Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

https://doi.org/10.23913/ride.v11i22.930

Artículos científicos

Análisis estadístico de gestión de riesgo en pacientes con Cacu en Cd. Juárez Chihuahua

Statistical Risk Management Analysis in Patients With CaCu on Cd. Juárez Chihuahua

Análise estatística da gestão de risco em pacientes com Cacu em Cd. Juárez Chihuahua

Luz Elena Terrazas Mata
Tecnológico Nacional de México, Campus Ciudad Juárez, México
lterrazas@itcj.edu.mx
https://orcid.org/0000-0003-1396-1751

Alfonso Aldape Almillo
Tecnológico Nacional de México, Campus Ciudad Juárez, México
aaldape@itcj.edu.mx
https://orcid.org/0000-0001-8921-9479

Manuel Arnoldo Rodríguez Medina
Tecnológico Nacional de México, Campus Ciudad Juárez, México
manuel_rodriguez_itcj@yahoo.com
https://orcid.org/0000-0002-8922-4718

Juan Pedro Benítez Guadarrama
Universidad Autónoma del Estado de México, México
jpbenitezg@uaemex.mx
https://orcid.org/0000-0002-2826-6359

Inocente Yuliana Meléndez Pastrana
Tecnológico Nacional de México, Campus Ciudad Juárez, México
yuliana.mp@cdjuarez.tecnm.mx
https://orcid.org/0000-0002-5260-7977

Resumen

Durante los últimos años se ha incrementado el desarrollo de modelos probabilísticos y de análisis estadísticos aplicados a datos de supervivencia. El modelo de riesgos proporcionales de Cox se ha convertido en uno de los más empleados, principalmente en temas relacionados con la salud, pues resulta particularmente útil para comparar grupos en los que se estudia el tiempo transcurrido hasta la ocurrencia de un evento, pudiendo analizar conjuntamente el efecto de varias variables. Por ende, el objetivo de este estudio es presentar los parámetros de validación de los supuestos del modelo de riesgos proporcionales de Cox como una herramienta para la toma de decisiones respecto al tiempo de ocurrencia del fenómeno, comparación de grupos, efectos de diferentes factores sobre el evento de desenlace o falla, así como destacar características que identifican a grupos de riesgo con el fin de proporcionar a los profesionales de la salud información estadística que les ayude a fundamentar decisiones en materia de diagnóstico, pronóstico y tratamiento. Esta, por tanto, es una investigación transversal exploratoria en donde se reportan los resultados obtenidos de un análisis estadístico realizado a los datos derivados de expedientes y entrevistas a un grupo de 127 pacientes diagnosticadas con cáncer cervicouterino durante el periodo 2013-2017 en la clínica de colposcopia de la Jurisdicción Sanitaria II en Cd. Juárez Chihuahua. Se analizaron los factores *edad, diabetes, escolaridad, estado civil* y *cantidad de parejas* con el fin de determinar su tasa de riesgo mediante el modelo de riesgos proporcionales de Cox. El estudio identificó que los factores que tiene mayor efecto en la supervivencia de pacientes son la edad y la diabetes. El modelo de Cox permite utilizar información de sujetos que se pierden durante el seguimiento, como sucede en los análisis de supervivencia, además de que toma en cuenta todos los factores de riesgo que pueden intervenir en un resultado dado.

Palabras clave: aplicación parámetros de validación, modelo de Cox, razón de riesgo, supervivencia.

Abstract

In recent years, the rapid development of probabilistic models and statistical analysis applied to survival data has increased. The Cox proportional hazards model has become one of the most widely used methods, mainly in health-related issues, it is particularly useful for comparing groups in which the time elapsed until the occurrence of an event is studied, being able to jointly analyze the effect of several variables.

The objective of this study is to present the validation parameters of the assumptions of the Cox proportional hazards model as a tool for decision-making, regarding the time of occurrence of the phenomenon, comparison of groups, effects of different factors on the event of outcome or failure, as well as highlighting characteristics that identify groups at risk; in order to provide health professionals with statistical information that helps them inform decisions regarding diagnosis, prognosis and treatment; thus also highlighting these strengths in different fields of study that seek to identify these behaviors.

This is an exploratory cross-sectional investigation where the results obtained from a statistical analysis performed on the data derived from records and interviews with a group of 127 patients diagnosed with cervical cancer are reported during the period from 2013 to 2017 in the colposcopy clinic of the Jurisdicción Sanitaria II in Cd. Juárez Chihuahua.

Age, diabetes, education, marital status and number of partners were analyzed in order to determine their risk rate using the Cox proportional hazards model. The study identified that the factors that have the greatest effect on patient survival are age and diabetes.

The Cox model allows the using information from subjects who are lost to follow-up, as in survival analyzes, in addition to taking into account all the risk factors that may intervene in a given result, it helps to substantiate clinical judgment.

Keywords: application validation parameters, Cox model, hazard ratio, survival.

Resumo

Nos últimos anos, o desenvolvimento de modelos probabilísticos e análises estatísticas aplicadas a dados de sobrevivência tem aumentado. O modelo de riscos proporcionais de Cox tornou-se um dos mais utilizados, principalmente em questões de saúde, pois é particularmente útil para comparar grupos em que se estuda o tempo decorrido até a ocorrência de um evento, podendo analisar conjuntamente o efeito. de várias variáveis. Portanto, o objetivo deste estudo é apresentar os parâmetros de validação dos pressupostos do modelo de riscos proporcionais de Cox como ferramenta para a tomada de decisões quanto ao tempo de ocorrência do fenômeno, comparação de grupos, efeitos de diferentes fatores no desfecho ou falha evento, bem como destacar características que identificam grupos de risco, a fim de fornecer aos profissionais de saúde informações estatísticas que os auxiliem a informar decisões sobre diagnóstico, prognóstico e tratamento. Trata-se, portanto, de uma investigação exploratória transversal onde os resultados obtidos a partir de uma análise estatística realizada com os dados provenientes de prontuários e entrevistas com um grupo de 127 pacientes com diagnóstico de câncer de colo uterino durante o período de 2013-2017 no ambulatório de Colposcopia da Vara Sanitária II em Cd. Juárez Chihuahua. Os fatores idade, diabetes, escolaridade, estado civil e número de parceiros foram analisados para determinar sua taxa de risco por meio do modelo de riscos proporcionais de Cox. O estudo identificou que os fatores que têm maior efeito na sobrevida do paciente são a idade e o diabetes. O modelo de Cox permite o uso de informações de sujeitos perdidos no seguimento, como nas análises de sobrevida, além de levar em consideração todos os fatores de risco que podem interferir em um determinado desfecho.

Palavras-chave: aplicação de parâmetros de validação, modelo de Cox, razão de risco, sobrevivência.

Fecha Recepción: Noviembre 2020 **Fecha Aceptación:** Mayo 2021

Introduction

Cervical cancer (Cacu) is the fourth most common neoplasm in women, with an estimated 570,000 new cases in 2018, representing 6.6% of all female cancers. In fact, approximately 90% of cervical cancer deaths occur in low- and middle-income countries (Donatus et al., 2019). For example, in Latin America, Cacu is the second most common neoplasm in women, with 68,818 annual cases, while in Mexico it is the second cause of cancer death in women. Annually, an occurrence of 13,960 cases is estimated in women, with an incidence of 23.3 cases per 100,000 women. In 2013, in the specific group of women aged 25 years and over, 3,771 deaths were registered in women, with a rate of 11.3 deaths per 100,000 women (Government of Mexico, 2019a).

In the state of Chihuahua, the mortality rate from this disease was 12.2 in 2013, according to data from the National Institute of Statistics and Geography (Inegi) (Government of Mexico, 2019b). In addition, it stands out that in the aforementioned state, malignant tumors presented a high incidence, being the fourth cause of death, with a total of 2338 deaths, which represents 10.5% of the deaths of the entity with a rate of 68.93 per 100,000 inhabitants . In 2009, in Cd. Juárez, of the 725 deaths due to malignant tumors among men and women, 9.25% corresponded to the Cacu (Ministry of Health, 2017).

However, the high cervical cancer mortality rate worldwide could be reduced with a comprehensive approach that includes prevention, early diagnosis, screening programs, and effective treatment (Donatus et al., 2019). In this sense, it can be indicated that the survival analysis and the different parametric or non-parametric statistical models contribute to decision-making during medical care.

However, in a survival analysis — apart from the event of interest, which is time — it is important to identify how the parameters within this analysis provide information on the comparison of groups and characteristics that identify groups at risk.

Survival

Fernández (1995) points out that survival is related to the measurement of time to a response, failure, death, relapse or development of a certain disease or event. The term survival is due to the fact that in the first applications of this analysis method the death of a patient was used as an event.

Survival analysis

Velasco (2016) defines the concept of survival analysis as an inferential technique whose main objective is to examine and model the time it takes for a certain event to occur. Generally, given the innumerable applications of these techniques in the biomedical field, the event is commonly called death and time as life time. This analysis consists of a continuous follow-up of a series of individuals from the beginning of the study until the end of the study.

According to Fernández (1995), survival analysis is a statistical area in which the response variable is the time that elapses between an initial event (which determines the inclusion of the individual in the study) and a final event (generically called failure) that occurs when the individual presents the characteristic to finish the study (death, discharge from the disease, etc.).

Statistical methods in survival analysis are similar to those used in other areas where censored data are not presented; for example, descriptive analysis, population comparison, regression-type models, etc. (Fernández, 1995).

On the other hand, Gómez and Cobo (2004) explain that in the process of obtaining empirical evidence on which clinical decisions are based, survival analysis addresses the three classic situations that resolve statistical inference: univariate study, bivariate study and multivariate study.

In a survival analysis, two variables are needed: the time of follow-up and a variable that indicates whether the event of failure or death occurred; Likewise, for the analysis and handling of the data we can use parametric and non-parametric techniques. In this regard, it is important to highlight - as Flores-Flores (2011) points out - that the survival time and clinical characteristics must be observed in each patient (these are called covariates). If we are interested in determining the effect of these on survival time, then the study focuses on the analysis of the relationships between survival time and covariates using a regression model.

Reyes-Méndez, Rodríguez-Medina, Reyes-Uribe and Terrazas-Mata (2019) define the survival function as the probability that a person survives (without experiencing the event of interest) at least until time t. A more formal definition is shown in equation 1.

$$S(t) = Pr[T > t] = 1 - F(x) = \int_{t}^{\infty} f(x)dx \quad (1)$$

However, survival analyzes are not limited to biosanitary or medical application studies; In this research, the parameters that must be observed and interpreted in an analysis using the Cox regression model are explained in a simple way, and how the result describes the characteristics of the analyzed population.

Cox proportional hazards model

Within survival analysis, in different disciplines, one of the most widely used statistical models is Cox's proportional hazards. This is a semi-parametric model that allows us to describe the effect of covariates on survival. In addition, it extends the methods of analysis of this to simultaneously evaluate the effect of several risk factors over time and examines how they influence the risk rate of an event. This technique works for both quantitative predictor variables and categorical variables. It also allows us to examine how factors influence the risk rate of an event (Reyes-Méndez et al., 2019).

The Cox model is the survival analysis equivalent of the linear regression model. It is a semi-parametric model, since it does not require any form for the risk function (it can be increasing or decreasing), but, instead, it defines a parameter that is the ratio between both risk functions (from the English hazard ratio). This approach responds in a natural way to the question of the clinician, who does not want to know the exact life span of a given patient (part of the non-parameterized model) and focuses his efforts on identifying the interventions that can increase it (part of the parameterized model) (Gómez and Cobo, 2004). The Cox model does not impose a function as a premise for the risk function, but it does assume that the risk ratio takes the same value throughout the monitoring period.

In the model, the hazard rate function of the failure time of a model with a vector of covariates given by X is defined by the equation (2) (Boj del Val, 2017):

$$h(t_i, X_i) = h(t_i, X_{i1}, \dots, X_{ip}) = h_0(t_i) e^{\sum_{j=1}^{p} \beta_j X_{ij}} \quad (2)$$

The following hypotheses can be raised in a survival analysis through the Cox model, which help to identify in a simple and practical way the difference between the analyzed groups, identifying the determining characteristics in a comprehensive way, so they can be considered for prevention and detection programs, as well as effective treatments if applicable, if there is a difference in survival times between two groups analyzed and if subsets of covariates help to explain survival time.

The Wald test, likelihood ratio and log-rank test parameters help to answer these questions according to the significance of the model and the verification of the assumptions of the Cox model that are explained in detail in this work.

Materials and methods

The present investigation conforms to the provisions of the Declaration of Helsinki of 1995 revised in World Medical Association (2000). In addition, all the ethical guidelines required to conduct research in humans were followed, including adherence to the legal requirements of Mexico.

The analysis was carried out with the information obtained from the standardized format used by the Ministry of Health in all its dependencies within the Cervical Cancer Prevention and Control program (Cacu). This format is called Cervical cytology request and report, which observes categories such as identification of the registration unit, patient identification, HPV capture, history observed in cytology, gyneco-obstetric situation, beginning of sexual life, result of cytology, among others that are not mentioned because they lack relevance for this analysis. In addition to this format, information from the registered medical history was considered, as well as the biopsy performed to confirm the diagnosis of cervical cancer.

These data correspond to the records of the clinical history, cervical cytology report and biopsy of the 127 women who were detected, registered and treated with this disease through the colposcopy clinic of the Health Jurisdiction II during the period 2013-2017. The files with this information are active in said agency.

The study presents prognostic factors that can be considered to have a greater effect on the survival of patients with Cacu that affect women who receive care from the Sanitary Jurisdiction II of Cd. Juárez Chih., Their risk rate and verification of the assumptions that verify the proportional hazards model.

The R-program software was used to manage the information and to program the Cox model in order to determine that the data submitted met the parameters of the proportional hazards model assumption, where the relative risk or risk ratio (HR) of an individual from the alternative group with respect to an individual from the reference group is assumed to be constant over time (Gómez, October 3, 2012).

In the output of the R-program for this model, the risk rate (HR), the β coefficients and their significance within the model were identified, verified through the p-value. In this regard, Reyes-Méndez et al. (2019) explain that if a risk rate is greater than one (HR> 1), the risk increases. If HR <1, the risk is reduced. If HR = 1, then there is no effect. In cancer studies, if there is a covariate with HR> 1, it is called a poor prognostic factor, and HR <1 is called a good prognostic factor.

Contrast evaluation for parameter verification

There are three tests that are applied to the Cox model in order to validate the assumptions of the model: the likelihood ratio test, the Wald test, and the log-rank test.

1. The likelihood ratio contrast allowed verifying that the covariates (in this study called factors) are independent of each other (Reyes-Méndez et al., 2019).
2. The Wald test is used to determine the true value of the parameters based on the sample estimate. This test, of parametric statistics, is based on the fact that when there is a relationship between or within the data, it is possible to build a model capable of describing the effect of changes in the explanatory variables (probability of the event of interest called success). The Wald test has a more direct interpretation than the plausibility test. However, the maximum likelihood test converges faster towards the normal distribution. When in doubt about which one to use, it is advisable to decide on the maximum likelihood test (Boj del Val, 2017). In summary, it was verified

whether the coefficients β ≠ 0. That is, the variable has multiplicative effects on the resulting Cox model.

3. Log-rank test: This test was used to compare the survival curves of two or more groups (in this case, patients diagnosed with diabetes and also separated by the educational level of the patients). In this test, the null hypothesis is that there is no difference in group survival. It is a nonparametric test, since it does not make assumptions about the distributions. The distribution of the test statistic is approximately chi-square. The test compares the observed number of events in each group against what would be expected if the null hypothesis were true (that is, if the survival curves were identical) (Reyes-Méndez *et al*., 2019).

Results

A first analysis was carried out that we call Cox1. In this study, the covariates diabetes and age were involved, considering the variable diabetes as dichotomous, cataloged with 1 when they do not suffer from it and 0 when they do. Table 1 shows an output from R-program, where the significance of the model can be verified only through the likelihood ratio. Figure 1 presents the schedule.

Figura 1. Programación Cox1 para nivel de significancia

```
survdiff(Surv(datos20$tiempos,datos20$evento)~datos20$diabetes)
cox1<-coxph(Surv(datos20$tiempos,datos20$evento)~datos20$diabetes+datos20$edad)
datos20<-read.csv("CANCER2.CSV",header = T)
> cox1<-coxph(Surv(datos20$tiempos,datos20$evento)~datos20$diabetes+datos20$edad)
> cox1
Call:
coxph(formula = Surv(datos20$tiempos, datos20$evento) ~ datos20$diabetes +
    datos20$edad)
```

Fuente: Elaboración propia

Tabla 1. Cox 1 para nivel de significancia

	Coef	exp(coef)	se(coef)	Z	p
datos20$diabetes	-0.91453	0.40070	0.4709	-1.945	0.0518
datos20$edad	0.02646	1.02681	0.01315	2.012	0.0442
Likelihood ratio test= 8.99					0.01118

Fuente: Elaboración propia

The estimate obtained directly through the output presented in Table 2 is the estimate of the relative risks from the exp (coef), with which we can say that the two variables were significant, and diabetes may not be considered; however, the results of the Cox analysis allow it to be taken into account. Figure 2 presents the programming for Cox1.

Figura 2. Programación Cox1 para estimación de riesgos relativos

```
> cox1<-coxph (Surv(datos20$tiempos, datos20$evento)~datos20$diabetes+datos20$edad)
> summary(cox1)
Call:
coxph(formula = Surv(datos20$tiempos, datos20$evento) ~ datos20$diabetes +
   datos20$edad)
 n= 127, number of events= 32
```

Fuente: Elaboración propia

Tabla 2. Cox1 Estimación de riesgos relativos

	Coef	exp(coef)	se(coef)	z	p
datos20$diabetes	-0.91453	0.40070	.4709	-1.945	0.0518
datos20$edad	0.02646	1.02681	0.01315	2.012	0.0442
Likelihood ratio test= 8.99					0.01
Wald Test= 10.49					0.005
Score(logrank)test= 11.33					0.003

Fuente: Elaboración propia

The presence of not having diabetes reduces the risk by a factor of 0.4007 for this covariate; the β coefficient for diabetes is -0.91453, so its negative sign indicates a decrease in risk; also thus we can verify a value p = 0.05.

Similarly, age with a coefficient β = 0.026 and a value of p = 0.04 represents a risk ratio HR = 1.02, indicating a significant relationship to a higher risk of death, keeping the other covariate constant. Older age is associated with poor survival.

Regarding the contrast of the likelihood ratio $p = 0.01 < 0.05$, it indicates that the factors are independent of each other.

The Wald contrast with $p = 0.005 < 0.05$ indicates that the value of the coefficients is different from zero, and the contrast of the log-rank score concludes that the survival curves of the groups of patients with diabetes differ significantly with $p = 0.003 < 0.05$.

Table 3 presents the verification of the assumptions of the Cox model that are summarized for the Cox1 analysis.

Tabla 3. Verificación de los supuestos del modelo de Cox1

	rho	chisq	p
datos20$diabetes	-0.0222	0.0149	0.903
datos20$edad	0.1218	0.4177	0.518
GLOBAL	NA	0.5025	0.778

Fuente: Elaboración propia

There is no significant evidence at 5% that the assumption is violated that the coef ficient between the risk for two subjects with the same vector of covariates is constant over t ime neither from the global point of view nor for each covariate. Since p values> 0.05 for th e respective factors means that the analyzed covariates are independent of each other.

Also, a second analysis was done, which we will refer to as Cox2. To do this, the following were considered as dichotomous covariates: education (1 if the patient had attended primary or secondary school, and 0 if she had high school or professional); marital status (0 if she w as single and 1 not single), as well as age and number of partners.

Table 4 shows an output from R program, where the significance of the model can be verifi ed only through the likelihood ratio. The estimate of relative risks for the Cox2 analysis is a lso shown with the coxph command. Figure 3 presents the programming for Cox2.

Figura 3. Programación Cox2 para estimación de riesgos relativos

```
> summary(cox2)
Call:
coxph(formula = Surv(datos20$tiempos, datos20$evento) ~ datos20$escolaridad +
   datos20$edad + datos20$parejas + datos20$ecivil)
  n= 127, number of events= 32
```

Fuente: Elaboración propia

Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Tabla 4. Cox2 Estimación de riesgos relativos

	Coef	exp(coef)	se(coef	z	p
datos20$escolaridad	1.30243	3.67822	1.03420	1.259	0.2079
datos20$edad	0.02982	1.03027	0.01454	2.051	0.0403
Datos20$parejas	0.04398	1.04496	0.12549	0.350	0.7260
Datos20$ecivil	0.03863	1.03936	0.39254	0.098	0.9216
Likelihood ratio test= 8.27					0.08
Wald Test=7.04					0.1
Score(logrank)test=7.73					0.1

Fuente: Elaboración propia

The estimate obtained directly through the output presented in table 4 is the estimate of relative risks.

The age variable is the only one that is significant with a p value = 0.04, unlike schooling, number of partners and marital status, which are not significant.

The p = 0.08 value of the maximum likelihood ratio indicates that the influences of these analyzed factors are not of great influence in this model.

Schooling with a β = 1.30 represents a risk ratio HR = 3.67, although this is not significant with p = 0.207, which indicates a significant relationship to a higher risk of death.

Age with a β = 0.02 with a risk ratio HR = 1.03, an additional year of age induces a daily risk of death, although not with a significant contribution given its β coefficient, thus keeping the other variables constant.

Regarding the number of couples, it does not represent significance in the model, since p = 0.72 with a coefficient β = 0.04. Increasing the number of partners by 1 induces a daily risk of death at the risk ratio of 1,044.

The marital status has no significance, since its p-value = 0.92 demonstrates it; the coefficient β = 0.03 and its HR = 1.03 indicate a higher risk of death.

Regarding the maximum likelihood ratio contrast, the value p = 0.08> 0.05 states that the analyzed values may not be independent of each other in general, without forgetting that age is; Likewise, the Wald test presents a value p = 0.1> 0.05, which indicates that the coefficient values can be zero, so they could not influence the model.

The contrast of the score with p = 0.1> 0.05 indicates that the survival curves of the groups with and without schooling do not differ significantly.

Table 5 presents the verification of the assumptions of the Cox model that is summarized for the Cox2 analysis. R-program's cox.zph command returns this summary.

Tabla 5. Verificación de los supuestos del modelo de Cox2

	Rho	Chisq	p
datos20$escolaridad	0.136308	0.6111511	0.434
datos20$edad	0.156627	0612584	0.434
datos20$parejas	0.155342	0.655322	0.418
datos20$ecivil	-0.000863	0.000024	0.996
GLOBAL	NA	1.518387	0.823

Fuente: Elaboración propia

The effect of the covariates of schooling, marital status, age and partners does not vary over time, which means constant risk compliance.

Discussion

Velasco (2016) states that "the Cox model defines the nature of the relationships between the variables considered with respect to cause-effect situations" (p. 13). In this analysis we review the effect of having diabetes or not having diabetes, as well as the level of education, factors that distinguish two groups with the common fact of suffering from Cacu; In addition to highlighting risk indicators, it is verified through the evaluation of contrasts to verify parameters and the assumptions of the model that different conditions of the patients may pose, and the difference in survival times between the groups of women suffering from diabetes is distinguished and those that do not, indicated by the Score log-rank parameter. Likewise, it was verified that the factors are independent of each other through the likelihood ratio.

In this regard, Pérez-Rodríguez, Rivas-Ruiz, Palacios-Cruz and Talavera (2014) indicate:

> Much of the clinical practice related to medical care, decision-making involves knowledge of the clinical course of the disease and the factors that affect it; and estimating the time elapsed until an event occurs. One of the most common examples is when a clinician wants to estimate how long a lung cancer patient might live with or without treatment. The Kaplan-Meier method is used to make this estimate. However, this method does not allow proposing a model that takes into account the different conditions of the patient that could modify the outcome, such as age, clinical stage, etc. Furthermore, it assumes that the groups are homogeneous with respect to all variables; For example, it assumes that all subjects are the same age, which is not necessarily

> the case of patients for whom we want to issue a reliable prognosis that allows us to make an adequate clinical judgment, which is assumed as an advantage in the model of Cox proportional hazards (p. 430).

On the other hand, Rivas-Ruiz, Pérez-Rodríguez, Palacios and Talavera (2014) agree with regard to decision-making:

> During medical care implies knowledge of the clinical course of the disease. The concern for estimating the time that elapses until an event; This has been going on for a long time; Even the first descriptions of such a mathematical exercise were made in other areas of knowledge, such as physics and astronomy.
>
> The knowledge of the clinical evolution of the disease allows estimating the possibility of occurrence of a phenomenon in a given time or its duration. Survival curves allow calculating the probability of a phenomenon occurring at different intervals and, in the same way, allow estimating the median survival of any phenomenon of interest (although the term survival is used, the outcome does not necessarily have to be death, but the occurrence of any other phenomenon (p. 308).

Pérez-Rodríguez *et al*. (2014) they also highlight the reference to a multivariate model:

> For a multivariate model to be considered complete, it must weight the effect of the different variables previously considered, for the prediction of one or more dependent variables, whose outcome is dichotomous, such as time to cure, time to relapse, time for a disease to start, etc. For this case, the Cox proportional hazards model constitutes a multivariate model that can weight the effect of a series of qualitative or quantitative variables on a dichotomous outcome over time, which is easily adapted to situations with incomplete data. (p. 430).

In accordance with the aforementioned, it is emphasized that the Cox molding is a robust model for the analysis of survival and detection of risk indicator variables, which has been applied to various sciences.

Báez, Rodríguez, De la Vega and Tlapa (2013) present a human reliability model for manual workers on assembly lines using the Cox proportional hazard model, with which nine factors were identified in 120 assembly line operators through tests psychometric. Subsequently, stress, motivation, memory and personality factors were identified through a multiple linear regression analysis, as well as those that contribute significantly to the occurrence of human error, which, together, are considered as the operating environment of the employee. The parameters for the base failure distribution were defined to model the human risk rate. The model obtained allows to establish the contribution of each factor to the probability of human errors being committed in a given period of time.

Conte, Domínguez, García Felipe, Rubio and Pérez Prados (2010) address the study of noise in the metal industry, as well as the common presence of chemical and physical pollutants that affect workers through the Cox regression model. In this work, the character of the cause-effect relationships between the variables considered with respect to three situations was defined: healthy / altered; recoverable / non-recoverable; without drops in conversational / with drops in conversational.

García-Bolívar (2012) — in order to evaluate survival times— uses two models in the study of graft mortality in inchi plants. The non-parametric model of Kaplan and Meier and the Cox regression model, which were applied in the description of the behavior of the graft survival time. The time considered to make observations on the presence of failures (dead graft) or not (stuck graft) was 340 days. The factors studied were sex of the donor plant, nitrogen dose and removal time.

Guendelman, Samuels, and Ramírez-Zetina (1999) examine health, labor, and social factors that contribute to giving up work in two transnational maquiladoras of the electronic branch of Tijuana, for which they followed the women until their resignation or the end of the period. observation, which was 2 years.

Gómez-González, Orozco-Hinojosa and Zamudio-Gómez (2006) study the main determinants of the risk rate or conditional probability of default of financial obligations of Colombian private sector firms using a duration model and taking as inputs the variables of the Camel model, where estimates are made by partial maximum likelihood.

Ayala, Borges and Colmenares (2007) use survival analysis as a statistical technique to investigate the probability that a financial entity changes state. In particular, the time in which the change of state occurred and which were the variables that most influenced this process were of interest. A classic survival analysis was carried out, an estimate by the Kaplan and Meier method and an adjustment through the Cox model, using data from the Venezuelan commercial banks between 1996 and 2004. The bank merger is the event of interest for the status change. In this way, it was possible to identify four risk reasons consistent with the theoretical review of the early warning models.

Finally, it is worth mentioning as a weakness of this research the fact that only five of the sociodemographic and clinical factors obtained from the medical history are presented in the evaluation of the model. Even so, it is worth highlighting as a strength that the model considers the survival times between the different categories, since to formulate the risk model for women with cancer it is necessary to identify the censored data and the length of stay in the study. In addition, the limitation in the information on factors considered for treatment are not available in said files. In short, this research focused only on patients with Cacu registered, diagnosed and treated through the colposcopy clinic of the Sanitary Jurisdiction II in Cd. Juárez Chihuahua, during the period 2013 to 2017.

Conclusions

The contrast of the likelihood ratio $p = 0.01 < 0.05$ indicates that the factors are independent of each other for the Cox1 analysis, where age and the fact of having diabetes were evaluated.

The Wald contrast with $p = 0.005 < 0.05$ indicates that the value of the coefficients is different from zero, and the contrast of the log-rank score concludes that the survival curves of the groups of patients with diabetes differ significantly with $p = 0.003 < 0.05$.

The age variable is the only one that is significant with a p value = 0.04, unlike schooling, number of partners and marital status, which are not significant.

The effect of the covariates of schooling, marital status, age and partners does not vary over time, which means constant risk compliance.

The results of this analysis show that the Cox model helped identify the age of the patients as risk factors, as well as the fact that they have diabetes.

According to the results obtained in the Cox1 and Cox2 analyzes, it is verified that in both cases the assumptions of the model are fulfilled, both for each covariate and globally, although in the estimation of relative risks and level of significance they are not important due to the number of related covariates in the Cox2 model that do not influence the model.

If the assumptions made by the Cox proportional hazards model are not met, then it means that the model is not adequate and we could use the stratified Cox model or the proportional odds model and the log-logistic model as an alternative.

Future lines of research

The application of different statistical models for survival analysis opens the door to consider a greater number of variables that help to design more complex models, since it offers the possibility of designing a survival model with the Cox proportional hazards model or - Given the nature of some variables - to be analyzed by means of a logistic regression model that uses the categorical variables that are presented.

The application of survival analysis has spread to multiple areas of knowledge; For example, Bayesian statistics naturally and easily incorporate the treatment of censorship mechanisms and, above all, truncation, allowing the introduction of a priori knowledge (that is, having initial information).

The statistical procedures of this methodology began to be approached from the frequentist inferential framework, hence the use of the Bayesian methodology is recommended, since it provides a theoretical framework that allows any type of uncertainty associated with the problem under study to be quantified in a probabilistic way.

References

Ayala, M. A., Borges, R. E. y Colmenares, G. (2007). Análisis de supervivencia aplicada a la banca comercial venezolana, 1996-2004. *Revista Colombiana Estadística*, *30*(1), 97-113. Recuperado de https://revistas.unal.edu.co/index.php/estad/article/view/29323

Báez, Y. A., Rodríguez, M. A., De la Vega, E. J., Tlapa, D. A. (2013). Factores que influyen en el error humano de los trabajadores en líneas de montaje manual. *Información Tecnológica*, *24*(6), 67-78. Recuperado de https://dx.doi.org/10.4067/S0718-07642013000600010

Boj del Val, E. (2017). *El modelo de regresión de Cox*. Departamento de Matemática Económica, Financiera y Actuarial. Facultad de Economía y Empresa. Universidad de Barcelona. Recuperado de http://diposit.ub.edu/dspace/bitstream/2445/49070/6/El%20modelo%20de%20Cox%20de%20riesgos%20proporcionales.pdf

Conte, J. C., Domínguez, A. I., García Felipe, A. I., Rubio, E. y Pérez Prados, A. (2010). Modelo de regresión de Cox de la pérdida auditiva en trabajadores expuestos a ruido y fluidos de mecanizado o humos metálicos. *Anales del Sistema Sanitario de Navarra*, *33*(1), 11-21. Recuperado de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272010000100002&lng=es&tlng=es

Donatus, L., Nina, F. K., Sama, D. J., Nkfusai, C. N., Bede, F., Shirinde, J. and Cumber, S. N. (2019). Assessing the uptake of cervical cancer screening among women aged 25-65 years in Kumbo West Health District, Cameroon. *The Pan African Medical Journal*, *11*. Doi: 10.11604/pamj.2019.33.106.16975

Fernández, P. (1995). Análisis de supervivencia. *Unidad de Epidemiología Clínica y Bioestadística. Cad Aten Primaria*, 130-135. Recuperado de https://www.fisterra.com/gestor/upload/guias/analisis_supervivencia2.pdf

Flores-Flores C. (2011). *Modelo de regresión de Cox usando splines* (tesis de licenciatura). Universidad Nacional Mayor de San Marcos, Perú. Recuperado de https://www.semanticscholar.org/paper

García-Bolívar, J.J. (2012). Análisis de supervivencia aplicado al estudio. *Revista Científica UDO Agrícola*, *12*(4), 759-769. Recuperado de https://dialnet.unirioja.es/servlet/articulo?codigo=6104314

Gobierno de México (2019a). Recuperado de https://www.gob.mx/salud/acciones-y-programas/información-estadística

Gobierno de México (2019b). Recuperado de https://www.gob.mx/cms/uploads/attachment/file/12951/MortalidadCaCu2000a2013.pdf

Gómez-González, J. E., Orozco-Hinojosa, I. P. y Zamudio-Gómez, N. E. (2006). Análisis de la probabilidad condicional de incumplimiento de los mayores deudores privados del sistema financiero colombiano. *Financiera*, (19), 93-102. Recuperado de https://repositorio.banrep.gov.co/handle/20.500.12134/2104

Gómez, G. (3 de octubre de 2012). *Cuestiones de supervivencia. De la curva de Kaplan-Meier al modelo de Cox.* (video de Youtube). Recuperado de https://www.youtube.com/watch?v=Zcg2GAMB474&list=PLK0kCCQRJcU_oo7RyQztEH5jcDV46ER92

Gómez, G. y Cobo, E. (2004). Hablemos de… análisis de supervivencia. *Gastroenterología y Hepatología Continuada*, *3*(4), 185-191. Recuperado de http://aeeh.es/wp-content/uploads/2012/05/v3n4a203pdf001.pdf

Guendelman, S., Samuels, S. y Ramírez-Zetina, M. (1999). Relación entre salud y renuncia al empleo en trabajadores de la industria maquiladora electrónica de Tijuana. *Salud Pública de México*, *41*(4), 286-296.

Pérez-Rodríguez, M., Rivas-Ruiz, R., Palacios-Cruz, L. y Talavera, J. O. (2014). Investigación clínica XXII. Del Juicio clínico al modelo de riesgos proporcionales de Cox. *Revista Médica Instituto Mexicano del Seguro Social*, *52*(4), 430-435. Recuperado de https://www.redalyc.org/articulo.oa?id=4577/457745483016

Reyes-Méndez, M., Rodríguez-Medina, M., Reyes-Uribe, V. y Terrazas-Mata, L. E. (2019). Selección de parámetros para estimar supervivencia. *Academia Journals*, *11*(5), 1023-1028. Recuperado de https://drive.google.com/drive/folders/1uznIijkpZK2TiC9fWhNozpw-7db1wez6

Rivas-Ruiz, R., Pérez-Rodríguez, M., Palacios, L. y Talavera, J. O. (2014). Investigación clínica XXI. Del juicio clínico al análisis de supervivencia. *Revista Médica Instituto*

Mexicano del Seguro Social, *52*(3), 308-315. Recuperado de http://www.redalyc.org/articulo.oa?id=457745482015

Secretaría de Salud (2017). *Mortalidad de Género 2009*. Centro Nacional de Equidad de Género y Salud Reproductiva. Recuperado de http:cnegsr.salud.gob.mx/contenidos/ProgramasdeAccion/CancerdelaMujer/CaCu/introduccion.html

Velazco, P. (2016). *Modelo de regresión de Cox y sus aplicaciones biosanitarias* (tesis de maestría). Universidad de Sevilla. Recuperado https://idus.us.es/handle/11441/43493

World Medical Association (2000). *Declaration of Helsinki*. Retrieved from https://www.wma.net/wp-content/uploads/2018/07/DoH-Oct2000.pdf

Rol de Contribución	Autor (es)
Conceptualización	Dr. Manuel Arnoldo Rodríguez Medina.
Metodología	M.C Luz Elena Terrazas Mata
Software	M.C Luz Elena Terrazas Mata
Validación	Dr. Manuel Rodríguez Medina
Análisis Formal	Dr. Manuel Rodríguez Medina
Investigación	M.C Luz Elena Terrazas Mata
Recursos	Dr. Alfonso Aldape Alamillo
Curación de datos	M.C Luz Elena Terrazas Mata
Escritura Preparación del borrador original	Dr. Alfonso Aldape Alamillo
Escritura - Revisión y edición	M.C Luz Elena Terrazas Mata
Visualización	Dr. Juan Pedro Benítez Guadarrama
Supervisión	Dr. Juan Pedro Benítez Guadarrama
Administración de Proyectos	M.C Yuliana Meléndez Pastrana
Adquisición de fondos	M.C Luz Elena Terrazas Mata