

<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 Almilló

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



estadística 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

Introducción

El cáncer cervicouterino (Cacu) es la cuarta neoplasia más frecuente en las mujeres, con un estimado de 570 000 nuevos casos en 2018, lo que representa 6.6 % de todos los cánceres femeninos. De hecho, aproximadamente 90 % de las muertes por cáncer de cuello uterino ocurren en países de ingresos bajos y medianos (Donatus *et al.*, 2019). Por ejemplo, en América Latina, el Cacu es la segunda neoplasia más común en mujeres, con 68 818 casos anuales, mientras que en México es la segunda causa de muerte por cáncer en la mujer. Anualmente, se estima una ocurrencia de 13 960 casos en mujeres, con una incidencia de 23.3 casos por 100 000 mujeres. En el año 2013, en el grupo específico de mujeres de 25 años y más, se registraron 3771 defunciones en mujeres con una tasa de 11.3 defunciones por 100 000 mujeres (Gobierno de México, 2019a).

En el estado de Chihuahua, la tasa de mortalidad por esta enfermedad fue de 12.2 en el año 2013, según datos del Instituto Nacional de Estadística y Geografía (Inegi) (Gobierno de México, 2019b). Además, sobresale que en el mencionado estado los tumores malignos presentaron una alta incidencia, siendo la cuarta causa de muerte, con un total de 2338 decesos, lo que representa 10.5 % de los fallecimientos de la entidad con una tasa de 68.93 por 100 000 habitantes. En 2009, en Cd. Juárez, de los 725 decesos por tumores malignos entre hombres y mujeres, 9.25 % correspondió al Cacu (Secretaría de Salud, 2017).

Ahora bien, la alta tasa de mortalidad por cáncer de cuello uterino en todo el mundo podría reducirse con un enfoque integral que incluyera la prevención, el diagnóstico



temprano, los programas de detección y el tratamiento eficaz (Donatus *et al.*, 2019). En tal sentido, se puede indicar que el análisis de supervivencia y los diferentes modelos estadísticos paramétricos o no paramétricos contribuyen a la toma de decisiones durante la atención médica.

Sin embargo, en un análisis de supervivencia —aparte del evento de interés, que es el tiempo— es importante identificar cómo los parámetros dentro de este análisis proporcionan información sobre comparación de grupos y características que identifican a grupos de riesgo.

Supervivencia

Fernández (1995) señala que la supervivencia se relaciona con la medida de tiempo a una respuesta, fallo, muerte, recaída o desarrollo de una determinada enfermedad o evento. El término *supervivencia* se debe a que en las primeras aplicaciones de este método de análisis se utilizaba como evento la muerte de un paciente.

Análisis de supervivencia

Velasco (2016) define el concepto *análisis de supervivencia* como una técnica inferencial cuyo principal objetivo es examinar y modelar el tiempo que se toma para que ocurra un determinado suceso. Generalmente, dadas las innumerables aplicaciones de estas técnicas en el ámbito biomédico, el suceso es comúnmente denominado *muerte* y el tiempo como *tiempo de vida*. Este análisis consiste en un seguimiento continuo de una serie de individuos desde que comienza el estudio hasta que finaliza.

Según Fernández (1995), el análisis de supervivencia es un área estadística en la que la variable *respuesta* es el tiempo que transcurre entre un evento inicial (que determina la inclusión del individuo en el estudio) y un evento final (genéricamente llamado *falla*) que ocurre cuando el individuo presenta la característica para terminar el estudio (muerte, alta de la enfermedad, etc.).

Los métodos estadísticos en el análisis de supervivencia son similares a los que se utilizan en otras áreas donde no se presentan datos censurados; por ejemplo, análisis descriptivo, comparación de poblaciones, modelos tipo regresión, etc. (Fernández, 1995).

Por otro lado, Gómez y Cobo (2004) explican que en el proceso de obtención de pruebas empíricas donde se basan las decisiones clínicas, el análisis de supervivencia aborda

las tres situaciones clásicas que resuelven la inferencia estadística: estudio univariante, estudio bivariante y estudio multivariante.

En un análisis de supervivencia se necesitan dos variables: el tiempo de seguimiento y una variable que indique si sucedió el evento de fallo o muerte; asimismo, para el análisis y manejo de los datos podemos utilizar técnicas paramétricas y no paramétricas. Al respecto, es importante destacar —como apunta Flores-Flores (2011)— que en cada paciente se deben observar el tiempo de supervivencia y sus características clínicas (a estas se les llama *covariables*). Si nos interesa determinar el efecto de estas en el tiempo de supervivencia, entonces el estudio se centra en el análisis de las relaciones entre el tiempo de supervivencia y las covariables mediante un modelo de regresión.

Reyes-Méndez, Rodríguez-Medina, Reyes-Uribe y Terrazas-Mata (2019) definen la función de supervivencia como la probabilidad de que una persona sobreviva (sin que experimente el evento de interés) al menos hasta el tiempo t . Una definición más formal se muestra en la ecuación 1.

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

Sin embargo, los análisis de supervivencia no se limitan a estudios de aplicación biosanitaria o medicina; en esta investigación, se explican de manera sencilla los parámetros que se deben observar e interpretar en un análisis mediante el modelo regresión de Cox, y cómo el resultado describen las características de la población analizada.

Modelo de riesgos proporcionales de Cox

Dentro del análisis de supervivencia, en diferentes disciplinas, uno de los modelos estadísticos más utilizados es el de riesgos proporcionales de Cox. Este es un modelo semiparamétrico que permite describir el efecto de las covariables en la supervivencia. Además, amplía los métodos de análisis de esta para evaluar simultáneamente el efecto de varios factores de riesgo en el tiempo y examina cómo influyen en la tasa de riesgo de un evento. Esta técnica funciona tanto para las variables predictoras cuantitativas como para las variables categóricas. También permite examinar cómo los factores influyen en la tasa de riesgo de un evento (Reyes-Méndez *et al.*, 2019).

El modelo de Cox es el equivalente en análisis de la supervivencia al modelo de regresión lineal. Se trata de un modelo semiparamétrico, ya que no exige ninguna forma para



la función de riesgo (puede ser creciente o decreciente), pero, en cambio, define un parámetro que es la razón entre ambas funciones de riesgo (del inglés *hazard ratio*). Este planteamiento responde de forma natural a la pregunta del clínico, que no ambiciona conocer cuál es el tiempo exacto de vida de un paciente determinado (parte del modelo no parametrizado) y centra sus esfuerzos en identificar las intervenciones que pueden aumentarlo (parte del modelo parametrizado) (Gómez y Cobo, 2004). El modelo de Cox no impone como premisa una función para la función de riesgo, pero sí asume que la razón de riesgo toma el mismo valor durante todo el periodo de seguimiento.

En el modelo, la función de tasa de riesgo del tiempo de falla de un modelo con vector de covariables dado por X está definida en la ecuación (2) (Boj del Val, 2017):

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

Las siguientes hipótesis se pueden plantear en un análisis de supervivencia a través del modelo de Cox, las cuales ayudan a identificar de manera sencilla y práctica la diferencia entre los grupos analizados, identificando las características determinantes de manera integral, por lo que pueden ser consideradas para programas de prevención y de detección, así como tratamientos eficaces si es el caso, si existe diferencia en los tiempos de supervivencia entre dos grupos analizados y si los subconjuntos de covariables ayudan a explicar el tiempo de supervivencia.

Los parámetros test de Wald, razón de verosimilitud y test log-rank ayudan a contestar estas preguntas de acuerdo con la significancia del modelo y la verificación de los supuestos del modelo de Cox que en este trabajo se explican a detalle.

Materiales y métodos

La presente investigación se ajusta a las disposiciones de la Declaración de Helsinki de 1995 revisada en World Medical Association (2000). Además, se siguieron todas las pautas éticas requeridas para realizar investigaciones en humanos, incluida la adherencia a los requisitos legales de México.

El análisis se efectuó con la información obtenida del formato estandarizado que utiliza la Secretaría de Salud en todas sus dependencias dentro del programa de Prevención y Control de Cáncer Cervicouterino (Cacu). Este formato se llama *Solicitud y reporte de citología cervical*, el cual observa categorías tales como identificación de la unidad de registro, identificación del paciente, captura de VPH, antecedentes que se observan en la

citología, situación ginecobstétrica, inicio de vida sexual, resultado de la citología, entre otras que no se mencionan por carecer de trascendencia para este análisis. Además de este formato, se consideró información de la historia clínica registrada, agregada también la biopsia que se realiza para confirmar el diagnóstico de cáncer cervicouterino.

Estos datos corresponden a los registros de la historia clínica, reporte de citología cervical y biopsia de las 127 mujeres que se detectaron, registraron y atendieron con esta enfermedad a través de la clínica de colposcopia de la Jurisdicción Sanitaria II durante el periodo 2013-2017. Los expedientes con esta información se encuentran activos en dicha dependencia.

El estudio presenta factores pronósticos que pueden considerarse de mayor efecto en la supervivencia de pacientes con Cacú que aquejan a las mujeres que reciben atención por parte de la Jurisdicción Sanitaria II de Cd. Juárez Chih., su tasa de riesgo y comprobación de los supuestos que verifican el modelo de riesgos proporcionales.

Se utilizó el *software* R-program para el manejo de la información y para programar el modelo de Cox con el fin de determinar que los datos sometidos cumplieran con los parámetros del supuesto del modelo de riesgos proporcionales, donde el riesgo relativo o razón de riesgo (HR) de un individuo del grupo alternativo respecto a un individuo del grupo de referencia se supone constante a lo largo del tiempo (Gómez, 3 de octubre de 2012).

En la salida del R-program para este modelo se identificaron la tasa de riesgo (HR), los coeficientes β y la significancia de estos dentro del modelo, verificados a través del *p-value*. Al respecto, Reyes-Méndez *et al.* (2019) explican que si una tasa de riesgo es mayor que uno ($HR > 1$), se incrementa el riesgo. Si $HR < 1$, el riesgo se reduce. Si $HR = 1$, entonces no hay efecto. En los estudios de cáncer, si existe una covariable con $HR > 1$, se le denomina *factor de mal pronóstico*, y $HR < 1$ es nombrado *factor de buen pronóstico*.

Evaluación de contrastes para verificación de parámetros

Existen tres pruebas que se aplican al modelo de Cox con el fin de validar los supuestos del modelo: el contraste de la razón de verosimilitud, el contraste de Wald y la prueba log-rank.

1. El contraste de la razón de verosimilitud permitió verificar que las covariables (en este estudio llamadas *factores*) son independientes una de otra (Reyes-Méndez *et al.*, 2019).



2. La prueba de Wald se utiliza para poder determinar el verdadero valor de los parámetros basándose en la estimación de la muestra. Esta prueba, de estadística paramétrica, se basa en que cuando existe una relación entre o dentro de los datos es posible construir un modelo capaz de describir el efecto de los cambios de las variables explicativas (probabilidad del suceso de interés denominado *éxito*). El contraste de Wald tiene una interpretación más directa que el contraste de verosimilitud. Sin embargo, el test de la máxima verosimilitud converge más rápido hacia la distribución normal. Ante la duda de cuál utilizar, es recomendable decidirse por el test de la máxima verosimilitud (Boj del Val, 2017). En resumen, se verificó si los coeficientes $\beta \neq 0$. Es decir, que la variable tiene efectos multiplicativos sobre el modelo de Cox resultante.
3. Log-rank test: Esta prueba se utilizó para comparar las curvas de supervivencia de dos o más grupos (en este caso, los pacientes diagnosticados con diabetes y separados también por la escolaridad de las pacientes). En esta prueba, la hipótesis nula es que no hay diferencia en la supervivencia de los grupos. Es una prueba no paramétrica, ya que no hace suposiciones acerca de las distribuciones. La distribución del estadístico de prueba es aproximadamente chi-cuadrada. La prueba compara el número observado de eventos en cada grupo contra lo que se esperaría si la hipótesis nula fuera cierta (es decir, si las curvas de supervivencia fueran idénticas) (Reyes-Méndez *et al.*, 2019).

Resultados

Se efectuó un primer análisis que llamamos *Cox1*. En este se involucraron las covariables *diabetes* y *edad*, considerando la variable *diabetes* como dicotómica, catalogada con 1 cuando no la padece y 0 cuando sí la padece. En la tabla 1 se muestra una salida de R-program, en donde la significancia del modelo puede verificarse solo a través de la razón de verosimilitud. La figura 1 presenta la programación.

Figura 1. Programación Cox1 para nivel de significancia

```
surv diff(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

Fuente: Elaboración propia

	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

La estimación obtenida directamente a través de la salida presentada en la tabla 2 es la estimación de los riesgos relativos a partir de los exp(coef), con lo que podemos decir que las dos variables resultaron significativas, pudiendo no considerar la diabetes; sin embargo, los resultados del análisis de Cox permiten tomarla en cuenta. La figura 2 presenta la programación para 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

La presencia de no padecer diabetes reduce el riesgo en un factor del 0.4007 para esta covariable; el coeficiente β para diabetes es -0.91453, por lo que su signo negativo indica una disminución en el riesgo; también así podemos verificar un valor $p = 0.05$.

De manera similar, la edad con un coeficiente $\beta = 0.026$ y un valor $p = 0.04$ representa una relación de riesgo $HR = 1.02$, indica una relación significativa a un mayor riesgo de muerte, manteniendo constante la otra covariable. Una mayor edad se asocia con mala supervivencia.

Respecto al contraste de la razón de verosimilitud $p = 0.01 < 0.05$, indica que los factores son independientes uno de otro.

El contraste de Wald con $p = 0.005 < 0.05$ indica que el valor de los coeficientes es diferente de cero, y el contraste del “score” log-rank concluye que las curvas de supervivencia de los grupos de pacientes con diabetes difieren significativamente con $p = 0.003 < 0.05$.

La tabla 3 presenta la verificación de los supuestos del modelo de Cox que se resumen para el análisis Cox1.

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

No existe evidencia significativa a 5 % de que se viole el supuesto de que el coeficiente entre el riesgo para dos sujetos con el mismo vector de covariables es constante en el tiempo ni desde el punto de vista global ni para cada covariable. Dado que los valores $p > 0.05$ para los respectivos factores significa que las covariables analizadas son independientes entre sí.

Asimismo, se hizo un segundo análisis, al cual nos referiremos como Cox2. Para ello, se consideraron como covariables dicotómicas las siguientes: *escolaridad* (1 si la paciente había cursado primaria o secundaria, y 0 si tenía preparatoria o profesional); *estado civil* (0 si era soltera y 1 no soltera), así como *edad* y *cantidad de parejas*.

En la tabla 4 se muestra una salida de R program, en donde la significancia del modelo puede verificarse solo a través de la razón de verosimilitud. También se muestra con el comando *coxph* la estimación de riesgos relativos para el análisis Cox2. La figura 3 presenta la programación para 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

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

La estimación obtenida directamente a través de la salida presentada en la tabla 4 es la estimación de los riesgos relativos.

La variable *edad* es la única que se presenta significativa con un valor $p = 0.04$, a diferencia de *escolaridad*, *cantidad de parejas* y *estado civil*, que no se muestran significantes.

El valor $p = 0.08$ de la razón de máxima verosimilitud indica que las influencias de dichos factores analizados no son de gran influencia en este modelo.

La escolaridad con un $\beta = 1.30$ representa una relación de riesgo $HR = 3.67$, aunque esta no sea significativa con $p = 0.207$, que indica una relación significativa a un mayor riesgo de muerte.

La edad con un $\beta = 0.02$ con una razón de riesgo $HR = 1.03$, un año adicional de edad induce un riesgo diario de muerte, aunque no con una contribución significativa dado su coeficiente β , así manteniendo constantes las otras variables.

Respecto a la cantidad de parejas, no representa significancia en el modelo, dado que $p = 0.72$ con un coeficiente $\beta = 0.04$. Al incrementar en 1 la cantidad de parejas induce un riesgo diario de muerte a la razón de riesgo de 1.044.

El estado civil no tiene significancia, dado que su valor $p = 0.92$ así lo demuestra; el coeficiente $\beta = 0.03$ y su $HR = 1.03$ indican un mayor riesgo de muerte.

Respecto al contraste de razón de máxima verosimilitud, el valor $p = 0.08 > 0.05$ expone que los valores analizados pudieran no ser independientes uno de otro en forma general, sin olvidar que la edad sí lo es; asimismo, el contraste de Wald presenta un valor $p = 0.1 > 0.05$, lo que nos indica que los valores de los coeficientes pueden ser cero, por lo que podrían no influir en el modelo.

El contraste del score con $p = 0.1 > 0.05$ indica que las curvas de supervivencia de los grupos con escolaridad y sin ella no difieren de manera significativa.

La tabla 5 presenta la verificación de los supuestos del modelo de Cox que se resume para el análisis Cox2. El comando `cox.zph` de R-program arroja este resumen.

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	0.612584	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

El efecto de las covariables *escolaridad*, *estado civil*, *edad* y *parejas* no varía en el tiempo, lo que significa el cumplimiento de riesgo constante.

Discusión

Velasco (2016) afirma que “el modelo de Cox define el carácter de las relaciones existentes entre las variables consideradas con respecto a las situaciones causa-efecto” (p. 13). En este análisis revisamos el efecto de tener diabetes o no padecer diabetes, así como el nivel de la escolaridad, factores que distinguen a dos grupos con el hecho común de padecer Cáncer; además de resaltar indicadores de riesgo, se verifica mediante la evaluación de contrastes para verificación de parámetros y los supuestos del modelo que pueden plantear distintas condiciones de las pacientes, y se distingue la diferencia en los tiempos de supervivencia entre los grupos de mujeres que padecen diabetes y las que no, indicado por el parámetro Score log-rank. Asimismo, se verificó que los factores son independientes unos de otros a través de la razón de verosimilitud.

Al respecto, Pérez-Rodríguez, Rivas-Ruiz, Palacios-Cruz y Talavera (2014) indican:

Gran parte del ejercicio clínico relacionado con la atención médica la toma de decisiones implica el conocimiento del curso clínico de la enfermedad y los factores que la afectan; y la estimación del tiempo transcurrido hasta que un evento suceda. Uno de los ejemplos más comunes es cuando un clínico desea estimar cuánto tiempo podría vivir un paciente con cáncer de pulmón con o sin tratamiento. Para realizar esta estimación se usa el método de Kaplan-Meier. Sin embargo, este método no permite plantear un modelo que tome en cuenta las distintas condiciones del paciente que pudieran modificar el desenlace, como la edad, la etapa clínica, etc. Además, asume que los grupos son homogéneos con respecto a todas las variables; por ejemplo, asume que todos los sujetos tienen la misma edad, lo cual no es necesariamente el caso de los pacientes de los cuales queremos emitir un pronóstico confiable que nos permita realizar un juicio clínico adecuado, lo que se asume como ventaja en el modelo de riesgos proporcionales de Cox (p. 430).

Por otro lado, Rivas-Ruiz, Pérez-Rodríguez, Palacios y Talavera (2014) están de acuerdo en lo referente a la toma de decisiones:

Durante la atención médica implica el conocimiento del curso clínico de la enfermedad. La preocupación por estimar el tiempo que transcurre hasta un evento; esto se ha dado desde mucho tiempo atrás; incluso, las primeras

descripciones de tal ejercicio matemático se realizaron en otras áreas del conocimiento, como la física y la astronomía.

El conocimiento de la evolución clínica de la enfermedad permite estimar la posibilidad de ocurrencia de un fenómeno en un tiempo determinado o su tiempo de duración. Las curvas de supervivencia permiten calcular la probabilidad de que ocurra un fenómeno a distintos intervalos y, de igual forma, permiten estimar la mediana de supervivencia de cualquier fenómeno de interés (aunque se utiliza el término supervivencia, el desenlace no necesariamente tiene que ser muerte, sino la ocurrencia de cualquier otro fenómeno (p. 308).

Pérez-Rodríguez *et al.* (2014) también destacan lo referente a un modelo multivariado:

Para que un modelo multivariado sea considerado completo debe ponderar el efecto de las distintas variables previamente consideradas, para la predicción de una o más variables dependientes, cuyo desenlace es dicotómico, como tiempo de curación, tiempo de recaída, tiempo para que una enfermedad inicie, etc. Para este caso, el modelo de riesgos proporcionales de Cox constituye un modelo multivariado que puede ponderar el efecto de una serie de variables cualitativas o cuantitativas sobre un desenlace dicotómico a través del tiempo, que se adapta fácilmente a situaciones con datos incompletos (p. 430).

De acuerdo con lo mencionado anteriormente, se hace énfasis en que el moldeo de Cox es un modelo robusto para el análisis de supervivencia y detección de variables indicadoras de riesgo, el cual ha sido aplicado a diversas ciencias.

Báez, Rodríguez, De la Vega y Tlapa (2013) presentan un modelo de confiabilidad humana para trabajadores manuales en líneas de montaje utilizando el modelo de riesgo proporcional de Cox, con el cual se identificaron nueve factores en 120 operadores de líneas de montaje mediante pruebas psicométricas. Posteriormente, se identificaron los factores de estrés, motivación, memoria y personalidad, mediante un análisis de regresión lineal múltiple, así como aquellos que contribuyen de forma significativa a la ocurrencia del error humano, que, en conjunto, se consideran como el entorno operativo del trabajador. Se definieron los parámetros para la distribución de fallas base para modelar la tasa de riesgo

humano. El modelo obtenido permite establecer la contribución de cada factor a la probabilidad de que se cometan errores humanos en un periodo de tiempo determinado.

Conte, Domínguez, García Felipe, Rubio y Pérez Prados (2010) abordan el estudio del ruido en la industria del metal, así como la presencia común de contaminantes químicos y físicos que afectan a los trabajadores por medio del modelo de regresión de Cox. En dicho trabajo se definió el carácter de las relaciones causa-efecto entre las variables consideradas con respecto a tres situaciones: sano/alterado; recuperable/no recuperable; sin caídas en conversacionales/con caídas en conversacionales.

García-Bolívar (2012) —con el fin de evaluar tiempos de supervivencia— utiliza dos modelos en el estudio de la mortalidad del injerto en plantas de inchi. El modelo no paramétrico de Kaplan y Meier y el modelo de regresión de Cox, los cuales fueron aplicados en la descripción del comportamiento del tiempo de supervivencia del injerto. El tiempo considerado para realizar observaciones sobre la presencia de fallas (injerto muerto) o no (injerto pegado) fue de 340 días. Los factores estudiados fueron sexo de la planta donadora, dosis de nitrógeno y tiempo de remoción.

Guendelman, Samuels y Ramírez-Zetina (1999) examinan factores de salud, laborales y sociales que contribuyen a renunciar al trabajo en dos maquiladoras transnacionales del ramo electrónico de Tijuana, para lo cual efectuaron el seguimiento de las mujeres hasta su renuncia o final del periodo de observación, que fue de 2 años.

Gómez-González, Orozco-Hinojosa y Zamudio-Gómez (2006) estudian los principales determinantes de la tasa de riesgo o probabilidad condicional de incumplimiento de las obligaciones financieras de las firmas del sector privado colombiano usando un modelo de duración y tomando como insumos las variables del modelo Camel, en donde se hace estimaciones por máxima verosimilitud parcial.

Ayala, Borges y Colmenares (2007) utilizan el análisis de supervivencia como técnica estadística para investigar la probabilidad de que una entidad financiera cambie de estado. En particular, interesó el tiempo en que ocurrió el cambio de estado y cuáles fueron las variables que más influyeron en dicho proceso. Se hizo un análisis de supervivencia clásico, una estimación por el método de Kaplan y Meier y un ajuste a través del modelo de Cox, utilizando los datos de la banca comercial venezolana entre 1996 y 2004. La fusión bancaria es el evento de interés para el cambio de estado. De este modo se lograron identificar cuatro razones de riesgo consistentes con la revisión teórica de los modelos de alerta temprana.

Finalmente, se puede mencionar como debilidad de esta investigación el hecho de que solo se presentan en la evaluación del modelo cinco de los factores sociodemográficos y clínicos obtenidos del historial médico. Aun así, vale destacar como fortaleza que el modelo considera los tiempos de supervivencia entre las diferentes categorías, ya que para formular el modelo de riesgo para las mujeres que padecen cáncer es necesario identificar los datos censurados y el tiempo de permanencia en el estudio. Además, la limitante en la información sobre factores considerados de tratamiento no se encuentran disponibles en dichos expedientes. En definitiva, esta investigación se concentró únicamente en las pacientes con Cacú registradas, diagnosticadas y atendidas a través de la clínica de colposcopia de la Jurisdicción Sanitaria II en Cd. Juárez Chihuahua, durante el periodo 2013 a 2017.

Conclusiones

El contraste de la razón de verosimilitud $p = 0.01 < 0.05$ indica que los factores son independientes unos de otros para el análisis Cox1, donde se evaluó la edad y el hecho de padecer diabetes.

El contraste de Wald con $p = 0.005 < 0.05$ indica que el valor de los coeficientes es diferente de cero, y el contraste del “score” log-rank concluye que las curvas de supervivencia de los grupos de pacientes con diabetes difieren significativamente con $p = 0.003 < 0.05$.

La variable *edad* es la única que se presenta significativa con un valor $p = 0.04$, a diferencia de *escolaridad*, *cantidad de parejas* y *estado civil*, que no se muestran significantes.

El efecto de las covariables *escolaridad*, *estado civil*, *edad* y *parejas* no varía en el tiempo, lo que significa el cumplimiento de riesgo constante.

Los resultados de este análisis demuestran que el modelo de Cox ayudó a identificar como factores de riesgo la edad de las pacientes, así como el hecho de que padezcan diabetes.

Según los resultados conseguidos en los análisis Cox1 y Cox2, se verifica que en los dos casos se cumplen los supuestos del modelo, tanto para cada covariable como de manera global, aunque en la estimación de riesgos relativos y nivel de significancia no resultan importantes debido a la cantidad de covariables relacionadas en el modelo Cox2 que no influyen en el modelo.

Si los supuestos que plantea el modelo de riesgos proporcionales de Cox no se cumplen, entonces significa que el modelo no es el adecuado y podríamos utilizar como

alternativa el modelo de Cox estratificado o el modelo odds proporcional y el modelo log-logístico.

Futuras líneas de investigación

La aplicación de diferentes modelos estadísticos para el análisis de supervivencia abre la puerta para considerar un mayor número de variables que ayuden a diseñar modelos más complejos, ya que se ofrece la posibilidad de diseñar un modelo de supervivencia con el modelo de riesgos proporcionales de Cox o —dada la naturaleza de algunas variables— de analizar mediante un modelo de regresión logística que utilice las variables categóricas que se presenten.

La aplicación de los análisis de supervivencia se ha extendido a múltiples áreas del conocimiento; por ejemplo, la estadística bayesiana incorpora de forma natural y sencilla el tratamiento de mecanismos de censura y sobre todo truncamiento, permitiendo introducir conocimiento *a priori* (es decir, que tiene información de inicio).

Los procedimientos estadísticos propios de esta metodología empezaron a abordarse desde el marco inferencial frecuentista, de ahí que se recomiende la utilización de la metodología bayesiana, ya que proporciona un marco teórico que permite cuantificar de forma probabilística cualquier tipo de incertidumbre asociada al problema objeto de estudio.

Referencias

- 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_o07RyQztEH5jcDV46ER92
- 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://cneqsr.salud.gob.mx/contenidos/ProgramasdeAccion/CancerdeMujer/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

