

Medicent Electrón. 2015 oct.-dic.;19(4)

CARDIOCENTRO
«ERNESTO CHE GUEVARA»
SANTA CLARA, VILLA CLARA

ARTÍCULO ORIGINAL

Validación interna de modelo predictivo creado mediante nueva metodología aplicable en la atención primaria de salud

Internal validation of a predictive model created through a new methodology applicable in primary health care

Dra. Vielka González Ferrer¹, Dra. C. Milagros Alegret Rodríguez², Dra. Yainedy González Ferrer³, Dr. Adrián Moreno Arias⁴

1. Especialista de Primer Grado en Bioestadística. Instructora. Aspirante a Doctor en Ciencias de la Salud. Cardiocentro Ernesto Che Guevara. Santa Clara, Villa Clara. Cuba. Correo electrónico: vielkaqf@capiro.vcl.sld.cu
2. Doctora en Ciencias de la Salud. Profesora Titular. Centro Provincial de Higiene y Epidemiología. Santa Clara, Villa Clara. Cuba. Correo electrónico: malegret@capiro.vcl.sld.cu
3. Especialista de Primer Grado en Estomatología General Integral. Clínica Celia Sánchez Manduley. Santa Clara, Villa Clara. Cuba. Correo electrónico: yainedy75@yahoo.es
4. Especialista de Primer Grado en Ortodoncia. Instructor. Policlínico docente Octavio de la Concepción y la Pedraja. Camajuaní, Villa Clara. Cuba. Correo electrónico: adrianma@ucm.vcl.sld.cu

RESUMEN

Introducción: los modelos predictivos sirven de apoyo a la toma de decisiones en salud pública. Como parte del desarrollo de estos modelos, se debe contar con alguna forma de validación interna que permita cuantificar el optimismo en su desempeño predictivo. Para esta validación, se utiliza el mismo grupo de estudio empleado para su desarrollo y los resultados son reproducibles a la población subyacente.

Objetivo: validar un índice de necesidad de tratamiento ortodóntico, creado mediante una metodología que utiliza, para construir el modelo multivariante, los valores del estadígrafo *V de Cramer* de cada predictor.

Métodos: el modelo creado con la muestra de entrenamiento, se aplicó a 181 estudiantes de una escuela primaria de Santa Clara y se calcularon medidas del desempeño discriminatorio; estas fueron: área bajo la curva *Receiver Operating Characteristic* y parámetros calculados a partir de las matrices de confusión. Fueron comparados los modelos obtenidos mediante el nuevo método y la regresión logística.

Resultados: el nuevo modelo superó en todos los parámetros calculados a la regresión logística, con valores de sensibilidad, especificidad y validez de 79,3 %, 84,3 % y 81,2 %, respectivamente. El área bajo la curva fue de 0,886.

218

Conclusiones: estos resultados avalan el índice obtenido mediante V de *Cramer*, para su utilización en la población diana subyacente. La facilidad de cálculo y comprensión de esta metodología son argumentos a favor de su uso por decisores del sector en la atención primaria de salud.

DeCS: atención primaria de salud, índice de necesidad de tratamiento ortodóncico, valor predictivo de las pruebas.

ABSTRACT

Introduction: predictive models are support tools when it comes to decision making in public health. We should count on a specific form of internal validation, as a part of the development of these models, which allows us to quantify any optimism in their predictive performance. For this validation, the same group of study employed for its performance is used, and results are reproducible to the underlying population.

Objective: to validate an index of orthodontic treatment need, created by means of a methodology, that uses the values of Cramer's V of each predictor in order to build the multivariate model.

Methods: the model created with the training sample was applied to 181 students from a primary school of Santa Clara, and measures of discriminatory performance were calculated, such as, area under the receiver operating characteristic curve, as well as, parameters were calculated from the confusion matrices. Models obtained by means of the new method and the logistic regression were also compared.

Results: the new model exceeds logistic regression in all calculated parameters with values of sensitivity, specificity and validity of 79,3 %, 84,3 % and 81,2 %, respectively. Area under the curve was of 0,886.

Conclusions: these results support the obtained index through Cramer ' V in order to be used in the underlying target population. The easiness of calculation and comprehension of this methodology are arguments in favor of its use for health decision - makers in primary care.

DeCS: index of orthodontic treatment need, predictive value of tests, primary health care.

INTRODUCCIÓN

Los modelos o índices predictivos tienen varias aplicaciones en salud pública. Mediante estas herramientas se pueden identificar individuos con alto riesgo de presentar una enfermedad, información que puede ser útil para incluirlos en tamizajes intensivos y detectar esta precozmente, para focalizar intervenciones preventivas,¹ o bien, para priorizar servicios de salud.

El resultado de aplicar un modelo predictivo a un grupo de individuos pueden ser valores continuos que permiten predecir, a partir de la definición de un punto de corte, la pertenencia de los sujetos a uno de los grupos preestablecidos. Esto permite que se obtenga una regla de decisión, o sea, si el resultado de la aplicación del índice a un sujeto en particular supera el punto de corte, indicará la presencia del resultado de interés. Para su desarrollo se han utilizado varios enfoques metodológicos; el estándar utilizado en salud para resolver problemas binarios de clasificación es la regresión logística. Como requisito previo a la aplicación de un nuevo modelo, creado bajo cualquier enfoque, se encuentra su validación, para lo cual se pueden emplear bases de datos externas o internas.²

El interés fundamental para realizar una validación externa es utilizar el modelo en poblaciones diferentes a las empleadas para crearlo. Pero si el objetivo es crear un modelo demandado para resolver un problema de clasificación de una población específica, y el interés no radica en generalizarlo a otras áreas geográficas, la validación interna resultará ser la tarea primordial de este proceso de creación.^{3,4}

Cuantificar la habilidad predictiva de un modelo sobre los mismos datos a partir de los cuales este fue desarrollado, tiende a dar un estimado optimista de su desempeño, debido al sobreajuste (muy pocos resultados del evento con relación al número de predictores candidatos.⁵ La validación interna permite estimar el potencial del sobreajuste y optimismo en el desempeño del modelo,^{6,7} y consiste en no utilizar otros datos que los del grupo de estudio,² para lo cual se utilizan varias técnicas.

En este artículo, se describen los resultados del proceso de validación interna de un índice de necesidad de tratamiento ortodóntico, creado mediante una metodología que consiste en utilizar los valores del estadígrafo V de *Cramer* de cada predictor incluido en el modelo, para conformar una ecuación multivariante.⁸ De manera que el objetivo principal de este artículo es darle respuesta a la siguiente interrogante: ¿Es válida la utilización del modelo construido mediante la nueva metodología?

MÉTODOS

Los datos para desarrollar el índice de necesidad de tratamiento ortodóntico provienen de un estudio transversal, realizado en 371 estudiantes seleccionados por medio de un muestreo estratificado que incluyó a todas las escuelas primarias, secundarias y preuniversitarios de Villa Clara, durante el curso escolar 2011-2012.⁹ En un inicio, se tuvieron en cuenta 20 variables que estaban incluidas en el índice de estética dental (DAI)¹⁰ e índice de prioridades de tratamiento ortodóntico (IPTO)¹¹ (por ser empleados con el mismo fin) y su posible inclusión en el nuevo índice fue sometida al criterio de especialistas. Finalmente, se emplearon las nueve variables que mostraron resultados significativos en ambos índices y que, además, tuvieron correlaciones bajas entre ellas, una vez realizado el análisis de componentes principales. Los predictores incluidos fueron: afectación de la estética (X_1), apiñamiento (X_2), irregularidad anterior (X_3), rotaciones (X_4), relación molar anteroposterior (X_5), cierre labial anormal (X_6), mordida cruzada (X_7), mordida abierta dentaria (X_8) y resalte (X_9). La variable de respuesta fue «necesidad de tratamiento según el especialista». Todas las variables fueron medidas dicotómicamente.

La nueva metodología tiene un carácter puramente aritmético; es, de hecho, una suma de los predictores presentes en cada sujeto, pero se trata de una suma ponderada por la V de *Cramer* correspondiente a cada predictor. Los valores de V de *Cramer* son indicativos de la fuerza de asociación que tiene cada predictor con la respuesta de interés. El modelo que incluye las nueve variables del modelo ortodóntico quedaría de la siguiente forma:

$$R(V; X) = V_{X_1}(X_1) + V_{X_2}(X_2) + V_{X_3}(X_3) + V_{X_4}(X_4) + V_{X_5}(X_5) + V_{X_6}(X_6) + V_{X_7}(X_7) + V_{X_8}(X_8) + V_{X_9}(X_9)$$

Donde la función $R(V; X)$ calculada para un paciente en particular, constituye el riesgo de presentar el evento al tener en cuenta el «peso» de los factores presentes en él. Por ejemplo, el riesgo de tener necesidad de tratamiento ortodóntico para un paciente que presente los factores, X_2 , X_5 y X_6 sería:

$$r(y = \text{necesidad de tratamiento}) = V_{X_2} + V_{X_5} + V_{X_6}$$

Para realizar la validación interna de los modelos predictivos, una de las estrategias descritas en la literatura consiste en dividir aleatoriamente la base de datos de estudio en dos subgrupos: uno para desarrollar el modelo (muestra de entrenamiento) y el otro para validarlo (muestra de prueba).³ Sin embargo, algunos autores plantean que, a menos que la muestra sea particularmente grande (>20 000), la división para derivar y evaluar el modelo no debe hacerse al azar, ya que ambas son seleccionadas para ser similares y, por tanto, desempeñarse de manera muy favorable. Se sugiere entonces que en los modelos en los que se utiliza esta estrategia de validación interna,

un mejor enfoque es no aleatorizar la partición (esto es, utilizar ciertas escuelas o hacer una partición temporal).^{12, 13}

Basados en esta última recomendación, el índice ortodóntico creado se evaluó en una muestra de prueba conformada por 181 estudiantes elegidos al azar de una escuela primaria de Santa Clara, que formaron parte de un estudio piloto anterior.¹⁴

Si se sumaran los sujetos de la muestra de prueba y los de la muestra de entrenamiento, los primeros constituirían aproximadamente la tercera parte del total, tal y como se recomienda al emplear esta estrategia para la validación interna.^{6,15}

Este tipo de validación implica evaluar en los datos de prueba las mismas medidas de desempeño empleadas en la muestra de entrenamiento.

Se evaluaron las medidas clásicas de discriminación: el área bajo la curva (AUC) *Receiver Operating Characteristic* (ROC) y los parámetros calculados con las matrices de confusión y sus respectivos intervalos de confianza al 95 %. Mediante la discriminación, se analiza si los pacientes que presentan la respuesta de interés («necesidad de tratamiento ortodóntico» para este caso de estudio), tienen mayor riesgo, según el modelo predictivo elaborado, que los que no lo presentan.^{2,16}

Para la validación del índice desarrollado con la nueva metodología, se compararon, además, las medidas de discriminación obtenidas por él, con las obtenidas por la técnica estándar en salud para desarrollar índices predictivos, esta es, la regresión logística. Para ejecutar esta última, se empleó el programa SPSS versión 17.0 y el método «hacia atrás» fue el utilizado para seleccionar los modelos.

Para la construcción de las matrices de confusión, se utilizaron los mismos puntos de corte (p) empleados en el desarrollo de los modelos: ellos fueron 1,030 para el obtenido mediante la nueva metodología y 0,514 para el de la regresión logística. Estos fueron seleccionados por medio de un criterio que determina la sensibilidad y especificidad más alta conjuntamente (para un mismo punto), el cual se calcula por medio del índice de Youden.¹⁷ Se tuvo en cuenta que con los puntos de corte seleccionados, al utilizar cada método, se obtuvieran modelos con iguales valores de sensibilidad y especificidad.

RESULTADOS

El modelo ortodóntico obtenido por medio de la nueva metodología quedó de la siguiente forma:

$$r(y = \text{tto.}) = 0,709 * X_1 + 0,526 * X_2 + 0,491 * X_3 + 0,414 * X_4 + 0,324 * X_5 + 0,257 * X_6 + 0,232 * X_7 + 0,177 * X_8 + 0,109 * X_9$$

Mediante él se expresa el riesgo máximo que tiene un paciente de la población estudiada, de tener necesidad de tratamiento ortodóntico. Su valor es de 3,24.

La Tabla 1 muestra que el modelo obtenido mediante el nuevo enfoque superó, en los parámetros calculados a partir de las matrices de confusión, al modelo que se obtuvo por regresión logística, con valores de sensibilidad, especificidad y validez de 79,3 %, 84,3% y 81,2 %, respectivamente.

Tabla 1. Comparación de modelos según parámetros de las matrices de confusión.

Parámetros de las matrices de confusión	V de Cramer	Regresión logística
Sensibilidad	79,3 (78,6-80)	76,6 (75,8-77,3)
Especificidad	84,3 (83,3-85,3)	78,6 (77,4-79,7)
Validez	81,2 (80,8-81,6)	77,3 (76,9-77,8)

Al analizar las *AUC* de las técnicas utilizadas para obtener el índice, la del modelo nuevo fue de 0,886, superior a la del modelo de la regresión (Tabla 2).

Tabla 2. Descriptivos del área bajo la curva ROC de los clasificadores en comparación.

Variables resultado de contraste	Área	Error típico	Significación asintótica	Intervalo de confianza asintótico al 95 %	
				Límite inferior	Límite superior
V de Cramer	0,886	0,027	0,000	0,834	0,939
Regresión logística	0,883	0,026	0,000	0,833	0,934

DISCUSIÓN

El desempeño de los modelos predictivos representa un dominio de interés y de importantes aplicaciones. La habilidad de predecir correctamente la necesidad de tratamiento ortodóncico es de gran beneficio sanitario, sobre todo desde el punto de vista administrativo, ya que permite garantizar los recursos materiales y humanos para las personas que realmente los necesitan. Los bioestadísticos investigan sobre técnicas cada vez más sencillas y fáciles de interpretar por los decisores en salud, con el fin de que estos puedan utilizar sin prejuicios, estas herramientas de trabajo como apoyo a sus decisiones.

Los hallazgos obtenidos de la replicación durante el desarrollo del modelo original en diferentes datos, pero provenientes de la misma población diana subyacente, son clave.¹⁸⁻²⁰ Esto garantiza la reproducibilidad del modelo obtenido.⁷

Se espera que los modelos predictivos se desempeñen de manera muy favorable en la base de datos con la que fue desarrollado, comparado con el desempeño encontrado cuando se prueba en nuevos, pero comparables individuos. Esto es simplemente porque el modelo fue diseñado para ajustarse óptimamente en la muestra de desarrollo, pero se vuelve menos exacto cuando se prueba en nuevos, pero similares individuos (sobreajuste). El optimismo potencial del desempeño del modelo, aumenta cuando el número de resultados de interés en la muestra de desarrollo decrece y aumenta el número de predictores candidatos en ella (relacionados con el número de resultados de interés).^{6,7} Para estimar el potencial del sobreajuste y optimismo en el desempeño del modelo, se aboga por la aplicación de técnicas de validación interna; por este motivo, si se comparan los resultados obtenidos en las muestras de desarrollo⁹ con los presentes, se observará en estos últimos una disminución en la calidad de todos los parámetros evaluados. Sin embargo, en la validación de los modelos obtenidos mediante la nueva metodología, se mantienen muy buenos resultados. En el caso de los parámetros de las matrices, todos superan el 70 %, y con relación a las *AUC*, son superiores a 0,85; en todos los casos, los intervalos de confianza no

incluyen al 0, lo que significa que estos índices discriminan mucho mejor entre individuos con resultados opuestos, que si se hiciera al azar.

El modelo creado con la metodología que emplea los valores de V de *Cramer* logró resultados superiores a los alcanzados por técnicas más complejas, como la regresión logística. Por tanto, si bien no se puede afirmar que esta nueva metodología supera con creces a las clásicas, sí es una alternativa útil para el desarrollo y evaluación de modelos predictivos. Su mayor mérito consiste en la sencillez con la que se ejecuta, por lo que resulta muy accesible para el personal no especializado.⁸

No pretendemos que la simplicidad de este método lo convierta en una panacea para aquellos investigadores que deseen desarrollar nuevos modelos predictivos. Se advierte sobre la necesidad de evaluar sus resultados mediante la comparación con otros métodos de clasificación disponibles, así como sobre la necesidad de emplear varias muestras para hacerlo, ya sea interna como externamente, según el propósito de la creación del modelo, reproducirlo en la población que sirvió para crearlo o generalizarlo a otras similares.

A pesar de que la validación externa de los modelos predictivos es comúnmente considerada mejor que la validación interna, ya que esta implica más transportabilidad que reproducibilidad, cuando los modelos se crean para utilizarse a nivel local, como en nuestro caso, la validez interna resulta suficiente como requisito anterior a su puesta en práctica.

El hecho de haber utilizado datos reales para confeccionar y validar este modelo, hace que pueda ser aplicado a la población de origen, lo cual se ve reforzado por la facilidad de cálculo y la interpretación directa de sus resultados, lo que permite que pueda emplearse en la atención primaria de salud. Además, utiliza menos recursos que otros métodos estadísticos, pues no requiere de programas computacionales y el tiempo de ejecución es mínimo.

Se pone en manos de los decisores en salud una útil herramienta para el desarrollo o actualización de modelos predictivos, con similar desempeño al de métodos tradicionales.

REFERENCIAS BIBLIOGRÁFICAS

1. Pfeiffer RM. Extensions of criteria for evaluating risk prediction models for public health applications. *Biostatistics*. 2013;14(2):366-81.
2. Collins G, Groot JA, Dutton S, Omar O, Shanyinde M, Tajar M, *et al*. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(40):1-11.
3. Min Oh SM, Stefani K, Chang Kim H. Development and Application of Chronic Disease Risk Prediction Models. *Yonsei Med J*. 2014 Jul.;55(4):853-60.
4. Steyerberg EW. *Clinical prediction models: A practical approach to development, validation, and updating*. New York: Springer; 2009.
5. Collins GS, Reitsma JB, Altman DG, Moons K. GM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med*. 2015;162(1): 55-64.
6. Moons K. GM, Pascal Kengne A, Woodward M, Royston P, Vergouwe Y, Altman DG, *et al*. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98:683- 90.
7. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925-31.
8. González Ferrer V, Alegret Rodríguez M, González Ferrer Y, Vargas Yzquierdo J. Nueva estrategia metodológica para desarrollar índices predictivos en la atención primaria de salud y su impacto en la especialidad de Cardiología. *CorSalud [internet]*. 2015 ene.-mar. [citado 10 jul. 2015];7(1):[aprox. 1 p.]. Disponible en: <http://www.corsalud.sld.cu/sumario/2015/v7n1a15/cartasv7n1.html#indicepredictivo>
9. Moreno Arias A. Validación de índice de necesidad y prioridad de tratamiento ortodóntico [tesis]. Villa Clara: Universidad de Ciencias Médicas; 2014.

10. Cons N, Jenny J, Kohout F. DAI: The Dental Aesthetic Index. Iowa City, Iowa: University of Iowa, College of Dentistry; 1986.
11. Águila FJ. Manual de Ortodoncia: teoría y práctica. Madrid: Editorial Aguiram; 1999.
12. Altman DG, Vergouwe Y, Royston P, Moons GM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
13. Altman DG, Royston P. What do we mean by validating a prognostic model?. *Stat Med*; 2000;19:453-73.
14. Padilla Centeno ML. Propuesta de un índice de maloclusiones y priorización de tratamiento ortodóntico [tesis]. Villa Clara: Universidad de Ciencias Médicas; 2012.
15. Steyerberg EW. Prognostic modeling for clinical decision making. Theory and applications. Utrecht (The Netherlands): Elinkwijk BV; 1996.
16. Sanchis J, Avanzas P, Bayes-Genis A, Pérez de Isla L, Heras M. Nuevos métodos estadísticos en la investigación cardiovascular. *Rev Esp Cardiol*. 2011;64(6):499-500.
17. Rodríguez-Escudero JP, López-Jiménez F, Trejo-Gutiérrez JF. Cardiología "basada en la evidencia": aplicaciones prácticas de la epidemiología. IV. Modelos de predicción de riesgo cardiovascular. *Arch Cardiol Méx* [internet]. 2012 ene.-mar. [citado 10 ago. 2014];82(1):[aprox. 5 p.]. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-99402012000100011&lng=es
18. Ioannidis JP, Khoury MJ. Improving validation practices in "Omics" research. *Science*. 2011;334:1230-2.
19. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, *et al*. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166-75.
20. Ioannidis JP. Scientific inbreeding and same-team replication: type D personality as an example. *J Psychosom Res*. 2012;73:408-10.

Recibido: 31 de marzo de 2015

Aprobado: 25 de junio de 2015

Dra. Vielka González Ferrer. Especialista de Primer Grado en Bioestadística. Instructora. Aspirante a Doctor en Ciencias de la Salud. Cardiocentro Ernesto Che Guevara. Santa Clara, Villa Clara. Cuba. Correo electrónico: vielkagf@capiro.vcl.sld.cu