SOCIOECONOMIC ANALYTICS

Smoking recurrence: A Survival Analysis

Ana Beatriz Ramos da Silva [1] *
Débora de Sousa Cordeiro [2]
Giovana de Souza [3]

[1 2 3] Departamento de Estatística, State University of Paraíba, Campina Grande, Brazil
E-mails: [1] ana.beatriz.ramos@aluno.uepb.edu.br; [2] debora.cordeiro@servidor.uepb.com.br; [3] giovana.souza@aluno.uepb.edu.br.
* Corresponding author

How to cite this paper: da Silva, A. B. R., Cordeiro, D. S., Souza, G., (2024). Smoking recurrence: A Survival Analysis. *Socioeconomic Analytics,* 2(1), 144-163 .https://doi.org/10.51359/2965-4661.2024.265079

RESEARCH ARTICLE

Socioeconomic Analytics
https://periodicos.ufpe.br/revistas/SECAN/
ISSN Online: 2965-4661

Submitted on: 18.11.2024.
Accepted on: 15.12.2024.
Published on: 27.12.2024.

Abstract

In this work, we employ Survival Analysis to model the time until smoking relapse, considering social, demographic, and clinical variables, such as age group, gender, type of treatment, and time as a smoker. Survival curves were estimated using the Kaplan-Meier and Nelson-Aalen methods, showing group differences. The Logrank and Wilcoxon tests indicated significance only for the age group. In the Cox modeling, the type of treatment and age group stood out as relevant. It was concluded that combined treatments with three interventions are more effective, especially for young people who require greater support and awareness.

Keywords

Survival analysis, Hypothesis Tests, smoking, estimation, public health.

1. Introduction

Smoking is a disease characterized by dependence on the nicotine found in cigarettes. According to the Pan American Health Organization and the World Health (WHO) (2019) smoking is considered a pandemic, currently responsible for around five million deaths a year, most of them in men. If the prevalence Organization of tobacco use does not change, it is estimated that by 2025 this figure will rise to ten million deaths a year (Araújo *et al.*, 2004). This dependence causes physical, psychological damageand behavioral. The damage to health is exorbitant and, the ones that main occur most frequently are lung, cancerheart attacks, chronic bronchitis, pulmonary emphysema and strokes (WORLD HEALTH ORGANIZATION, 2023).

Quitting smoking has numerous benefits: After 20 minutes, blood pressure and pulse rate return to normal. After 2 hours, nicotine no longer circulates in the blood and within 12 to 24 hours the lungs function better (Araújo *et al.*, 2004). sense of smell and taste also become more acute. As it is an addiction, the difficulties of quitting are exorbitant, must include psychological support, medication, group therapies and changes in habits help so treatments for quitting tocurb anxiety addiction. Furthermore, according to Abreu and Caiaffa (2011), among adolescents and young adults the problem of smoking is even more influential, emphasis on the population with less schooling, males and hyperbolic alcohol consumption. Other studies, such as that by Malcon *et al.* (2003), have identified an association between tobacco addiction and a low level of schooling among adolescents.

Knowing this, in order to evaluate the behavior of patients who are quitting, the time until smoking relapse will be evaluated considering some characteristics of the patient's profile, such as the type of treatment, the patient's age, gender, type of addiction and length of time smoking, among other characteristics, which Silva *et al.* (2014) point out as essential to characterize the patient's profile. For this purpose, the analysis approach survival will be used, which is extensively applied in studies where it is necessary to monitor and understand the time until the occurrence a certain event, which in this case ofis smoking relapse (Klein & Moeschberger, 2003). This method seeks to understand the differences in abstinence times according to the patient's demographic, socioeconomic and clinical profile.

2. Materials and Methods

The data being analyzed refers to 125 patients being studied for treatment against smoking, the treatments applied being triple therapy and nicotine patch. This data was taken from the GitHub platform, published by Mota Vieira (2020). In addition, these data include other variables besides the type of treatment the patient is undergoing. The variables analyzed together with the time until smoking relapse are categorical.

With regard to the treatment offered to patients, there are two main: Types the exclusive use nicotine patches and a combined approach that uses three different pharmacological methods simultaneously, which is usually made up of nicotine replacement (NRT) which denotes the function of the nicotine patch, bupropion (BUP) and varenicline (VAR), where BUP helps with withdrawal symptoms and VAR acts on the brain's nicotine receptors by decreasing the pleasure associated with tobacco consumption, as cited by Dantas *et al.* (2016). With regard to age, the categories were defined based on factors associated with the typical characteristics of each age group. For patients under 25, believed to social factors are play a significant role in relapse. For those between aged 25 and 44, which includes young and middle-aged, adults there is less social, but if the addiction was acquired in youth, these individuals may have a more consolidated dependence on tobacco. Similarly, the 45 to 64 age group probably carries an even higher smoking burden due to greater exposure to tobacco throughout life. For patients aged 65 and over, the health problems resulting from prolonged tobacco use, as well as other conditions associated with ageing influence can serve as a strong incentive to give up smoking. According to Goulart *et al.* (2010) the patient's age group directly influences their recovery, considering that the groups to which the patient belongs have their own characteristics. In addition, smokers over the age of 50 in particular find it more difficult to give up the habit.

With regard to gender, the categories analyzed are male and female, respectively. Race was categorized as white, black, Hispanic and others. The time the patients worked in this study was classified into three groups: Full-time, part-time and other. Considering the length of time they had been smoking, the patients were divided into three groups: Those with less than 15 years of addiction, indicating intermediate; dependence between 16 and 30 years, where dependence is strongly established; and over 30 years, where the risk of relapse tends to be even greater due to a long history of dependence on tobacco.

The categories for the number of quit attempts were defined on based the following perceptions: Patients who have never tried to quit may have a strong dependence or lack of motivation to quit. Those who have tried (zero attempts) 1 or 2 times and failed indicate moderate dependence. Patients who have tried between 3 and 5 times without success suggest a higher level of dependence, possibly facing greater difficulties in the process of quitting. For those with more than 5 attempts, this may indicate severe dependence and major challenges in the process of quitting, despite having significant motivation to quit.

The categories for the number of days without smoking were defined follows: Those who can't abstain for even one day show strong dependence. Patients who go between 1 and 30 days without smoking show that they are trying to quit but cannot sustain abstinence for more than a month. Those who manage to go between 31 and 180 days without smoking indicate greater control over their addiction. Patients who abstain for more than 6 months show significant control over the situation. These characteristics of the patient's clinical profile were established based on the guidelines addressed by Lopes *et al.* (2023).

The exploratory analysis of the individuals under study was carried out in order to understand how the patients condition is constituted. According to Bussab and Morettin (2017), graphs and frequency tables are very useful when we seek to describe the behavior of data when working with qualitative characteristics, with this in mind, graphs of were plotted frequency so that this behavior could be understood.

2.1. Survival Analysis

Survival analysis is a powerful tool from a group of statistical techniques used when trying to assess the time until the recurrence of an expected event (Hosmer, Lemeshow and May, 2008). This technique is widely applied in various economic, social and industrial areas, as well as in medicine, which pioneered the studies that led to the widespread use of this tool. In this case, we will use Survival Analysis to assess the time to relapse of patients studying smoking, with the aim of identifying the possible causes that make it more difficult for those seeking to quit.

Survival analysis is extremely useful when the data lacks information during the analysis period, as this lack of information makes other analyses impossible (Colosimo and Giolo, 2021). This lack of information is referred to as censoring and the occurrence of censoring determines the entire analysis approach.

The survival data for a study subject are represented in general by the pair (τ_i, δ_i), where τ_i is the time of failure of the individual and δ_i it is indicative of the status of the individual at that time, that is, whether he failed or censored, thus δ_i will receive 1 if the individual has failed, relapsed into addiction, and 0 if the individual is censured.

Censoring can occur in different ways, in this study the survival approach is clinical, so censoring can occur for various reasons and some of them may be the loss of patient follow-up or the non-occurrence of the event until the end of the study.

2.2. Survival Function

The survival function is widely used to discuss survival analysis. This function denotes the probability that will the individual not fail by a certain time. This function is written as follows:

$$\boldsymbol{S(t) = P(T \geq t)} \qquad \textbf{(1)}$$

Thus, the cumulative distribution function that shows the probability of an individual failing in a given time t is defined by:

$$\boldsymbol{F(t) = 1 - S(t)} \qquad \textbf{(2)}$$

The survival function is a parametric approach that can only be used in cases where there is no censoring, which is uncommon in real studies. To avoid censoring, non-parametric approaches are used to estimate survival curves and cumulative risk. The most widely used estimator is the Kaplan-Meier (1958) estimator, followed by the Nelson-Aalen estimator, which was first proposed by Nelson (1972) and improved by Aalen (1978), both of which will be used in this study. The Kaplan-Meier estimator is defined as:

$$\boldsymbol{\hat{S} = \prod_{j:t_j<t} \left(\frac{n_j - d_j}{nj}\right) = \prod_{j:t_j<t} \left(1 - \frac{dj}{nj}\right)} \qquad (3)$$

where t is the time, n_j is the number of individuals at risk and d_j the number of failures that occurred. The Nelson-Aalen estimator is based on the survival function which is defined by:

$$\boldsymbol{S(t) = exp\{-\Lambda(t)\}} \qquad \textbf{(4)}$$

where $\boldsymbol{\Lambda}(t)$ is the accumulated risk function, which is estimated by:

$$\hat{\Lambda} = \sum_{j:j<t} \left(\frac{d_j}{n_j}\right) \quad (5)$$

Thus, the Nelson-Aalen estimator for the survival function is expressed as follows:

$$S(t) = exp\{-\hat{\Lambda}\} \quad (6)$$

In general, Kaplan-Meier estimates tend to be closer to the value of $S(t)$.

2.3. Comparison tests

In many studies, the main interest is to assess whether there is a difference in the survival of individuals analyzed by interest groups that share the same characteristics in the study. For this, non-parametric tests are used, such as the Logrank and Wilcoxon tests. The Logrank test, according to Mantel (1966), is the most used in survival studies, although the Wilcoxon test is more recent (Gehan, 1965). The calculation of the test statistic by the Logrank test is constructed as follows:

$$T = \frac{\left[\sum_{j=1}^{k}(d_{2J} - w_{2j})\right]^2}{\sum_{j=1}^{k}(V_j)_2} \quad (7)$$

For large samples, the statistic T follows a Chi-squared distribution with 1 degree of freedom. Other tests widely used in the literature, such as the Wilcoxon test, generalize this statistic by adding weights. These weights vary between the tests, which results in the differences observed between them. This generalization can be expressed as follows:

$$S = \frac{\left[\sum_{j=1}^{k} u_j(d_{2J} - w_{2j})\right]^2}{\sum_{j=1}^{k} u_j^2(V_j)_2} \quad (8)$$

in which in the Wilcoxon test the weight $\mathbf{u}_j = \mathbf{n}_j$, assigning greater weight to events that occur earlier, while in Logrank $\mathbf{u}_j = \mathbf{1}$. We consider that there is a difference between these curves when the p-values obtained from the test statistics are higher than the established level of significance.

2.4. Modeling

For this study, the approach modeling used will be semi-parametric modeling or models Cox, since parametric does not suit the behavior of the data. Cox regression models (Cox, 1972) have brought new perspectives to survival analysis, especially for clinical cases. To follow Cox's modeling, the groups evaluated need to have proportional risks (Struthers and Kalbfleisch, 1986). Thus, assuming proportionality of risks, we have:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = K \qquad (9)$$

Thus, K is the ratio of failure rates or relative risk, constant for the entire follow-up period. The Cox regression model is characterized by the coefficients that measure the effects of the covariates on the failure rate function. For an estimation method, necessary, inference is and one of the inferential methods widely used is the maximum likelihood method [see (Cox and Hinkley, 1974)]. However, the presence of the non-parametric component makes the maximum likelihood method unfeasible. With this in mind, Cox (1975) proposed a partial. maximum likelihood method. Thus, the modeling is semi-parametric. Considering p variables, Cox's model has the following form:

$$\lambda(t) = \lambda_o(t) g(x'\beta) \qquad (10)$$

where g is a function that must be specified, so that $g(0) = 1$. The model is composed of two components, one parametric and one non-parametric.

The interpretation of covariates is based on whether they accelerate or decelerate the risk function. The risk ratio of failure two sample items i and j, which have the same covariate values of except for the ith, can be evaluated as the risk ratio at time t, provided that the risks are proportional [see (COLOSIMO; GIOLO, 2021)].

Still on the validation of the model, metrics such as the coefficient of agreement offered by Harrell *et al.* (1982) will be used, this coefficient varies from 0.5 to 1, in which, above 0.5, we understand that the occurrence of the event of interest did not occur by chance between the observations. In addition, the Wald test, also proposed by Cox (1972), evaluates the significance of the coefficients that exist in the model, observing whether the null option that the coefficient is equal to zero is true. Finally, the Akaike criterion (AIC) is a measure used to measure the quality of models, the AIC penalizes models that have a large number of parameters seeking to avoid an overfit, considering as the best model the one that has the lowest AIC [see (Akaike, 1974)].

3. Results and Discussions

Initially, the frequency tables were evaluated so that the behavior of the characteristics of the patient's profile is understood.

Table 1 - Frequency distribution for patient profile characteristics.

Variable	Frequency
Age	
< 25 years	2
25 to 44 years	42
45 to 64 years	70
> 65 years	11
Gender	
Female	81
Male	44
Race/Color	
Black	38
White	77
Hispanic	8
Other	2
Type of Employment	
Part-time	14
Full-time	72
Other occupation	39

According to Table 1, there is a predominance of women in the patient population, and the majority of these patients are generally white, work full-time and are between 25 and 64 years old. Although the prevalence of people addicted to tobacco is higher among the male, both in this study and in the study by Lopes *et al.* (2023), it was observed that the proportion of women followed up in treatment was higher than that of men population.

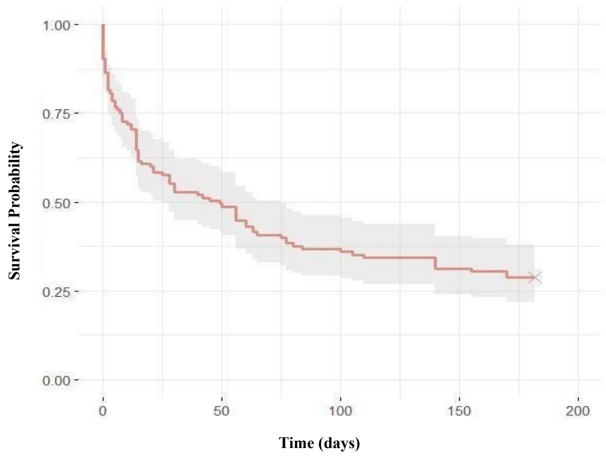
Table 2 - Frequency distribution for the patient's clinical profile.

Variable	Frequency
Treatment Group	
Nicotine patch	64
Combination	61
Smoking time	
< 15 years	12
16-30 years	52
> 30 years	59
Number of times tried to quit smoking	
Never tried to quit the habit	11
1-2 attempts to quit addiction	52
3-5 attempts to quit addiction	35
More than 5 attempts to quit the addiction	27

Number of days the patient has gone without smoking previously	
1-30 days without smoking	40
31-180 days without smoking	31
> 180 days without smoking	42
Not a day without smoking	12

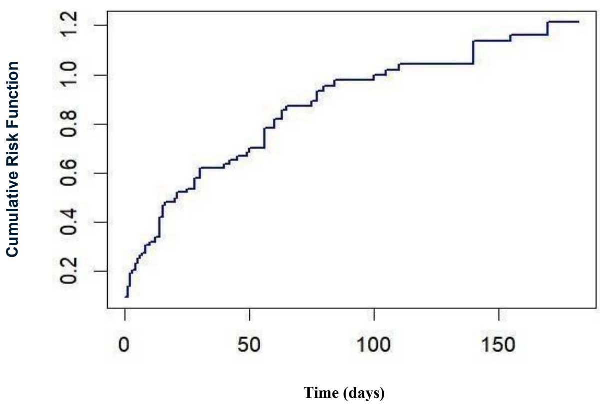
In Table 2, the treatments were applied proportionally to the patients. In addition, most of these patients have been smoking more than 15 years, in short, they have tried to quit 1-2 times and have gone up to 30 days and more than 180 days without smoking in previous attempts. Next, the Kaplan-Meier and Nelson-Aalen risk curves will be presented to assess the likelihood of the individual under study relapsing into smoking.

Figure 1 - Survival Curve (Kaplan-Meier) for time to smoking recurrence.



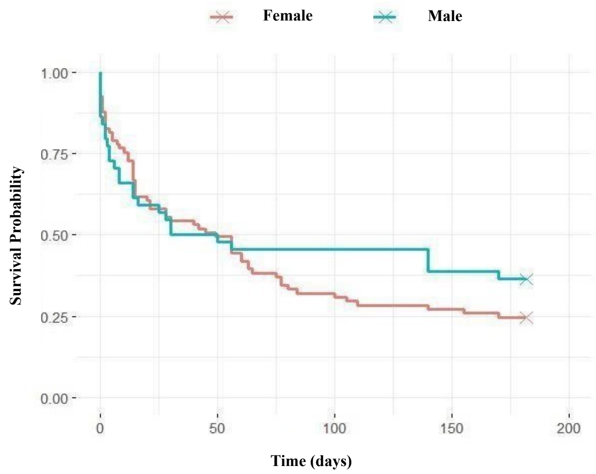
The Kaplan-Meier estimates of the time until smoking relapse indicate a high risk of relapse in the first days after the start of treatment, showing that by the 50th day of treatment, half of the patients had already returned to the habit. The survival curve for both methods shows a sharp decrease in this initial period, suggesting a weakness in recent abstinence. However, the relapse rate decreases significantly over time, indicating greater long-term abstinence stability.

Figure 2 - Nelson-Aalen Cumulative Risk Function.



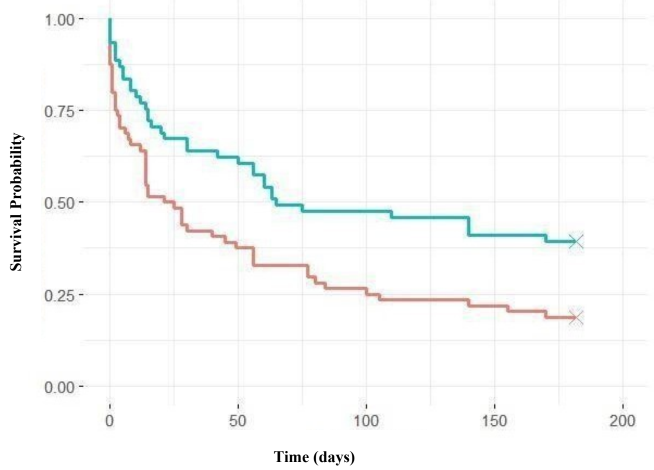
The cumulative risk function is an increasing estimate, in which the current risk is always added to that of the previous time. Evaluating the risk curve, the indications perceived in the survival curves are reiterated, which indicate that the risk of relapse in the first days of the intervention is high, so the function increases rapidly in the first days, indicating that many patients returned to smoking at the beginning of the treatment. It is possible to evaluate the decay behavior in relation to different categories and whether there are differences between these categories in terms of smoking recurrence.

Figure 3 - Survival curve by genre.



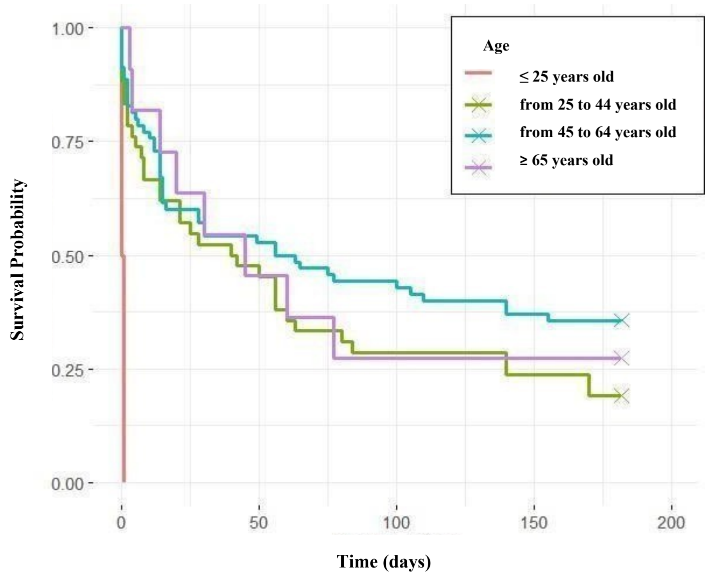
Evaluating the difference between gender, we have evidence that the survival curves are similar, especially at the beginning of the study, after a certain time, female patients are more likely to return to smoking.

Figure 4 - Survival curve for treatments.



Looking at the survival curves, the evidence observed is that combination treatment offers better results than treatment with the nicotine patch alone.

Figure 5 - Survival curve for the patient's age group.



For the age group, it is visually observed that patients aged 45-64 years have a higher probability of not recurring the disease. The evidence suggests that younger patients tend to relapse more easily. The other curves referring to the patient's profile, present in the appendix, showed slight or no evidence that they affect the probability of patients relapsing into the habit of smoking.

Table 3 - Comparison test for survival curves.

Variables	Log-rank	Wilcoxon
Gender	0.4	0.8
Race	0.5	0.4
Type of employment	0.3	0.4
Type of addiction	0.9	0.8
Age	< 0,01***	< 0.01***
Smoking time	0.2	0.3
Number of times that tried to quit	0.9	0.8
Number of days previously spent without smoking	0.2	0.1

From the tests comparing survival curves, it is possible to observe that, among all the covariates analyzed, only the age group proved to be significant in terms of the time until the recurrence of the smoking habit, i.e., the age of the patients possibly influences the cessation of cigarette addiction. Cox modeling was initially performed for all characteristics of the patient's profile. And the following results were obtained.

Table 4 - Model estimates considering all variables associated with the patient profile.

Variable	Coefficient	Odds Ratio	Standard Error	Z-value	P-value
Treatment: Combination	-0.5799	0.56	0.2361	-2.456	0.0141**
Age: 25-44 years	-1.113	0.3287	0.9676	-1.15	0.2502
Age: 45-64 years	-1.413	0.2434	1.002	-1.409	0.1587
Age: 65 or more	-1.452	0.2341	1.104	-1.316	0.1883
Days without smoking: 31-180 days	0.09603	1.101	0.3532	0.272	0.7857
Days without smoking: > 180 days	-0.2947	0.07447	0.3424	-0.861	0.3893
Days without smoking: < 1 day	-15.17	2.59E-07	2228	-0.007	0.9946
Gender: Male	-0.1711	0.8427	0.2575	-0.665	0.5063
Race: Hispanic	-0.3437	0.7091	0.5076	-0.677	0.4983
Race: Other	-1.121	0.3261	1.049	-1.069	0.2853
Race: Black	0.2685	1.308	0.2738	0.981	0.3268
Type of occupation: Other	-0.127	0.8807	0.443	-0.287	0.7744
Type of occupation: Full-time	-0.574	0.5633	0.4189	-1.37	0.1706
Addiction: Heavy	0.1136	1.12	0.2895	0.392	0.6947
Smoking time: > 35 years	-0.1014	0.9035	0.3134	-0.324	0.7462
Smoking time: < 15 years	0.2065	1.229	0.4451	0.464	0.6427
Attempts to quit smoking: 3-5 times	-0.1808	0.8346	0.3029	-0.597	0.5505
Attempts to quit smoking: > 5	0.0075	1.008	0.3152	0.024	0.981
Never tried	15.95	3111000	2228	0.007	0.9946

*Significant at 5% confidence.

When evaluating the proportionality of the risks of this model, it was observed that the genre does not have proportionality of the risks, thus, the assumption of the model was made unfeasible and thus, it was necessary to build another model that does not violate this assumption. For this, a new model was adjusted, which considers the gender variable as a stratification, so this variable is no longer direct in the analysis, does not have its own coefficient, but adjusts the model to deal with the differences between the strata, promoting a better fit without violating assumptions. The estimates obtained by this model will be presented in the following table.

Table 5 - Model estimates considering gender as stratification.

Variable	Coefficient	Odds Ratio	Standard Error	Z-Value	P-value
Treatment: Combination	-0.5764	0.5619	0.2371	-2.431	0.0150 **
Age: 25-44 years	-1.767	0.1708	1.006	-1.756	0.0791 *
Age: 45-64 years	-2.07	0.1262	1.045	-1.981	0.0476 **
Age: 65 or more	-2.165	0.1147	1.147	-1.888	0.0590 *
Days without smoking: 31-180 days	0.0704	1.073	0.3558	0.198	0.8432
Days without smoking: > 180 days	-0.3229	0.724	0.345	-0.936	0.3493
Race: Hispanic	-0.4011	0.6696	0.5114	-0.784	0.4329
Race: Other	-1.178	0.3078	1.05	-1.123	0.2616
Race: Black	0.2776	1.32	0.2751	1.009	0.3131
Type of occupation: Other	-0.0932	0.911	0.4485	-0.208	0.8354

Type of occupation: Full-time	-0.4995	0.6068	0.4225	-1.182	0.2371
Addiction: Heavy	0.087	1.091	0.2951	0.295	0.7683
Time of smoking: > 35 years	-0.0512	0.9501	0.3209	-0.16	0.8732
Time of smoking: < 15 years	0.1835	1.201	0.4476	0.41	0.6818
Attempts to quit smoking: 3-5	-0.1508	0.86	0.306	-0.493	0.6222
Attempts to quit smoking: >5	0.0294	1.03	0.3155	0.093	0.9258
Never tried	0.1553	5.539	2.729	0.006	-

*Significant at 5% of confidence. **Significant ate 1%.

As shown in Table 5, after stratifying gender, more factors began to influence the recurrence of smoking in patients. Before interpreting the results, the proportionality test was performed and the following results will be presented in the following table:

Table 6 - Risk proportionality test.

Variable	P-value
Treatment group	0.951
Age	0.425
Time in days without smoking	0.789
Race	0.313
Type of occupation	0.588
Type of addiction	0.243
Time of addiction	0.055
Attempts to quit smoking	0.318
Global	0.350

The graphs inherent to the proportional risk validations will be presented in the appendix. All p-values obtained were greater than 0.05. Thus, it is assumed that the risks are proportionate and the violation has been circumvented. Then, the other validations of the model will be presented.

Table 7 - Model validations.

Tests	Value	P-value
Concordance	0.653	-
Likelihood ratio	26.16	0.1
Wald	26.77	0.08 *
Score (log-rank)	39.27	0.003**

*Significant at 5% confidence. **Significant at 1%.

The model considering all variables, but stratifying gender was the best among the models tested, achieving better results without violation of assumptions and achieving the lowest AIC (Akaike Criterion). In the complete model without gender stratification, the AIC was 769.4827 and in the complete model with gender stratification, the AIC was 653.0448, indicating a better fit in which, although most of the variables are not directly significant, they influence the model.

As can be seen in Table 7, the model has a discriminatory power of 65.30%, and the score highly significant log-rank indicates that there are differences between the groups of patients analyzed, a fact that is corroborated by the marginal significance of the Likelihood Ratio and Wald test.

With the modelstratified Cox, it is possible to evaluate factors that influence the time until smoking recurs, considering gender as a stratification variable. The treatment applied to the patients showed significant results. In the combined treatment, the estimated coefficient was -0.576, and its exponential, which indicates the odds ratiogroup, corresponds to 0.562. This suggests that combined treatment reduces the patient's risk of returning to smoking by 44%. Dantas *et al.* (2016) point out that other drugs such as clonidine and nortripline can be used as an alternative second option for triple therapy since they are not as effective as BUP and VAR.

As for age group, the 25-44 age group had an odds ratio of 0.1708, indicating that patients in this age group have an approximately 82% reduction in the likelihood of relapse compared to patients aged up to 25. In the 45-64 age group, the results were even more significant, with an odds ratio of 0.126, suggesting a reduction of around 87% in the risk of smoking relapse. For patients aged 65 and over, the odds ratio indicates a reduction of approximately 88% in the likelihood of relapse compared to the younger group. These results are consistent with those obtained by Bertoni and Szklo (2021), who showed that the prevalence of using electronic devices to smoke is 10 times higher among younger people than among older age groups.

4. Conclusions

This study made it possible to assess the impact of the patient's clinical profile on smoking cessation treatment. The initial evaluation of the survival curves showed that most of the characteristics of the patient's profile influenced their treatment for smoking. For the patient's type of treatment in the survival curves, the indication was that there would be a difference between the patient's treatment and their recovery; after this suggestion, the curve comparison tests were carried out and no such difference was detected. However, after implementing the Cox model, this difference was significant, so that the treatment consisting of three different methods reduces the risk of the patient returning to smoking by around 44% when compared to the treatment with the nicotine patch alone.

Furthermore, among the sociodemographic characteristics of the patients, only the age group showed significant results in terms of the characteristics addressed. This difference was detected in the curve comparison tests and subsequently brought intriguing results in the modeling. The 45-64 age group proved to be the most relevant among the others, showing the greatest statistical significance according to its p-value, where, when compared to patients aged up to 25, the risk of returning to addiction is reduced by more than 80%. However, the age groups of 25-44 years and over 60 years also showed significant results in terms of reducing the risk of returning to smoking, where a reduction of around 82% and 87% in the risk of returning to smoking was detected respectively in relation to the age segment of up to 25 years.

Still on the subject of the patient's profile, evaluating their clinical profile, which includes measuring time in years that the patient has been exposed to tobacco, strength of the virus, the number of times they have tried to quit and the maximum level of abstinence showed no significant differences in relation to the patient's relapse. Finally, based on the results obtained in this analysis, the most successful treatment for quitting is a combination of therapeutic methods. This reinforces the importance of making combined treatment widely available, given its superior effectiveness in smoking cessation. As for the age group of the patients, the younger group, aged up to 25, proved to be more susceptible to relapsing into smoking. To address this challenge, we recommend developing awareness campaigns specifically aimed at this audience

Acknowledgments

We would like to thank the State University of Paraíba (*Universidade Estadual da Paraíba*) for its contribution to the research and writing of this article, since the study that provided this article arose from a work prepared in a discipline offered by the institution in the Statistics course. Special thanks to UFMG teacher Dr. Rivert Paulo Braga Oliveira for the availability of the data. There are no funding sources for outreach.

Conflict of Interest Statement

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the content of the manuscript and there are no financial interests to report.

References

1. Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics, 6*(4), 701–726. https://doi.org/10.1214/aos/1176344247
2. Abreu, M. N. S., & Caiaffa, W. T. (2011). Tabagismo entre adolescentes e adultos jovens de Belo Horizonte, Minas Gerais, Brasil: Influência do entorno familiar e grupo social. *Cadernos de Saúde Pública*, *27*(1), 77-85. https://doi.org/10.1590/S0102-311X2011000100009.
3. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705
4. Araújo, A. J., Pereira, E. C., Siqueira, S. A., Silva, J. C. R., Souza, L. A., & Dias, W. A. (2004). Diretrizes para cessação do tabagismo. *Jornal Brasileiro de Pneumologia, 30*(Supl. 2), S1–S76.
5. Bertoni, N., & Szklo, A. S. (2021). Dispositivos eletrônicos para fumar nas capitais brasileiras: prevalência, perfil de uso e implicações para a Política Nacional de Controle do Tabaco. *Cadernos de Saúde Pública*, 37(7), e00261920.
6. Bussab, W. O., & Morettin, P. A. (2017). *Estatística básica* (8ª ed.). Saraiva Educação.
7. Colosimo, E. A., & Giolo, S. R. (2021). *Análise de sobrevivência aplicada* (3ª ed.). Blucher.
8. Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–220.
9. Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. Chapman and Hall.

10. Cox, D. R. (1975). Partial likelihood. *Biometrika, 62*(2), 269–276.

11. Dantas, D. R. G., Pinheiro, A. H. B., Rossoni, A. L. M., Prado, L. O., & Barreira, S. N. (2016). Tratamento do tabagismo no Brasil, com bupropiona ou vareniclina: uma revisão sistemática. *Revista Saúde & Ciência Online*, *5(1)*, 61-75.

12. Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2), 203–223. doi:10.1093/biomet/52.1-2.203

13. Goulart, D., Engroff, P., Ely, L. S., Sgnaolin, V., Santos, E. F. dos., Terra, N. L., & De Carli, G. A. (2010). Tabagismo em idosos. *Revista Brasileira De Geriatria E Gerontologia*, 13(2), 313–320.

14. Harrell, F. E., Lee, K. L., & Mark, D. B. (1982). Multivariate prognostic models: Issues in developing models, evaluating models, and measuring and reducing errors. *Statistics in Medicine, 1*(2), 141–150.

15. Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd ed.). Wiley.

16. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

17. Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.

18. Lopes, F. M., Moraes, C. A. D., Rodrigues, G., Cardozo, L., Bezerra, J. F. D. O., & Szupszynski, K. P. D. R. (2023). Efeito do Programa de Cessação do Tabagismo: uma revisão dessa política pública para dependência tabágica. *Estudos de Psicologia (Campinas)*, *40*, e210170.

19. Malcon, M. C., Menezes, A. M. B., & Chatkin, M. (2003). Prevalência e fatores de risco para tabagismo em adolescentes. *Revista de Saúde Pública*, 37, 1-7.

20. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50(3), 163–170.

21. Mota Vieira, J. (2020). *Análise de sobrevivência: Dados* [Repositório GitHub]. Github. https ://github .com /Joziani /Analise -Sobrevivencia /tree /main /dados

22. Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics, 14*(4), 945–966. https://doi.org/10.1080/00401706.1972.10488991

23. Organização Mundial da Saúde. (2023). *Tobacco*. Disponível em: https://www.who.int/news-room/fact-sheets/detail/tobacco.

24. Organização Pan-Americana da Saúde & Organização Mundial da Saúde. (2019). *Tabaco*. Disponível em: https://www.paho.org/pt/topicos/tabaco.

25. Silva, S. T. D., Martins, M. C., Faria, F. R. D., & Cotta, R. M. M. (2014). Combate ao Tabagismo no Brasil: a importância estratégica das ações governamentais. *Ciência & Saúde Coletiva*, *19*, 539-552.

26. Struthers, C. A., & Kalbfleisch, J. D. (1986). Misspecification of proportions in Cox regression model. *Biometrika, 73*(2), 363–36

Appendix

Figure 6: Survival curves for the race.

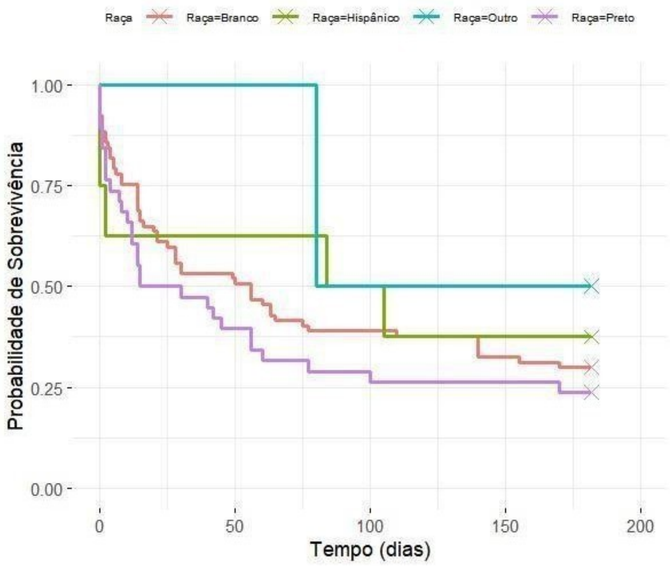


Figure 7: Survival curves for the type of addiction.

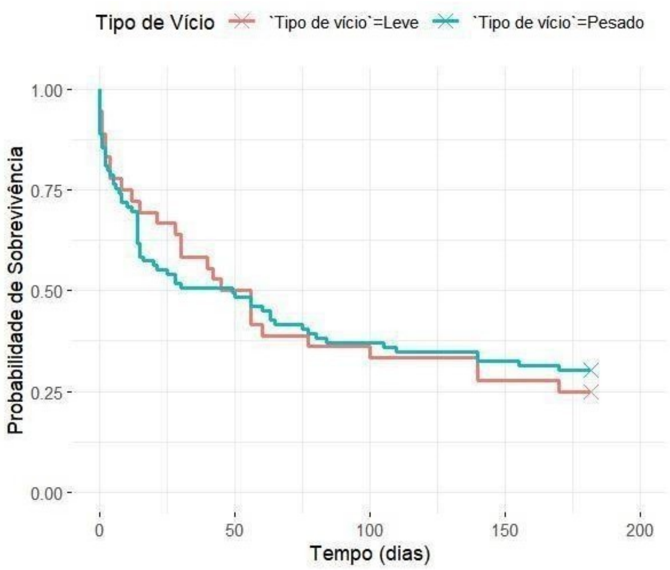


Figure 8: Survival curve for the patient's type of employment.

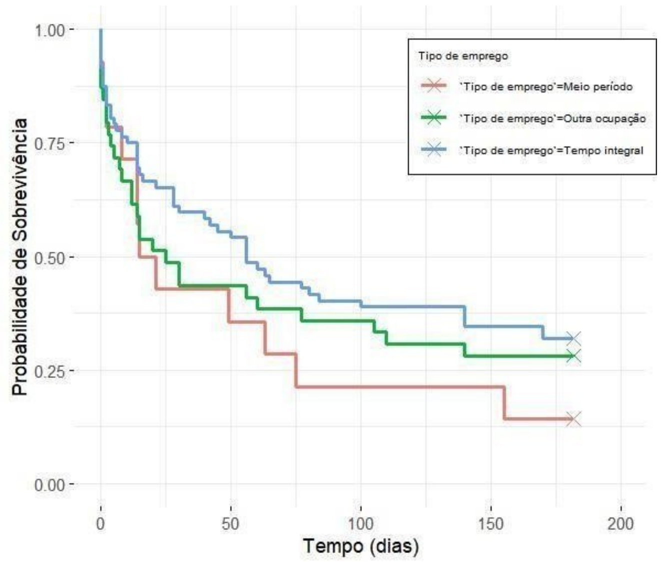


Figure 9: Survival curve for the number of times the patient tried to quit.

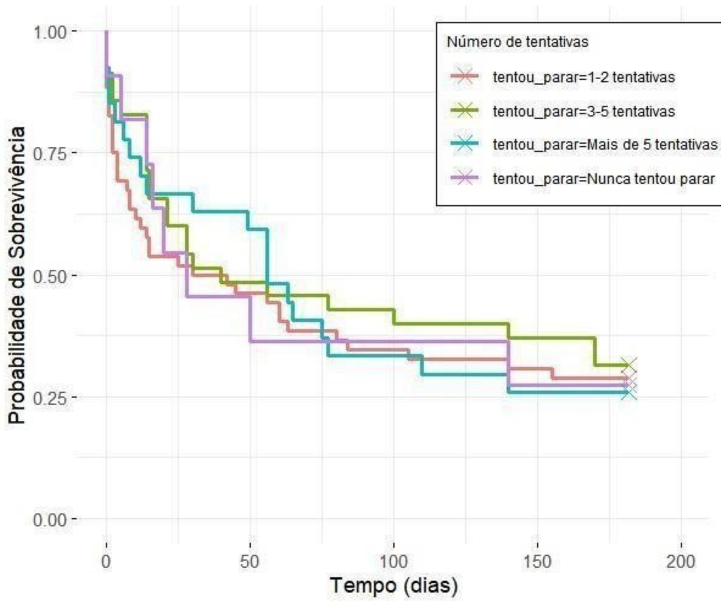


Figure 10: Survival curve for the number of days the patient had not smoked previously.

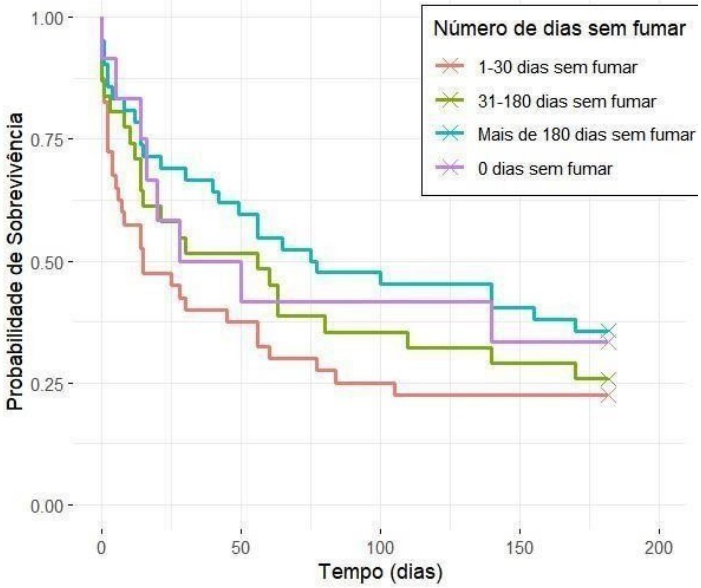


Figure 11: Survival curve for the time in years that the patient has smoked.

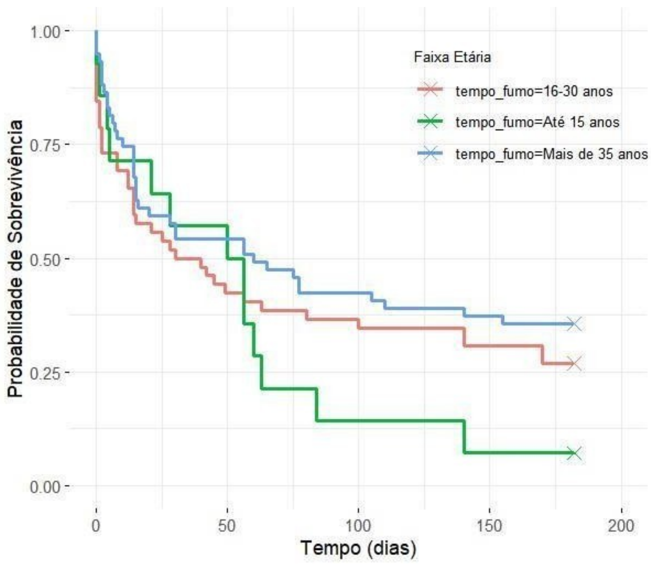


Figure 12: Risk proportionality graph for the age group.

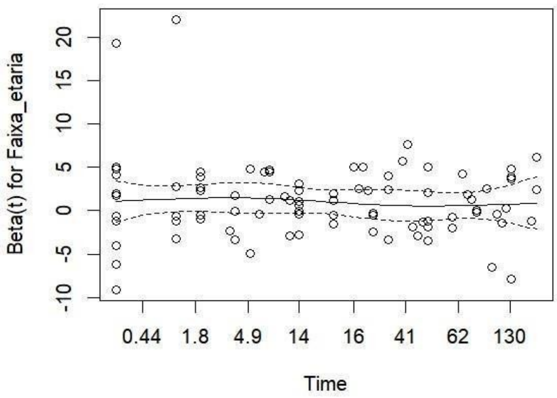


Figure 13: Risk proportionality graph for the treatment group.

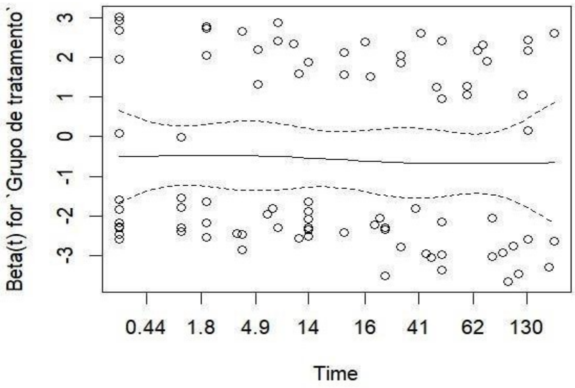


Figure 14: Risk proportionality graph for the type of addiction.

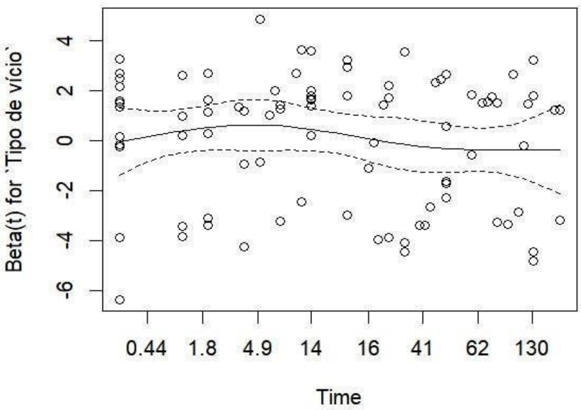


Figure 15: Proportionality graph of the risks for the number of times you tried to quit.

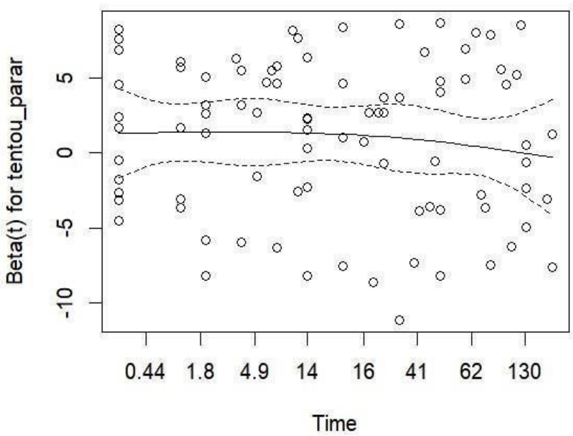


Figure 16: Proportionality graph of the risks for the time in years that the patient has spent smoking.

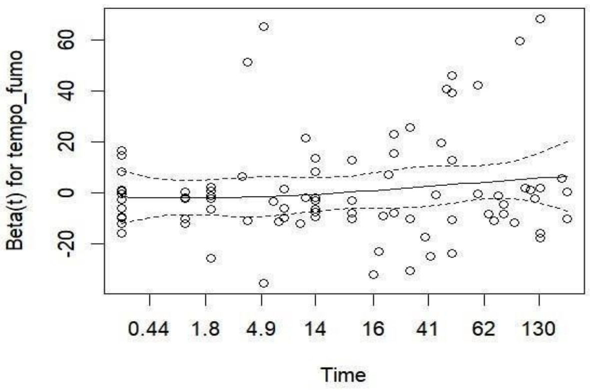


Figure 17: Proportionality graph of the risks for the patient's race.

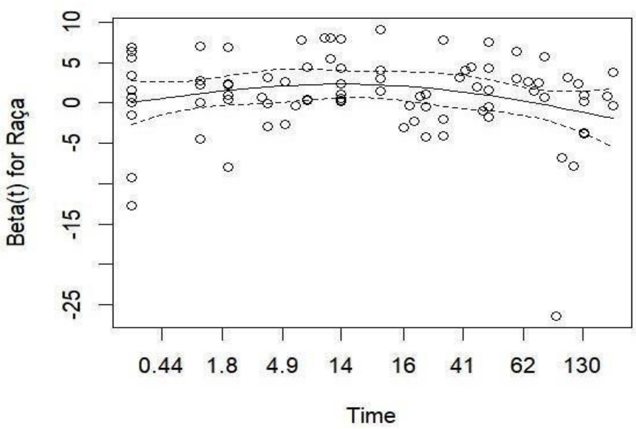


Figure 18: Graph showing the the proportionality of the risks to time in days that the patient had not smoked previously.