (re)directing its commercial and underwriting policies.

▪ KEYWORDS: Applied survival analysis; health insurance providers; supplementary health.

1 Introduction

The Constitution of the Federative Republic of Brazil, 1988, in its Title VIII – of The Social Order, Chapter II – of Social Welfare, Section II – of Health, specifies the legislation on health and regulates it through its articles (VIOT, 2019). Its art. 196 establishes that health is a right of all and a duty of the State, while art. 197, that it is incumbent upon the Government to provide for their regulation, supervision and control, and they shall be carried out directly or by third parties and also by individuals or private legal entities. As for the participation of private institutions in health care, art. 199 states that health assistance is open to private enterprise. Thus, in short, the Brazilian health system comprises of public (Unified Health System – SUS) and private (Supplementary health) sub-sectors.

[1] Universidade Federal da Paraíba – UFPB, Departamento de Finanças e Contabilidade, Curso de Ciências Atuariais, CEP 58051-900, João Pessoa, PB, Brasil. E-mail: *carolinecslima@gmail.com*, *luiz.atuario@gmail.com*

[2] Universidade Federal de Sergipe- UFS. Departamento de Estatística e Ciências Atuariais, Curso de Ciências Atuariais, CEP 49.100-000, Aracaju, SE, Brasil. E-mail: *mcs.atuarial@gmail.com*.

Supplementary health is supervised by the National Agency for Supplementary Health (ANS), which "promotes the defence of the public interest in supplementary health care, regulates sectoral agencies, including their relations with providers and consumers, and contributes to the development of Health actions in the country" (ANS, 2019a, p. 1). In other words, ANS is the agency that regulates and supervises the supplementary health sector in Brazil, under the Ministry of Health.

Object of the present study, supplementary health in Brazil is regulated, in general, by Act 9,656/98 (BRASIL, 1998), also known as the Health Plans Act, which provided for private health care plans and insurance. According to Viot (2019), the referred act: a) employed the main requirements for the sector and determined standards and rules for its operation; b) determined the control of readjustments and the prohibition of unilateral termination of contracts of individual plans; c) imposed maximum vesting periods and greater rigidity in the care segmentation; d) created age groups for readjustment and prohibited readjustment by age change to those who are over 60 years old; e) prohibited the categorization of risk by age and health condition, and established rules for retirees and dismissed workers.

In order to have a better understanding of the magnitude of this sector, let us highlight its coverage and monetary volume in 2018: it had 47 million beneficiaries (ANS, 2019b) and moved R$ 180 billion, approximately 2.7% of the GDP (FENASAÚDE, 2019).

In 2016, despite the loss of more than 1.5 million customers - due to Brazilian economic stagnation and rising unemployment (ABRAMGE, 2016) -, health insurance providers managed to increase their revenue - profit jumped 66% (compared to the previous year) with consumer readjustments above inflation (GLOBO, 2017).

In a few words, Brazilian studies that investigate the supplementary health sector commonly address the functioning of the sector (OCKÉ-REIS and CARDOSO, 2011), the economic effects caused by regulation changes in the sector (SOUZA, 2014), the identification of health plan user profiles (KANAMURA and VIANA, 2007; ANDRADE, 2014), the determining factors for customer evasion from the aforementioned health insurance providers (HÖRBE, 2012), etc.

In other fields of study, it is common to use survival models for investigating the time before an event of interest happens: the time of municipal civil servant delayed retirement (SANTOS JÚNIOR, 2018), the period of time (survival) a company stays on the market (PIAIA, JACOBI and VENTURINI, 2018), the time before a taxpayer worker retires (PORTILHO, 2013), the time before private bank insolvency (ALVES, 2009), the time before a production failure (BATTISTELLA, 2008), the time before a patient under treatment dies (BARROS and MENEZES, 2008), the time before cancelling the subscription of a magazine (BARROS, 2002), among others.

As stated, in survival analysis the response variable is the time before an event of interest happens, also called failure time, in which failure represents the occurrence of an event of interest. The main feature of survival data is the presence of censoring, which is the partial observation of the response (COLOSIMO and GIOLO, 2006). In this case, it refers to situations in which, for some reason, the follow-up of the subject was interrupted without the occurrence of failure. Without the presence of censoring, classical statistical techniques, such as regression analysis, could be used for analyzing this type of data (COLOSIMO and GIOLO, 2006).

Given the context of coverage reduction and a methodological proposition that is poorly researched and that addresses supplementary health, the objectives of this present research are presented based on data from the insured and former insured parties of a health insurance provider (who prefers to remain anonymous.). In general, we intend to analyze member retention to the health insurance provider, based on traditional survival models, and also to describe the profile of beneficiaries and how some covariates may affect member retention.

2 Theoretical constructs

We hereby discuss the field in which this research is included - the supplementary health sector - by means of presenting data and studies to help with the problem framework - evasion of insured parties from health insurance providers.

2.1 Supplementary health in Brazil

This topic brings an overview of the sector, that is, the establishment of a plan, provider, and portfolio, the types of providers, coverage, contracts designation (old and new), the relation between the signing of new contracts and economic growth, the classification of new contracts, and the distribution of beneficiaries (by gender and age group) that purchase such product.

Act 9,656/98, amended by Provisional Measure 2,177-44/2001 (BRASIL, 2001) came into force with the following amendments:

> Art. 1. The provisions of this act are submitted to the private legal entities that operate health care plans, without failing to comply with the specific legislation that governs their activity, adopting, for the purposes of applying the norms established herein, the following definitions:
> I – Private Health Insurance: continuous provision of services or coverage of care costs at pre or post established price, for an indefinite period, in order to guarantee, without financial limit, health care, by the faculty of access and care by health professionals or services, freely chosen, members or not of the accredited, contracted or referenced network, aimed at medical, hospital and dental care assistance, to be paid in full or in part at the expense of the contracted provider, by reimbursement or direct payment to the provider, by the account and order of the consumer;
> II – Health Insurance Provider: legal entity constituted under the modality of civil or commercial society, cooperative, or self-management entity, which operates product, service, or contract that refers to paragraph I in this article;
> III – Portfolio: the set of contracts for the coverage of health care costs or health care services in any of the modalities covered by paragraph I and § 1 of this article, with all the rights and obligations contained therein.

Thus, it is clear that a provider can market several plans, with different coverages and prices, in order to protect the insured against morbidity and according to their needs and budget constraints. Therefore, the insured parties can make payments, based on a agreement that describes monthly fees, contractual term, and health insurance plan coverage. After signing the contract, the beneficiary will have access to the services offered by the health unit that their health insurance covers.

The private health insurance provider, in other words, is the legal entity constituted under the business modality, association, foundation, cooperative, or self-management entity, compulsorily registered within ANS, which operates or markets private health insurance.

The providers are sorted according to the following modalities: administrator, self-management, health care cooperative, dental cooperative, philanthropy, medical group, dental group practice, or specialized health insurer. The following are the definitions of each of these modalities, according to Viot (2019).

The administrator of benefits is the legal entity that proposes hiring a private health insurance to a given group, to provide as policyholder or offer services to legal entities that have contracted private health insurance to a given group, and who develops activities outlined in specific regulations (RN ANS no. 196, of 07/14/2009).

Self-management is the modality of a provider that operates private health care plans to a closed group of people who must necessarily belong to the same professional position or relate to the company instituting and/or sponsoring and/or maintaining the provider of health insurances.

The health care cooperative is the modality in which a provider is established as a non-profit association formed by doctors operating health insurances, as provided for in Act No. 5,764, of December 16, 1971.

Philanthropy is the type of provider defined as a legal and non-profit organization, a public interest entity, and holder of social assistance charity certificate issued by the competent authorities.

Medical group is a provider incorporated in a company that operates private health care plans, with the exception of the following modalities: administrator, health care cooperative, self-management, philanthropy and specialized health insurer.

The specialized health insurer is an insurer partnership that operates health insurance and has an exclusive corporate purpose for acting in the supplementary health sector, in accordance with Act No. 10,185, of February 12, 2001.

A dental cooperative is the modality of provider that represents a non-profit association, in accordance with Act No. 5,764, of December 16, 1971, comprised by dentists that exclusively operate dental health insurances.

The modality of dental group practice is a provider formed in partnership that exclusively operates dental private health insurance, except for dental cooperatives.

Table 1 shows the modality distribution of active providers in Brazil. These July 2017 data have been given by ANS.

Considering the health care coverage (medical care with or with no dentist service included) of private health insurances in Brazil, we may observe that the Southeast (35.1%) and South (24.8%) regions have the highest coverage rates, followed by the Midwest (21.4%), Northeast (12.2%) and North (10.6%) regions, according to ANS (2019c).

Contracts concluded before the creation of Act no 9,656/98 are designated as old, while contracts concluded after the creation of such act are designated as new, in accordance with the nomenclature used by the ANS. As stated by ANS data, up to December 2018 the number of old contracts have amounted to almost 5 million (8.69%), while the new ones to more than 43 million (91.31%).

The categorization of new contracts according to Pauxis (2015) includes: a) Individual/Family Contracts, which address the natural person, and beneficiaries pay full premium to the health insurance provider; b) Collective Agreements through Membership, in which the professional association or the union contracts the health insurance to beneficiaries; that is, agreements made to the legal entity, but beneficiaries pay full premium to the health insurance provider; c) Collective Corporate Agreements, which are also agreed to the legal entity and the premium is partially or fully paid by the contracting legal entity to the provider. If partially paid, beneficiaries must pay the remaining amount.

According to data provided by ANS (2019c) as of December 2018, Collective Corporate Agreements accumulate the majority of insured parties (66.9%), followed by Individual/Family Contracts (19.2%), Collective Agreement through Membership (13.6%), and not informed (0.3%).

2.2 Member evasion

Health insurance providers, as well as other businesses that offer services and/or products, analyze and use customer retention mechanisms in order to maximize revenues. Barros (2002) states that business profits can be boosted by almost 100% if they are able to keep 5% of their total customers.

In order to achieve the ideal “customer retention”, their profiles must be analyzed (CHIAMULERA, 2017). This can be done by means of dynamic models that are especially employed in the financial and telecommunications areas to estimate the risk of customers stopping the purchase of products (so that measures can be promptly taken to avoid it).

From the perspective of customers evasion, several studies were carried mainly to identify the determining factors in the maintenance or loss of customers. For this purpose, several mathematical-statistical models were applied.

Hohgraefe (2015) estimated the risk of customers evasion at a fleet management business in Brazil using the model of Cox (1972) and considering time-dependent variables based on a counting process. The estimated model selected five significant independent variables (herein the name of variables was not informed due to confidentiality reasons), two of them fixed in time (X2 and X3) and three time-dependent variables (Z2, Z4 and Z5). Traditionally, logistic regression, discriminant, and survival analysis models are used to predict customer evasion.

Barros (2002) has analyzed the risk of customers evasion from a Brazilian company that sells periodical subscriptions that have had the first subscription or any renewal performed from 1994 to 2001, by means of the Cox (1972) proportional risk model. The sample resulting from simple random sampling includes only natural persons and includes 6,034 subscribers. Results found from this sample show that 50% of subscribers cancel their subscription before the 28th month. It has been noted that gender, marital status, age, and most recent amount paid may affect the length of time of subscriptions.

Portilho (2013) has examined the cancelling of supplementary pension insurance contracts. A random sample of 68,968 participants was used for the analysis (70% of which for estimation, and 30% for model validation), considering the survival analysis applied to member retention in a health insurance provider 5 from January 3, 2005 to August 14, 2011. The variables analyzed were marital status, age group, gender, type of insurance, type of payment, payment method, made contribution, contribution range. Among the main results, we found that: married or widowed participants tend to keep the health insurance for longer; people over 60 make up the group with the highest probability of keeping their health insurance, as well as those up to 19 years old; and gender did not show significant differences; the most persistent type of insurance is the Open Benefit Generating Plan (PGBL, Plano Gerador de Benefício Livre); with regard to payment method, the longest subscription period have been found for those who have chosen direct debit; and regarding the range of contribution, the more persistent are the ones who make largest contributions over R$10,000.00.

Chiamulera (2017) chose to estimate Customer Lifetime Value (CLV, Valor Vitalício do Cliente) based on survival analysis. The study used as database a company that sells clothes and shoes to physical stores and e-commerce website (in Brazil). Firstly, grounded on the model of Cox and adjusted to the database, variables associated to customer lifetime were identified. Next, the survival curve of the (11,283,031) bank customers was obtained and the CLV was estimated. A registration database analyzed from July 2010 to September 2017 was used along with a transactional database, which refers to purchases made from January 2014 to September 2017.

From a binary logistic regression, Karam, Silva and Schmidt (2008) related explanatory variables (transactional and demographic) to the probability of the evasion of customer from economic classes A and B, who were subscribers of newspapers in the state of Rio de Janeiro. For this purpose, subscriptions cancelled from June 2004 to May 2005 and active subscriptions in May 2005 were analyzed. With a total of 230,858 active subscriptions and 105,524 cancelled subscriptions, two simple random samples were taken: a training sample with 35,549 cases, and a sample for validation consisting of 4,796 cases. The adjusted model showed that the relevant variables to explain the evasion of newspaper subscribers were: payment method, complaint indicator, type of subscription, source of sale, age group, socio-demographic classification score (SD&W), segmentation score of customers by lifetime value (LTV), region, indicator of participation in loyalty actions, and quantity of aggregate products purchased. After analyzing the variables mentioned, it was concluded that the younger the client the greater the chances of cancellation. For this sample, the gender variable did not affect the cancellation of subscription.

Marín (2005), in turn, methodologically contributed by extending survival models to identify factors that may affect consumer loyalty to insurance providers. By using applications, Marín compared the developed and proposed methodology with that widely used by the market, the Tobit model, and concluded that the direction and significance of the effects did not differ among methods.

Brockett et al. (2008) analyzed the time before the cancellation of insured parties from European insurance policy portfolios. In order to do so, they used households that had multiple insurance policies (of different types) with the same provider as a unit of analysis, to try answering two questions: A) How long after the family cancellation of the first policy does the provider have to retain the client and prevent them from withdrawing from all policies to shift to the competitor service provider? b) What profile features are associated to customer loyalty? Using logistic regression analysis techniques and survival analysis, they found that the desertion time significantly depends on the method used to contact the company, depends on family demographics, and the nature of family insurance policy portfolio. Surprisingly, key customers with three or more policies besides the cancelled one are more vulnerable to total desertion on all policies than non-key customers. In addition, potential effects repellent to customer of premium increases seem to wear off after 12 months.

Huigevoort (2015) used data mining techniques to predict customer turnover at CZ health insurer. To do so, a 2013 data set was used, from four data mining techniques for forecasting modelling in KNIME software: logistic regression, decision tree, neural networks, and support vector machine. Thus, it was found that age, the number of times the client is insured by CZ, and the total consumption in health are the most important aspects to identify dropouts.

This present study has identified a gap related to the evasion of insured parties, and thus it contributes to the acknowledgement and discussion of factors that may favor the withdrawn of insured parties from a health insurance provider in Brazil.

3 Methodological Aspects

The methodological aspects of this study are described in this section. To begin, this is a quantitative, applied, retrospective and descriptive research, developed from a case study.

The population investigated consists of the insured (active and inactive) members of a health insurance provider (who opted for anonymity), defined as a medical cooperative, within the period from 1980 to 2018. Since secondary data were made available on the 122,324 insured members, the sample corresponds to such population.

Herein, the response variable "member retention in the health insurance provider" is investigated considering: data collected from the insurance provider (subsection 3.1), the traditional survival models (subsection 3.2), and the R software, version 3.4.0, with emphasis on the functions in the survival package (THERNEAU, 2015) made available by Moore (2016).

3.1 Data characterization

Only ten (10) variables were provided for each of the 122,324 insured (active and inactive) parties: contract number, user number, insurance number, date of birth, entry date into the insurance, date of evasion from the health insurance provider (when applicable), dependency (ownership), marital status, gender, and type of contract. Independent variables (covariates) shown in Table 1 were selected among these.

Of the covariates presented in Table 1, only the age covariate was not directly available, and it was calculated as follows: minimum between the date of follow-up termination and date of evasion from the insurance provider minus date of birth.

Table 1 – Member data and health insurance provider, 2018

Covariates	Description	Type	Level
Dependency	User dependency code.	Qualitative	Dependent; Holder.
Marital status	User marital status.	Qualitative	Common-law married; Married; Did not respond; Divorced; Single; Widowed.
Age	User age at the date of leaving the plan or at the end of the survey.	Qualitative	Adult: $Idade \geq 30$; Youth: $Idade < 30$.
Gender	User gender.	Qualitative	Female; Male.
Type of contract	Type of link to the user contract.	Qualitative	Collective agreement through membership; Collective corporate agreements; Individual/Family Contracts.

In addition, due to the characteristic of the model of Cox (which displays results as relative risks, that is, it presents whether the risk of the explicit group is higher or lower than that of the hidden group in the model's output), we chose to sort two age groups: youth and adults.

Still considering the visualization of Table 1, it is possible to verify that some variables at first relevant to explain member retention in the health insurance were not given by the insurance provider, such as the price of the service, the coverage, the form, and periodicity of billing, in addition to those indicated by Brockett et al. (2008) and Huigevoort (2015).

It should be noted that no information was also provided about the reasons why insured members had cancelled, such as failure to pay, shift of another provider, shift to another insurance, death, etc., making it impossible to elaborate a more thorough study that would use, for example, competitive risk models (KLEIN *et al*., 2014).

Since a provider offers several insurances (each with its own coverage and price), insured members only have one of the insurances. In this sense, the data provided do not detail whether the evasion corresponds to a shift of insurance (for instance, migrating from an insurance of smaller coverage to another of a greater one), which makes it impossible to use multiple state models (BEYERSMANN *et al.*, 2012).

Thus, the covariates presented, along with the failure indicator (which assumes value 0 in case of censoring and value 1 in case of failure), enable the survival analysis applied to the response variable entitled time before member evasion from the health insurance provider (or member retention in the health insurance provider), defined as minimum between the date of follow-up termination and date of evasion from the insurance provider minus the date of admission to the health insurance.

Thus, the aforementioned failure time is analyzed considering the presence of censoring, of five covariate candidates, and traditional parametric and semi parametric survival models are used to estimate member probability of keeping the health insurance and the respective risk of evasion.

3.2 Survival analysis

According to Colosimo and Giolo (2006), the aspects that rendered the outcome of the survival analysis are: failure times, and very often, censoring.

The time of failure (or the time before an event of interest happens) constitutes the response of the survival analysis. In our case, particularly, it corresponds to member time of retention (or the time before evasion) in a health insurance.

Censoring, which is the partial observation of response - due to some event the follow-up of the insured is interrupted - it is the main aspect of survival data. Censoring times should be used in statistical analysis, because although they are incomplete information, they provide us with lifetime information (COLOSIMO and GIOLO, 2006). Among the different types of existing censoring, the only type present in the current study is the progressive right censored, when the study has a predefined date to end - so individuals who did not observe the event of interest at the end of the follow-up are considered censored (BORGES, 2014). Hence, insured individuals who did not withdraw from the health insurance at the end of the follow-up, that is, in December 2018, were censured on that date.

As explained by Colosimo and Giolo (2006), survival data for the individual i $(i = 1, \dots, n)$ can be represented by the pair (t_i, δ_i), where t_i is the failure time and δ_i the variable failure indicator $(\delta_i = 1)$ or censoring $(\delta_i = 0)$. In the presence of covariates measured in the i-th individual such as, x_i = (dependency, marital status, age, gender, and type of contract), data are represented by $(t_i; \delta_i; x_i)$.

For Colosimo and Giolo (2006), the random and non-negative variable T, which represents the failure time, is commonly specified in survival analysis by the survival function or by the risk function.

As for the survival function $S(t)$, it is defined as the probability of an observation not failing before a certain time t, that is, the probability of an observation surviving through time t. In probabilistic terms, this is written as $S(t) = P(T \geq t)$.

The risk function $\lambda(t)$, in turn, represents the instantaneous failure rate in time t conditional to survival before a given time t, that is, it describes how the instantaneous failure rate changes with time (COLOSIMO and GIOLO, 2006), and is given by the expression.

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq \mathrm{t})}{\Delta t}. \quad (1)$$

Increasing, constant and decreasing risk function represents, respectively, that the risk of insured evasion from the insurance provider increases, maintains, and decreases over time. There is also a simplified representation for the risk function: $\lambda(t) = f(t)/S(t)$, in which $f(t) = [F(t)]'$, that is, $f(t)$ is the probability density function of the random variable T. The cumulative distribution function, i.e., $F(t) = 1 - S(t) = P(T \leq t)$, defines the probability of an observation not surviving over time t (CARVALHO *et al*., 2011).

Given this introduction to the Survival Analysis, we present briefly: A) The Kaplan-Meier estimator (KAPLAN and MEIER, 1958), a nonparametric estimator for the survival function; B) the risk functions according to the exponential, Weibull and Log-normal probabilistic models (COLOSIMO and GIOLO, 2006); c) and the Cox regression model (COX, 1972), a semi-parametric estimator for the risk function. In items (B) and (c), survival functions are calculated from risk functions.

As specified by Colosimo and Giolo (2006) and Carvalho *et al*. (2011), the nonparametric Kaplan-Meier estimator (KM) of the survival function is defined as

$$\hat{S}(t) = \prod_{j:t_j<t} \left(\frac{n_j - d_j}{n_j}\right) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right). \tag{2}$$

To present its general expression, the following are considered: $t_1 < t_2 <\ldots< t_k$, the k distinct and ordered failure times; d_j the number of failures in t_j, $j = 1, \ldots, k$, and; n_j the number of individuals at risk t_j, that is, individuals who did not fail and were not censored before the moment immediately preceding t_j.

The main properties, the asymptotic variance, as well as the confidence interval for the KM estimator of the survival function, are presented by Colosimo and Giolo (2006).

The logrank test was developed to compare KM estimates for the survival function among groups (or levels) of covariates. This information helps us understand whether there is some difference between the groups with regard to the probability of retaining members in the health insurance plan. If the risk of evasion is greater for men, for instance, the insurance provider can concentrate commercial efforts that stimulate the retaining of male members. This is an approximate test for the equality of two survival functions based on statistics expressed by (3), according to Colosimo and Giolo (2006),

$$T = \frac{\left[\sum_{j=1}^{k}\left(d_{2j} - w_{2j}\right)\right]^2}{\sum_{j=1}^{k}\left(V_j\right)_2}, \tag{3}$$

where d_{2j} is the number of deaths in the second group by cause j, which follows hypergeometric distribution, in a contingency table; w_{2j} is the mean of d_{2j}; $\left(V_j\right)_2$ is the variance of d_{2j}; T, under the null hypothesis $H_0: S_1(t) = S_2(t)$, it has a chi-square distribution with 1 degree of freedom for large samples.

Regarding the probabilistic models, the risk functions of some of the main probability distributions used in survival analysis are presented in accordance with Colosimo and Giolo (2006) and Carvalho *et al*. (2011): the equation (4) denotes the exponential distribution (λ), equation (5), Weibull (λ, γ) and equation (6), log-normal (μ, σ):

$$\lambda(t) = \frac{1}{\alpha}, \tag{4}$$

$$\lambda(t) = \frac{\gamma}{\alpha^{\gamma}} t^{\gamma-1}, \tag{5}$$

$$\lambda(t) = \frac{\frac{1}{\sqrt{2\pi}\sigma t}\exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right]}{\Phi\left(\frac{-\log(t)+\mu}{\sigma}\right)}. \quad (6)$$

where α and σ are scale parameters; γ shape parameter; μ location parameter (SILVA, 2015).

According to Colosimo and Giolo (2006) and Carvalho *et al.* (2011), the Cox regression model allows data analysis from lifetime studies in which the response is the time before the occurrence of an event of interest, by adjusting the covariates.

Considering p covariates so that x is a vector with the components $x = (x_1, \ldots, x_p)'$, the expression of the Cox regression model considers

$$\lambda(t) = \lambda_0(t) g(x'\beta), \quad (7)$$

where g is a function that must be specified, so that $g(0) = 1$. This model consists of the product of two components: a nonparametric and a parametric. The nonparametric component, $\lambda_0(t)$, is not specified and is a non-negative function of time. It is usually called base function, since $\lambda(t) = \lambda_0(t)$ when $x = 0$. The parametric component is often used in the multiplicative form exposed in

$$g(x'\beta) = \exp(x'\beta) = \exp(\beta_1 x_1 + \cdots + \beta_p x_p), \quad (8)$$

where β is the vector of parameters associated with covariates. The basic assumption to use the Cox regression model (or proportional risk model) is that the failure rates are proportional, i.e. the ratio of the failure rates of two different individuals is constant over time.

In this sense, the Schoenfeld residual test (Schoenfeld, 1982) verifies the proportionality assumption required in the Cox model. To define such residues, consider that there are $k \geq n$ distinct failure times $t_1 < t_2 < \ldots < t_k$. If the individual i with covariate vector $x_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ fails, one has for this individual a Schoenfeld residue vector $r_i = (r_{1i}, r_{2i}, \ldots, r_{pi})$ in which each component r_{qi} $(q = 1, \ldots, p)$ is defined by

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jq} \exp(x_j'\hat{\beta})}{\sum_{j \in R(t_i)} \exp(x_j'\hat{\beta})}. \quad (9)$$

For each of the p covariates considered in the model, there is a corresponding Schoenfeld residue for each individual i. Since the residuals are defined on each failure, the Schoenfeld residuals set is thus an array with k rows and p columns. Each row corresponds to a distinct failure time and each column to one of the p covariates considered in the model.

In line with Colosimo and Giolo (2006), the Cox regression model is widely used and the main reason for its popularity is the presence of the nonparametric component, which makes the model quite flexible.

The selection of parametric and semiparametric models is performed here by the Mean Absolute Percentage Error (MAPE), a measure that expresses the accuracy of the error in percentage. To calculate the MAPE we use equation (10) (PORTAL ACTION, sd).

$$MAPE = \frac{\sum_{t=1}^{n} |(y_t - \hat{y}_t)/y_t|}{n} \times 100, \qquad se\ y_t \neq 0. \tag{10}$$

y_t corresponds to the observed data, $\hat{y}_t$ to the adjusted data and n is the number of observations. The smaller the MAPE, the better our fit.

4 Results

This section presents the descriptive and modelling results in survival analysis. We used herein the database of a Brazilian health insurance provider, defined as a medical cooperative. The database consists of 122,324 active and inactive insured individuals (with censoring of 22.75%), assisted from 1984 to 2018.

The variables for analysis were collected from the user files. Of these, five (5) covariates (qualitative explanatory variables) were selected, and are shown in Table 1. They can have a significant effect on the evasion of the insured individuals from the insurance provider: dependency, marital status, age, gender, and type of contract.

In addition to these, there are the fundamental variables, those that characterize the survival models: the failure indicator (insured individual status), which states if the insured has remained in (0) or evaded from (1) the health insurance; and the variable response defined as member retention in the health insurance, which is determined by the difference between the age at the end of follow-up (or the age at the date of evasion) and the age at the time of subscription to the health insurance.

4.1 Descriptive analysis

Due to the existence of right censoring, we choose to present the median, and not the mean (which would probably be underestimated), as indicated by Colosimo and Giolo (2006), the member retention (in years) of the insured in the health insurance: 4.62 years. It represents a suitable central trend statistic as long as the number of censured data does not exceed that of failures (the censoring rate in this case is 22.75%).

When viewing Figure 1, it is possible to note that the distribution of member retention in health insurance is shifted to the left, thus, it registers a higher frequency for shorter retention periods. This indicates that one is more likely to withdraw from the health insurance plan within little time after enrolling. As put further, young and single male, financially dependent, and with collective agreements through membership contribute to the reduction of such time.

According to the objectives of this study and for exploratory purposes, Table 2 shows the number of people (in absolute and relative terms) by Type of Contract.

With respect to gender, the largest proportion of insured parties belong to the group of women contracting the individual/family modality (29.73%), while the smallest proportion corresponds to men contracting the collective agreement through membership (2.94%).

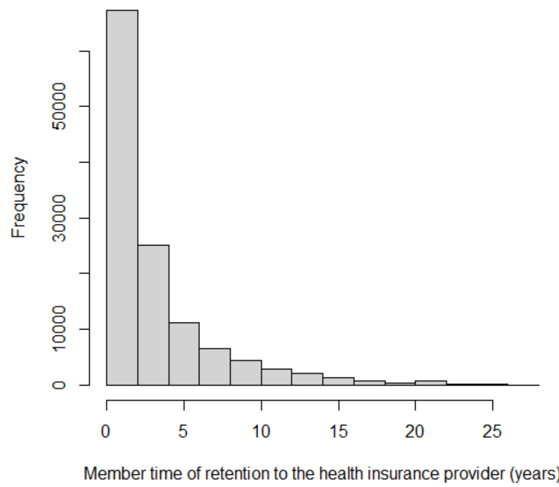

Figure 1 - Distribution of the member time of retention to the health insurance provider, 1984-2018.

As for age (categorized), the largest proportion of insured belong to the group of young members of individual/family contract modality (35.08%), while the smallest proportion corresponds to adult members of the collective agreement through membership (2.54%).

In relation to marital status, the largest portion of insured belongs to the group of single contractors of the individual family modality (36.12%), while the smallest share corresponds to common-law marriage contractors of the collective agreement through membership (0.01%).

As for dependency, the largest proportion of policyholders belong to the group of individual/family contract holders (42.69%), while the smallest proportion corresponds to collective agreement through membership (1.23%).

Thus, Table 2 shows that the individual / family modality, for any covariate considered, has higher prevalence, while the collective agreement through membership has the lowest. This ratio varies according to the type of provider and according to the provider's commercial and risk underwriting policy. A self-management provider, for example, only markets enterprise agreements, while a cooperative can market all three types.

Another factor that helps explain the low proportion of collective agreements through membership in relation to the total of contracts is that, as presented in subsection 2.1, it represents the least acquired contract in Brazil, because it requires affiliation. Besides, there is the fact that in the cooperative that we analyzed, collective agreements through membership are relatively new, with a maximum of 5 years of operation.

Table 2 - Number of policyholders by Type of Contract x Covariates, 2018

Variable	Level	Collective Agreements through Membership	Collective Corporate Agreements	Individual / Family Contracts
Dependency	Dependent	1505 (1,23%)	21971 (17,96%)	10856 (8,87%)
	Holder	7010 (5,73%)	28757 (23,51%)	52225 (42,69%)
Marital status	Common-law married	8 (0,01%)	248 (0,20%)	41 (0,03%)
	Married	1345 (1,10%)	18389 (15,03%)	15903 (13,00%)
	Did not respond	1526 (1,25%)	1710 (1,40%)	860 (0,70%)
	Divorced	59 (0,05%)	414 (0,34%)	989 (0,81%)
	Single	5565 (4,55%)	29764 (24,33%)	44186 (36,12%)
	Widowed	12 (0,01%)	203 (0,17%)	1102 (0,90%)
Age	Adult	3113 (2,54%)	24224 (19,80%)	20174 (16,49%)
	Young	5402 (4,42%)	26504 (21,67%)	42907 (35,08%)
Gender	Female	4918 (4,02%)	24848 (20,31%)	36369 (29,73%)
	Male	3597 (2,94%)	25880 (21,16%)	26712 (21,84%)

4.2 Analysis of member retention time in health insurance

After this brief description of the data, we proceed with the main analysis of this work, that is, the survival analysis applied to the issue of member retention in health insurance. Our results are divided into three parts: a) nonparametric analysis-derived from the Kaplan-Meier estimator (1958) applying logrank test; b) parametric analysis - derived from the exponential, Weibull and Log-normal models; c) semi-parametric analysis - derived from Cox model (1972) and Schoenfeld test.

Initially, Figure 2 shows the estimation of the survival function for the group analyzed, by means of the Kaplan-Meier estimator.

Figure 2 confirms low member retention in the health insurance provider included in the study. For example, we have $\hat{S}(4{,}61) = 28{,}08\%$, that is, the probability of a member keeping the health insurance plan within the median time is less than 30%.

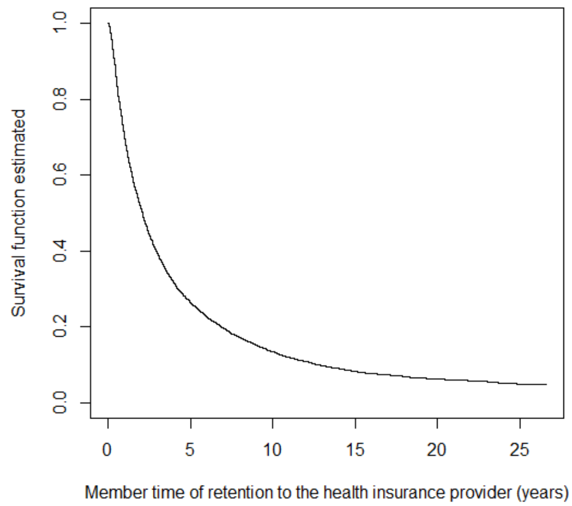

Figure 2 - Number of policyholders by Type of Contract x Survival curves estimated by the Kaplan-Meier estimator, referring to member time of retention to the health insurance provider, 1984-2018.

Figure 3 informs that the policyholder member retention is higher than that of the dependent for any time – after all, there can only be a dependent if there is a policyholder. This difference increases over time and stabilizes in $t = 7$ years. At such time, the probability that the policyholder will continue in the health insurance plan is greater than 20%, and that of the dependent is less than 20%.

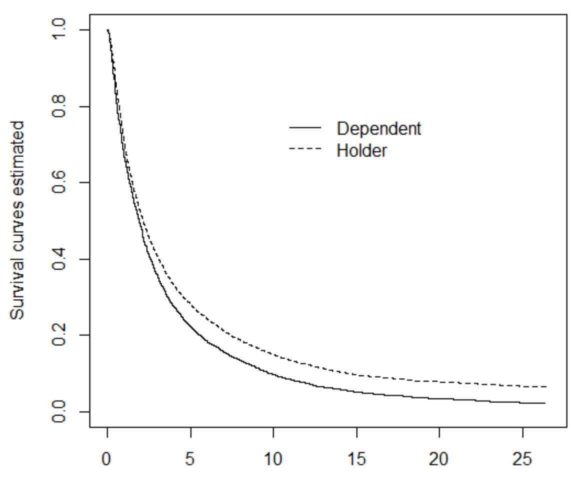

Figure 3 - Survival curves estimated by the Kaplan-Meier estimator, by levels of dependence, referring to member time of retention to the health insurance provider, 1984-2018.

Figure 4 shows that the probability of an adult continuing in the plan is higher than that of a youth at any time. That happens because adults have better cost-benefit outcomes compared to the youth, since younger enrolees demand less of plan and have lower wages.

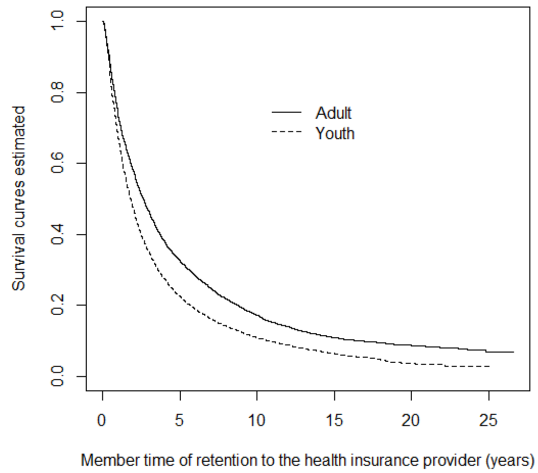

Figure 4 - Survival curves estimated by the Kaplan-Meier estimator, by levels of age, referring to member time of retention to the health insurance provider, 1984-2018.

Figure 5 shows that the probability of women continuing their enrolment is slightly higher than that of men, at any time, which makes sense, since it is empirically known that women, in general, uses more medical services than men.

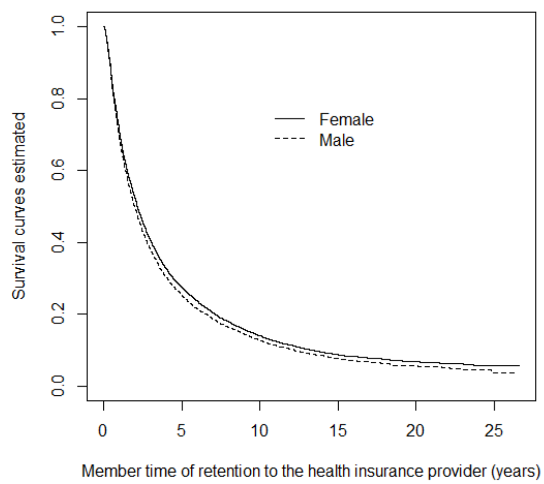

Figure 5 - Survival curves estimated by the Kaplan-Meier estimator, by levels of gender, referring to member time of retention to the health insurance provider, 1984-2018.

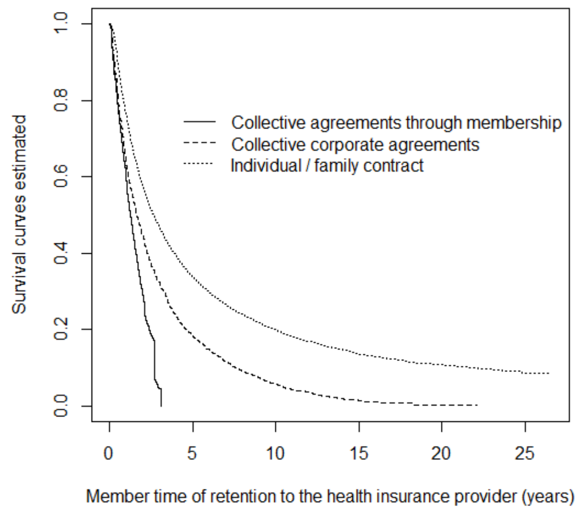

Figure 6 - Survival curves estimated by the Kaplan-Meier estimator, by contract type levels, referring to member time of retention to the health insurance provider, 1984-2018.

As for the type of contract, shown in Figure 6, the higher probability of member maintenance by enrolees can be due to: a) the commercial policy of the provider, b) the free membership subscription – it is only necessary for an individual to look for the provider and pay the insurance, C) in addition to the termination of the contract, which occurs only in case of fraud and/or failure to pay. The collective agreement through membership, in turn, presents the greatest risk of member withdraw, since its membership requires a link with a professional association or union and this relationship can be terminated at any time.

According to Figure 7, concerning marital status, the "did not respond" group presented, for the first years, a higher risk of cancelling their health insurance plan, while widows showed the lowest risk. It seems reasonable to believe that the common-law married, the married and the divorced members have higher expenses that compete with the monthly contribution to the health insurance. Accordingly, single and widowed individuals, with generally more individualized expenses (therefore reduced), would have more resources to remain on the health insurance plan.

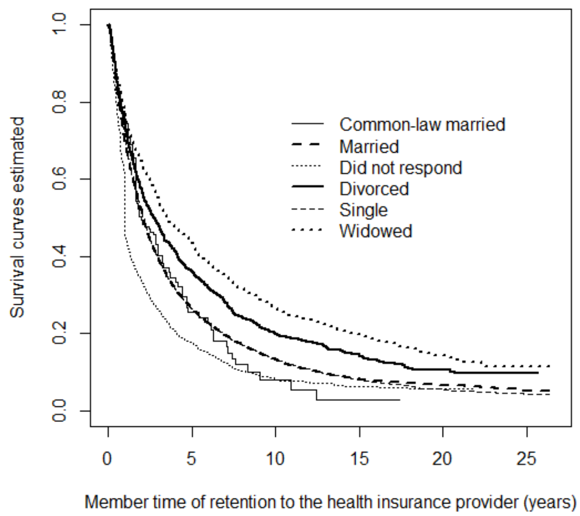

Figure 7 - Survival curves estimated by the Kaplan-Meier estimator, by levels of marital status, referring to member time of retention to the health insurance provider, 1984-2018.

Thus, in short: a) member retention to the provider, in general, is relatively low; b) regarding gender, there seems to be no difference in the first years of enrolment in the health insurance plan, with a change in behavior over time - the probability of female member retention seems to be a little bit higher; c) regarding dependency, the estimated survival for the policyholder is, in general, superior to that of the dependent enrolee d) as for marital status, cohabiting and separated members have higher probability of member retention; e) with regard to the type of contract, we note that the individual/family agreement is the type of contract with the highest survival, while the collective agreement through membership has the lowest survival; in relation to the categorized age, adults have shown a longer enrolment period of time compared to the youth. It is important to mention that all the evidence from the graphic needs to be confirmed through a logrank test.

After the graphical demonstration of the estimated survival functions, it is possible to observe through the logrank test whether there is (statistical) equality of the applied curves estimated by covariate levels (if the survival of men and women is equal, if the survival of policyholders and dependents is equal, etc.). For this test, the null hypothesis is that the survival estimates are equal, while the decision rule is: when the p-value is less than or equal to 5%, the null hypothesis is rejected, otherwise it is not rejected. The analysis of data has shown that the p-values for the five covariates (dependency, age, marital status, gender, and type of contract) are below 5%. Therefore, null hypotheses are rejected, i.e. there is evidence to reject the hypothesis that the estimated survivals of the levels are equal.

In addition, in order to further explore the data, the mean residual life on time t ($vmr(t)$) was calculated, as expressed in (11). For insured on time t, this amount measures the mean time remaining in the plan.

$$vm(t) = \frac{\int_t^{\infty} S(u)du}{S(t)} \qquad (11)$$

Illustratively, we have that: a) the remaining time of enrolment for someone who has been a member for 1 year is, on average, equal to 6.13 years; b) for those who have been a member for 13 years, it is equal to 8.57 years; c) for those who have been member for 25 years, it is equal to 1 year. Within the first 10 years, the longer the member retention time the greater the average remaining time as a member, which may be associated with advancing age and consequent greater usage of the services provided by the health insurance, in addition to raises on the wages of the insured individual over time, which would be in accordance with Brocket et al. (2008) – "the potential customer-repelling effects of increasing the premiums appear to wear out after 12 months." For memberships longer than 11 years, the membership mean remaining time begins to decrease, probably due to the death of the insured.

After the nonparametric analysis, the results of the parametric analysis are presented, explaining the failure risk due to some candidates to independent variables. Thus, the risk of evasion from the health insurance was reverted several times, due to the ensemble of the covariates dependency, marital status, age, gender, type of contract, and under the perspective of three probabilistic models: Exponential, Weibull, and Log-normal.

Thereby, the adjustments with the estimated survival curves most similar to that observed empirically via Kaplan-Meier were selected. In this sense, the covariates ensemble that best aligned to the data in accordance to the MAPE, for the three models, was comprised of dependency, marital status, age and gender, thus excluding type of contract, which in addition to worsening the adjustment, considerably increases computational cost. Such adjustments are shown in Figure 8.

When analysing Figure 8, it is possible to note that the best adjustment was that of the Log-normal model.

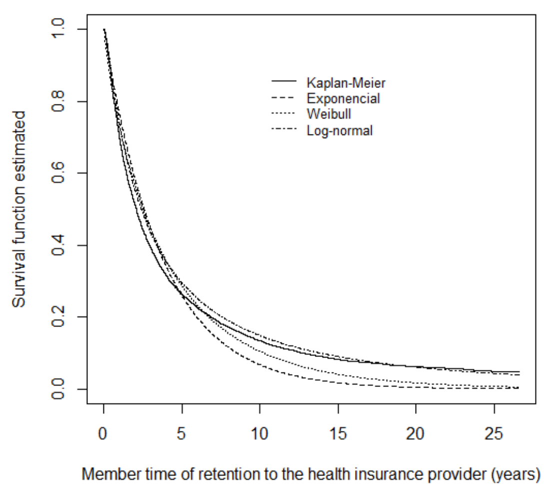

Figure 8 - Survival curves estimated by parametric models, referring to member time of retention to the health insurance provider, 1984-2018.

Once the parametric model is chosen, the results of the semi-parametric analysis are presented, and that also explains the failure risk due to candidates for independent variables.

Similarly, the risk of evasion from the health insurance was reduced due to the ensemble of the covariates dependency, marital status, age, gender, type of contract. Thus, we select the Cox adjustment that displays the estimated survival curve closest to that empirically observed via Kaplan-Meier. Such adjustments are shown in Figure 9.

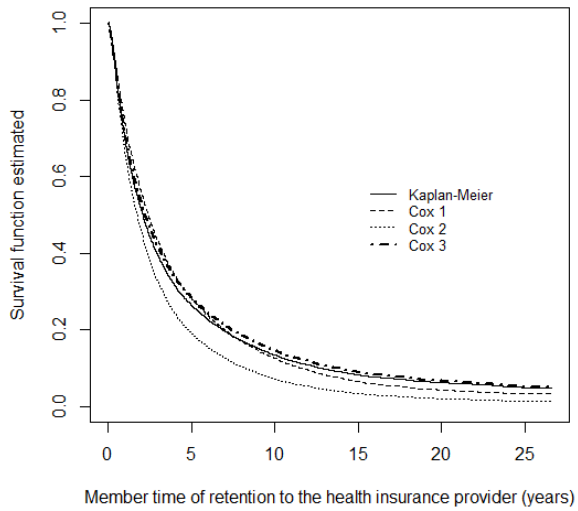

Figure 9 - Survival curves estimated by Cox semiparametric models, referring to member time of retention to the health insurance provider, 1984-2018.

Thus, an overview of Figure 9 allows us to conclude that for three different combinations of covariates, the best adjustment was provided by dependency, marital status, age and gender, thus excluding type of contract.

Subsequently, we chose to compare, according to Figure 10, the pre-selected parametric and semi-parametric settings for the choice of the model considered, among those adjusted in this study, the most appropriate to analyze member retention in the health insurance provider under analysis.

In accordance to Figure 10, it can be observed that the adjustments capture well the level and shape of empirical survival, with the Cox model presenting a slight advantage, observed both graphically and through the Mean Absolute Percentage Error (MAPE): log-normal and Cox show, respectively, 8.36% and 7.22%.

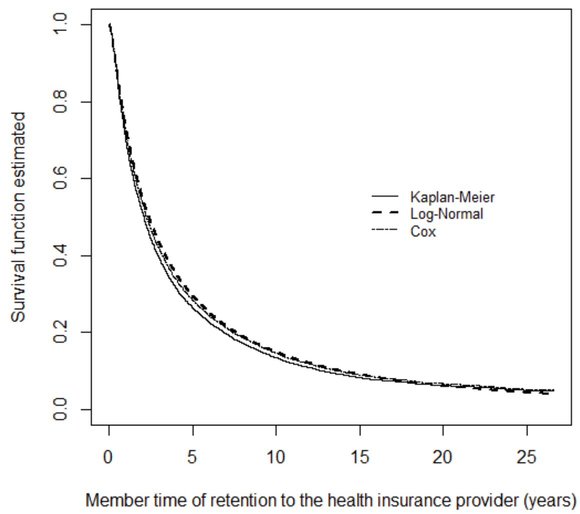

Figure 10 - Survival curves estimated by the Log-normal and Cox models, referring to member time of retention to the health insurance provider, 1984-2018.

After electing the Cox model, Table 3 presents the estimated effects (e^{β}) that the covariates exert, by level, on the risk of evasion from the health insurance. Effects greater than 1 point that such level, relative to the level omitted, presents higher evasion risks, while effects less than 1, lower risks. The asterisk indicates that the effect is relevant at the significance level of 5%.

Table 3 - Summary of Cox model fitting for member time of retention to the health insurance provider, 2018

Variable - level	e^{β}
Dependency - holder	0,890488*
Marital status - married	1,047463
Marital status - did not respond	1,494600*
Marital status – divorced	0,894566
Marital status - single	0,778608*
Marital status - widowed	0,766061*
Age - young	1,537879*
Gender - male	1,061594*

According to the results presented in Table 3, a) being a policyholder reduces the risk of evasion from the insurance (by 10.96%), when compared to being a dependent, B) being single and widowed reduces the risk of evasion, when compared to the "common-law marriage" category by 22.14 and 23.40%, respectively, C) being male, compared to female, increases the risk of evasion by 6.15%, d) being young, compared to adult, increases the risk of evasion by 53.78%.

Table 4 shows the overall and covariate p-values for the Schoenfeld test. As these values are lower than the significance level of 5%, the hypothesis of proportionality of risks required by the model is rejected, despite the good adjustment provided, little overrated.

Table 4 - Schoenfeld test for proportionality of risks of the Cox model fitted to the member time of retention to the health insurance provider, 2018

Variable	χ^2	P-value
Dependency	18.0	<0.001
Marital status	78.8	<0.001
Age	6.2	0.013
Gender	24.8	<0.001
Global	129.2	<0.001

Despite not considering this assumption, according to Hjort (1992), the estimator derived from a poorly specified model is consistent, although not for the parameter of the poorly specified model, but for the so-called less false parameter. The less false parameter is "less false" in the sense that it produces the best approximation of the incorrectly specified model relative to the true model that generated the data. For this reason, we opt to use this model, given its simplicity during the modelling phase, as well as its interpretation.

Thus, we complete the results section highlighting that the selected model, despite not obeying the premise of proportionality of risks, the model fitted well the data, according to graphical analysis, and identified some risk factors related to the evasion of the insured from the health insurance provider analyzed: male, young, common-law marriage, and dependent.

Conclusions

The relevance of customer evasion is evident when we see on the news that enrolees of health insurance plans have been withdrawing due to the Brazilian economic recession. Aware of the possible impact of this variation on the solvency of the institution, we chose, in this study, to address member retention in a health insurance, from data given by a provider who chose to remain anonymous.

Our exploratory analysis allowed to identify that most of the insured individuals are women, policyholders, young, single women, and member of the individual / family contract. It is noteworthy that member profile may be influenced by the type of provider, the cooperative, and the commercial and risk underwriting policy adopted by the provider.

The use of survival models was driven by a gap identified in studies that address customer evasion from health insurance providers in Brazil. Logistic regression models are commonly used in the evaluation of customer evasion in other areas. This "new feature" allowed to consider partial information (censoring) about the insured individuals, which is positive in terms of sample size. Through it, it was possible to estimate the survival function via Kaplan-Meier and compare the estimation of the levels of covariates.

In addition, the risk of evasion from the health insurance was estimated through parametric and semi-parametric regressions, considering the effects of four qualitative variables, in which all had significant effects over the aforesaid risk: two will increase the risk of evasion (male and youth) and two will reduce the risk of evasion (being a policyholder and single/widowed).

Limitations include:

- No Brazilian studies about member retention in health insurance were found, nor studies using survival models, which hindered the comparison of results.
- Despite the availability of data, it can be observed that it is difficult to access them: companies often impose limitations in relation to the use of important variables, such as the income of the insured, the price of the product, the method and periodicity of billing, etc.
- The lack of these variables probably worsens the quality of the adjustment and impairs to ascertain some research hypotheses.

In this context, we suggest the elaboration of a bibliometric study, focusing on the International review, in addition to the consideration of data from other health insurance providers, other variables that may explain the evasion from the health insurance, as well as other survival models, whether of competitive risks, multiple states, cure fraction or fragility, depending on the issue and the sample.

Finally, despite the limitations we have mentioned, this current research should contribute to the improvement of studies in supplementary health, given its scarcity and importance for the Brazilian people well-being and economy. Furthermore, we hope this application will be able to guide the risk mapping of health care providers, and also that these results help the provider we have analyzed to (re)direct its commercial and risk underwriting policies.

Acknowledgements

We would like to thank editors and reviewers for their comments and suggestions.

LIMA, C. S., SANTOS JÚNIOR, L. C., SÁ, M. C. Análise de sobrevivência aplicada ao tempo de permanência em planos de saúde. *Rev. Bras. Biom*. Lavras, v.39, n.3, p.372-398, 2021.

- *RESUMO*: *Apesar dos lucros auferidos no segmento de saúde suplementar, o número de segurados vinculados a ele oscilou entre 2015 e 2018, em contraste com o comportamento apresentado entre 2000 e 2014, quando só cresceu. Para compreender parte desse fluxo, isto é, a saída dos segurados desse mercado, objetiva-se analisar o tempo de permanência do segurado em planos de saúde, a partir de dados compostos por 122.381 segurados (e ex-segurados) acompanhados entre os anos de 1984 e 2018. Utilizando-se da análise de sobrevivência tradicional, por meio do estimador de Kaplan-Meier e de modelos paramétricos e semiparamétricos, destacam-se os seguintes resultados: a) a mediana do tempo de permanência no plano é de 4,62 anos; b) a massa de seguradoras é composta (ao longo dos anos observados) predominantemente por mulheres, solteiras, jovens, titulares e aderentes ao contrato de individual/familiar; c) conforme o modelo de Cox selecionado, ser homem (em relação à mulher), ser jovem (em relação ao adulto), ser dependente (em relação ao titular) e ser casado (em relação ao amasiado) aumentam o risco de saída da operadora analisada. Espera-se que esses resultados auxiliem a operadora analisada a (re)direcionar suas políticas comerciais e de subscrição de riscos.*

- *PALAVRAS-CHAVE: Saúde suplementar; operadora de planos de saúde; análise de sobrevivência aplicada.*

References

ABRAMGE - ASSOCIAÇÃO BRASILEIRA DE PLANOS DE SAÚDE. Cenário Saúde. *ABRAMGE*, year 2, n.6, 2016. Available at: https://www.abramge.com.br/portal/files/cenario-saude/cenario_da_saude_ed6.pdf (Accessed: 23 March 2019).

ALVES, K. L. F. *Análise de sobrevivência de bancos privados no Brasil*. 2009. 83f. Dissertação (Mestrado em Engenharia de Produção) - Universidade de São Paulo, São Carlos, 2009.

ANDRADE, V. D. *A importância dos planos de saúde conhecerem o perfil dos usuários do plano empresarial.* 2014. 76f. TCC (Graduação em Serviço Social) - Universidade Federal do Rio Grande do Norte, Natal, 2014.

ANS – Agência Nacional de Saúde Suplementar. *Quem somos*. 2019a. Available at: http://www.ans.gov.br/aans/quem-somos (Accessed: 07 April 2019).

ANS – Agência Nacional de Saúde Suplementar. *ANS TABNET: Informações em saúde suplementar*. 2019b. Beneficiários por UFs, Regiões Metropolitanas (RM) e Capitais. Available at: http://www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_br.def (Accessed: 02 May 2019).

ANS – Agência Nacional de Saúde Suplementar. *Caderno de informação da saúde suplementar: beneficiários, operadoras e planos*. 2019c. Available at: http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Dados_e_indicadores_do_setor/total-cad-info-jun-2019.pdf (Accessed: 07 April 2019).

BARROS, J. E. *Análise de sobrevivência: modelo de risco de desligamento de clientes*. 2002. 124f. Dissertação (Mestrado em Administração) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002.

BARROS, V. F A.; MENEZES, J. E. Aplicação da teoria da análise de sobrevivência no cálculo do risco de morte por infecção hospitalar. In: SIMPÓSIO EDUCAÇÃO TECNOLOGIA E SOCIEDADE, 1, 2008, Inhumas. *Anais*... CEFET-GO. 2008, p. 81-85. Available at: https://www.researchgate.net/publication/307645288_APLICACAO_DA_TEORIA_DA_ANALISE_DE_SOBREVIVENCIA_NO_CALCULO_DO_RISCO_DE_MORTE_POR_INFECCAO_HOSPITALAR (Accessed: 07 April 2019).

BATTISTELLA, P. M. D. *Análise de sobrevivência aplicada à estimativa da vida de prateleira de salsicha*. 2008. 115f. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Santa Catarina, Florianópolis, 2008.

BEYERSMANN, J.; ALLIGNOL, A.; SCHUMACHER, M. *Competing Risks and Multistate Models with R*. New York: Springer, 2012. 245p.

BORGES, Alexandra Isabel Monteiro. *Análise de sobrevivência com o R*. 2014. 63f. Dissertação (Mestrado em Matemática) - Universidade da Madeira, Funchal, Portugal, 2014. Available at: http://hdl.handle.net/10400.13/732 (Accessed: 09 June 2019).

BRASIL. Lei nº 9.656, de 3 de junho de 1998. Dispõe sobre os planos e seguros privados de assistência à saúde. *Diário Oficial da União*, Brasília, DF, 4 jun. 1998.

BRASIL. Medida provisória nº 2.177-44, de 24 de agosto de 2001. Altera a Lei no. 9.656, de 3 de junho de 1998, que dispõe sobre os planos privados de assistência à saúde e dá outras providências. *Diário Oficial da União*, Brasília, DF, 27 ago. 2001.

BROCKETT, P. L. *et al.* Survival Analysis of a Household Portfolio of Insurance Policies: How Much Time Do You Have to Stop Total Customer Defection? *The Journal of Risk and Insurance*, v.75, n.3, p.713-737, 2008.

CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; BARBOSA, M. T. S; SHIKAMURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. 2.ed. Rio de Janeiro: Fiocruz, 2011. 434p.

CHIAMULERA, M. L. *Modelagem do valor vitalício do cliente via abordagem de análise de sobrevivência*. 2017. 46f. TCC (Graduação em Estatística) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Editora Edgard Blücher, 2006. 370p.

COX, D. R. Regression models and life tables. *Journal of the Royal Statistical Society*, v.34, n.2, p.187-220, 1972.

FENASAÚDE - Federação Nacional de Saúde Suplementar. *Entenda o setor*. Available at: http://fenasaude.org.br/conheca-a-fenasaude/sobre-o-setor/entenda-o-setor.html (Accessed: 27 Frebuary 2019).

GLOBO. *Quantidade de segurados de planos de saúde diminuiu no Brasil em 2015*. 2017. Available at: http://g1.globo.com/jornal-nacional/noticia/2015/10/quantidade-de-segurados-de-planos-de-saude-diminuiu-no-brasil-em-2015.html (Accessed: 20 March 2019).

HJORT, N. L. On inference in parametric survival data models. *International Statistical Review*, v.60, p.355-387, 1992.

HOHGRAEFE, M. W. *Estimação do risco de cancelamento de clientes no setor de gerenciamento de frotas por meio de modelo de Cox com variáveis tempo-dependentes*. 2015. 41f. TCC (Graduação em Estatística) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

HÖRBE, C. V. *Comportamento do consumidor após falhas em serviços: uma pesquisa com usuários de planos de saúde*. 2012. 116f. Dissertação (Mestrado em Administração) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2012.

HUIGEVOORT, C. W. J. M. *Customer churn prediction for an insurance company*. 2015. 84f. Dissertação (Mestrado em Ciências) – Eindhoven University of Technology, Eindhoven, 2015.

KANAMURA, A. H.; VIANA, A. L. D. Gastos elevados em plano privado de saúde: com quem e em quê. *Revista de Saúde Pública*, v.5, n.41, p.815-820, jun. 2007.

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

KARAM, K. A.; SILVA, J. F.; SCHMIDT, F. H. *Regressão logística*: um modelo de risco de cancelamento de clientes. 2008. Available at: http://www.anpad.org.br/diversos/down_zips/38/MKT-B672.pdf (Accessed: 06 April 2020).

KLEIN, P. J. et al. *Handbook of Survival Analysis*. Boca Raton: CRC Press, 2014. 632p.

MARÍN, A. M. P. Survival methods for the analysis of customer lifetime duration in insurance. 2005. 149f. Tese (Doctorado em Estudios Empresariales) – Universidad de Barcelona, Barcelona, 2005.

MOORE, D. F. *Applied Survival Analysis Using R*. Piscataway: Springer International Publishing, 2016. 226p.

OCKÉ-REIS, C. O.; CARDOSO, S.S. A regulamentação dos preços dos planos individuais de saúde. *Revista de Economia Política*, v.31, n.3, p.455-470, jun. 2011.

PIAIA, R.; JACOBI, L. F.; VENTURINI, J. C. Estudo da sobrevida de empresas de Santa Maria - RS no período de agosto de 2008 a agosto de 2013. *Revista Thema*, v.15, n.4, p.1200-1216, 2018.

PORTAL ACTION. *Medidas de acurácia*. sd. Disponível em: http://www.portalaction.com.br/series-temporais/35-medidas-de-acuracia (Accessed: 17 December 2020).

PORTILHO, C. M. *Estimação da persistência de segurados de planos de previdência privada via modelos de sobrevivência*. 2013. 61f. Dissertação (Mestrado em Engenharia Elétrica) - Pontifica Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

SANTOS JÚNIOR, L. C. *Análise de sobrevivência aplicada a premissas atuariais: o caso da previdência pública municipal de Cabedelo/PB*. 2018. 162f. Tese (Doutorado em Biometria) – Universidade Estadual Paulista, Botucatu, São Paulo, 2018.

SCHOENFELD, D. Partial residuals for the proportional hazards regression mode. *Biometrika*, v.69, n.1, p.239-241, 1982.

SILVA, R. P. Modelos flexíveis de sobrevivência com fração de cura: implementação computacional. 152f. Dissertação (Mestrado) - Programa de Pós-Graduação em Matemática e Estatística Aplicada, Universidade Federal do Rio Grande do Norte, Natal, 2015.

SOUZA, R. M. L. *O mercado de saúde suplementar no brasil: regulação e resultados econômicos dos planos privados de saúde*. 2014. 296f. Tese (Doutorado em Políticas Públicas) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

THERNEAU, T. M. *A Package for Survival Analysis in S*. 2015. Available at: https://CRAN.R-project.org/package=survival (Accessed: 21 March 2019).

VIOT, M. *Seguro de pessoas, de saúde e planos de saúde*. Rio de Janeiro: ENS, 2019. 100p.

Received on 13.05.2020

Approved after revised on 04.01.2021