



ORIGINAL

# Utilización de los modelos de regresión múltiple en estudios observacionales (1970-2013) y requerimiento de la guía STROBE en revistas científicas españolas



J. Real<sup>a,b</sup>, R. Cleries<sup>c,d</sup>, C. Forné<sup>e,f</sup>, A. Roso-Llorach<sup>a,g</sup> y J.M. Martínez-Sánchez<sup>b,h,i,\*</sup>

<sup>a</sup> Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, España

<sup>b</sup> Facultat de Medicina i Ciències de la Salut, Universitat Intenacional de Catalunya, Sant Cugat, Barcelona, España

<sup>c</sup> Pla Director d'Oncologia de Catalunya, Institut Català d'Oncologia, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Barcelona, España

<sup>d</sup> Departament de Ciències Clíniques, Universitat de Barcelona, Campus de Bellvitge, L'Hospitalet de Llobregat, Barcelona, España

<sup>e</sup> Departament de Ciències Mèdiques Bàsiques, Universitat de Lleida, Lleida, España

<sup>f</sup> Oblikue Consulting, Barcelona, España

<sup>g</sup> Universitat Autònoma de Barcelona, Bellaterra, Barcelona, España

<sup>h</sup> Unitat de Control del Tabaquisme, Programa de Prevenció i Control del Càncer, Institut Català d'Oncologia, L'Hospitalet de Llobregat, Barcelona, España

<sup>i</sup> Grup de Prevenció i Control del Càncer, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Barcelona, España

Recibido el 10 de diciembre de 2014; aceptado el 5 de junio de 2015

Disponible en Internet el 6 de noviembre de 2015

## PALABRAS CLAVE

Análisis multivariante;  
Análisis de regresión;  
Modelos logísticos;  
Modelos lineales;  
Modelos de riesgos proporcionales;  
Modelos de Poisson;  
Estudios observacionales;  
Epidemiología

## Resumen

**Fundamentos:** En el ámbito de la investigación médica los modelos de regresión logística, lineal, Cox y Poisson son técnicas estadísticas ampliamente conocidas. El objetivo de este trabajo es describir la evolución de estas técnicas de regresión en los artículos observacionales indexados en PubMed (1970-2013) y revisar los requerimientos de las normas de autor de revistas españolas para conocer si requieren el cumplimiento de la guía STROBE.

**Métodos:** Se realizó una búsqueda dirigida en PubMed para identificar los artículos que utilizaron modelos de regresión logística, lineal, Cox y Poisson. Además, se revisaron las normas de autor de las revistas editadas en España indexadas en PubMed e incluidas en Web Of Science.

**Resultados:** El 6,1% de los artículos de estudios observacionales contenían algún término relativo a los modelos seleccionados, pasando del 0,14% en 1980 hasta un 12,3% en 2013. Este último año, un 6,7% de los artículos contenían algún término referido a regresión logística, un 2,5% a lineal, un 3,5% a Cox y un 0,31% a Poisson. Por otro lado, el 12,8% de las normas de autor de las revistas revisadas recomendaban explícitamente seguir la guía STROBE, y el 35,9%, la guía CONSORT.

\* Autor para correspondencia.

Correo electrónico: [jmmartinez@iconcologia.net](mailto:jmmartinez@iconcologia.net) (J.M. Martínez-Sánchez).

**Conclusiones:** Los modelos de regresión multivariantes en estudios observacionales publicados, tales como la regresión logística, lineal, Cox y Poisson, son cada vez más utilizados tanto a nivel global como en revistas publicadas en lengua española. Además, un porcentaje bajo de las revistas científicas españolas indexadas en PubMed incluyen en las normas de autoría el requerimiento de la guía STROBE.

© 2015 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

## KEYWORDS

Multivariate analysis;  
Regression analysis;  
Logistic models;  
Linear models;  
Proportional hazards models;  
Poisson models;  
Observational studies;  
Epidemiology

## Use of multiple regression models in observational studies (1970-2013) and requirements of the STROBE guidelines in Spanish scientific journals

### Abstract

**Background:** In medicine and biomedical research, statistical techniques like logistic, linear, Cox and Poisson regression are widely known. The main objective is to describe the evolution of multivariate techniques used in observational studies indexed in PubMed (1970-2013), and to check the requirements of the STROBE guidelines in the author guidelines in Spanish journals indexed in PubMed.

**Methods:** A targeted PubMed search was performed to identify papers that used logistic linear Cox and Poisson models. Furthermore, a review was also made of the author guidelines of journals published in Spain and indexed in PubMed and Web of Science.

**Results:** Only 6.1% of the indexed manuscripts included a term related to multivariate analysis, increasing from 0.14% in 1980 to 12.3% in 2013. In 2013, 6.7, 2.5, 3.5, and 0.31% of the manuscripts contained terms related to logistic, linear, Cox and Poisson regression, respectively. On the other hand, 12.8% of journals author guidelines explicitly recommend to follow the STROBE guidelines, and 35.9% recommend the CONSORT guideline.

**Conclusions:** A low percentage of Spanish scientific journals indexed in PubMed include the STROBE statement requirement in the author guidelines. Multivariate regression models in published observational studies such as logistic regression, linear, Cox and Poisson are increasingly used both at international level, as well as in journals published in Spanish.

© 2015 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights reserved.

## Introducción

En el ámbito de la investigación médica los modelos de regresión logística, lineal, de Cox y de Poisson son técnicas estadísticas ampliamente conocidas y utilizadas, ya que permiten evaluar las relaciones entre distintos factores de exposición e indicadores de salud de diversa naturaleza (dicotómicos, continuos, eventos dependientes del tiempo o recuentos)<sup>1,2</sup>. Además, en los estudios observacionales es habitual utilizarlos como herramienta de ajuste debido al potencial sesgo de confusión existente en este tipo de diseño<sup>2,3</sup>. Sin embargo, la mayoría de estos modelos demandan asunciones muy estrictas sobre el ajuste de los datos (linealidad de los predictores, normalidad, homocedasticidad, incolinealidad, etc.), cuyo incumplimiento puede invalidar el modelo y la inferencia realizada<sup>1,4</sup>.

En los últimos años, a fin de mejorar la comunicación y transparencia de los trabajos científicos, han surgido distintas guías de recomendaciones sobre cómo reportar los resultados científicos (CONsolidated Standards Of Reporting Trials [CONSORT]<sup>5</sup>, Statistical Analyses and Methods in the Published Literature [SAMPL]<sup>6</sup>, Strengthening The Reporting of OBServational studies in Epidemiology [STROBE]<sup>7</sup>).

La guía STROBE referencia los aspectos metodológicos esenciales que deben reportar los estudios epidemiológicos observacionales<sup>7</sup>. Actualmente, revistas de reconocido prestigio internacional, como *The Lancet* o *The British Medical Journal*, en sus guías de autor recomiendan seguir la guía STROBE e incluso exigen proporcionar una lista de verificación del cumplimiento de sus 22 puntos para poder enviar los trabajos.

El objetivo del presente trabajo es describir la evolución del uso de modelos de regresión estándares (logística, lineal, Cox y Poisson) en estudios observacionales indexados en PubMed y revisar las normas de autor de revistas españolas con el fin de conocer si requieren el cumplimiento de la guía STROBE.

## Material y métodos

Se realizó una búsqueda en PubMed dirigida a identificar todos los trabajos originales indexados en el repositorio bibliográfico con un diseño observacional que incluyeran entre sus métodos de análisis los modelos multivariantes diferenciando por tipo de modelo de regresión (logístico, lineal, Cox y Poisson). La búsqueda fue limitada a todos los

estudios realizados en la especie humana incluidos desde el 1 de enero de 1970 hasta el 31 de diciembre de 2013, excluyendo ensayos clínicos, editoriales, comentarios o series de casos. Para discriminar entre los tipos de regresión utilizados se añadieron a la búsqueda los términos que hacían referencia a los tipos de análisis multivariante principales: «logistic», «linear», «Cox», «Poisson» y/o «multivariate», «regression», «statistical regression» (tabla 1). Finalmente, y con el fin de evaluar la sensibilidad de la estrategia de búsqueda, se incorporaron como texto libre los siguientes sinónimos de las técnicas multivariantes: «adjusted odds ratio», «adjusted OR», «adjusted relative risk», «adjusted RR», «adjusted hazard ratio» y «adjusted HR».

Se describió la tendencia temporal de la frecuencia del uso de estos modelos de regresión y se calculó el porcentaje de cambio anual (PCA). También se realizó una revisión de las normas de autor de las 39 revistas editadas en España indexadas en PubMed e incluidas en Web Of Science para cuantificar las revistas que recomiendan el uso de las guías STROBE, CONSORT o las recomendaciones del Comité Internacional de Directores de Revistas Médicas, que incluye la guía STROBE, para la elaboración y presentación de los manuscritos<sup>8</sup>.

La gestión de datos se realizó mediante el paquete estadístico IBM® SPSS® v22, y para el análisis gráfico se utilizó la librería ggplot2 del paquete estadístico R3.1.2<sup>9</sup>.

## Resultados

De un total de 2.559.903 artículos observacionales indexados en PubMed, un 9,3% contenía un término relativo al análisis multivariante, pasando del 0,17% en 1970 al 16% en 2013 (PCA 11,15%) (tabla 2). Esta tendencia también se observó en los trabajos en lengua española (PCA 8,56%). El 6,1% de los trabajos reportaron la utilización de modelos de regresión lineal, logística, Cox o Poisson (del 0,14% en 1980 hasta el 12,3% en 2013; PCA 14,5%). Este porcentaje fue inferior para los trabajos en lengua española en todos los años de estudio (fig. 1 y tabla 2). Por otro lado, entre los 238.093 trabajos que contenían el término multivariante, el 48,5% contenía alguno de los modelos estándares estudiados (logístico, lineal, Cox o Poisson), presentando también una tendencia ascendente (del 9,5% en 1980 al 75% en 2013; PCA 6,5%). Se observó una tendencia similar para los artículos en lengua española (datos no mostrados).

El uso del término multivariante aumentó en los artículos indexados en PubMed en los años de estudio de 0,96 a 14,55% (PCA 11,15%). El término relativo al modelo logístico fue el modelo más utilizado durante todo el periodo de estudio (3,39%; PCA 18,47%). Le siguieron los términos referentes al modelo de Cox (1,51%; PCA 20,09%), al modelo lineal (1,31%; PCA 15,37%) y al de Poisson (0,13%; PCA 17,73%) (tabla 2). El último año analizado (2013), un 6,7% contenía términos referidos a regresión logística, un 2,5% a regresión lineal, un 3,49% a regresión de Cox y un 0,31% a regresión de Poisson. Por otro lado, al incluir los sinónimos de las técnicas multivariantes en el texto libre el número de artículos indexados aumentaron en un 0,99%, siendo el aumento del 0,25% en el modelo logístico y del 0,09% en el modelo de Cox.

El 12,8% de todas las revistas indexadas en Web Of Science y PubMed y editadas en España (n=39)

recomendaban explícitamente seguir la guía STROBE antes de enviar el manuscrito, y el 35,9%, seguir la guía CONSORT. El 30,8% recomendaban implícitamente la guía STROBE porque aconsejaban cumplir los requisitos de uniformidad de presentación de manuscritos del Comité Internacional de Directores de Revistas Médicas.

## Discusión

Nuestro trabajo refleja un aumento del uso de las técnicas estadísticas multivariantes en los estudios observacionales indexados en PubMed, especialmente los modelos de regresión logística. Por otro lado, las recomendaciones STROBE para mejorar la comunicación de los resultados científicos de los estudios observacionales solo se indican en el 12,8% de las normas de autor de las revistas editadas en España. En contraste, la guía de recomendaciones para los ensayos clínicos (CONSORT) se propone en un 35,9% de las mismas normas de autor.

El aumento observado del uso de las técnicas de regresión en los estudios observacionales puede deberse a la capacidad computacional de los ordenadores actuales y de los paquetes estadísticos para realizar estos análisis. Una revisión de la metodología estadística empleada en artículos publicados en 2 revistas con alto factor de impacto mostró que en un 16% de los artículos revisados se había utilizado metodología multivariante, siendo la regresión logística la más utilizada (10%) entre los años 2000 y 2007<sup>10</sup>, resultados consistentes con los de nuestro estudio. Por otro lado, una revisión más reciente de estudios donde la fuente de datos primaria fue la Encuesta Nacional de Salud de Canadá también indicó un predominio de la utilización del modelo logístico y un incremento del uso de las técnicas de regresión a lo largo de los años, como nuestro estudio<sup>11</sup>. Sin embargo, el porcentaje global que observaron de la utilización de técnicas de regresión fue notablemente superior al de nuestro estudio (80 vs. 14,5%). Existen motivos que pueden explicar tal discrepancia, como la diferencia en la estrategia de búsqueda o el tipo de referencias incluidas entre ambos estudios. Además, nuestro estudio englobó un universo de artículos de mayor variedad de estudios, especialidades, diseños, muestras y tipos de revistas.

El aumento del uso de las técnicas multivariantes en los artículos científicos observacionales en nuestro estudio también coincide con la cada vez mayor disponibilidad del software para llevarlas a cabo en los últimos años. Sin embargo, el uso de estas técnicas no está libre de potenciales errores y no siempre es apropiado, pues sobre ellas descansan fuertes asunciones que no siempre se cumplen. El incumplimiento de las asunciones formales de los modelos puede invalidar los resultados que se derivan del estudio, produciendo los conocidos errores de tipo I y/o tipo II<sup>1</sup>, o importantes sesgos de las estimaciones<sup>12</sup>. En este sentido, existen herramientas estadísticas que permiten evaluar si los modelos cumplen las condiciones de aplicación<sup>1,4</sup>.

La necesidad de evaluar críticamente la calidad metodológica de los estudios ha puesto de manifiesto graves deficiencias en los artículos de investigación. Estas deficiencias dificultan el desarrollo de revisiones sistemáticas, que posteriormente influyen en el desarrollo de guías de práctica clínica y, en última instancia, sobre el cuidado de los

**Tabla 1** Lista de descriptores utilizados en las distintas estrategias de búsqueda en PubMed

Criterio de búsqueda	Campos PubMed			
	MeSH	ptyp	tiab	text
<i>Estudios observacionales (+)</i>				
Evaluation studies		+		
Evaluation studies as topic	+			
Evaluation study			+	
Evaluation studies			+	
Intervention studies	+			
Intervention study			+	
Intervention studies			+	
Case-control studies	+			
Case-control		+		
Cohort studies	+			
Cohort			+	
Longitudinal studies	+			
Longitudinal			+	
Longitudinally			+	
Prospective			+	
Prospectively			+	
Retrospective studies	+			
Retrospective			+	
Follow up			+	
Comparative study		+		
Comparative study			+	
Observational			+	
<i>Humanos (+)</i>				
Humans	+			
<i>Exclusión de ensayos clínicos, editoriales, comentarios o series de casos (-)</i>				
Editorial/Letter/Comment/Case report		-		
Case report			-	
Case series			-	
Clinical trial		-		
<i>Modelo multivariante</i>				
Multivariate analyses				+
Multivariate/statistical regression				+
Regression/regressions				+
<i>Modelo logístico</i>				
Logistic model/s				+
Logistic regression				+
Logistic regressions				+
<i>Regresión lineal</i>				
Linear model/s				+
Linear regression				+
Linear regressions				+
<i>Modelo de regresión de Cox</i>				
Cox regression				+
Cox regressions				+
Cox models				+
<i>Regresión de Poisson</i>				
Poisson model/s				+
Poisson regression				+
Poisson regressions				+

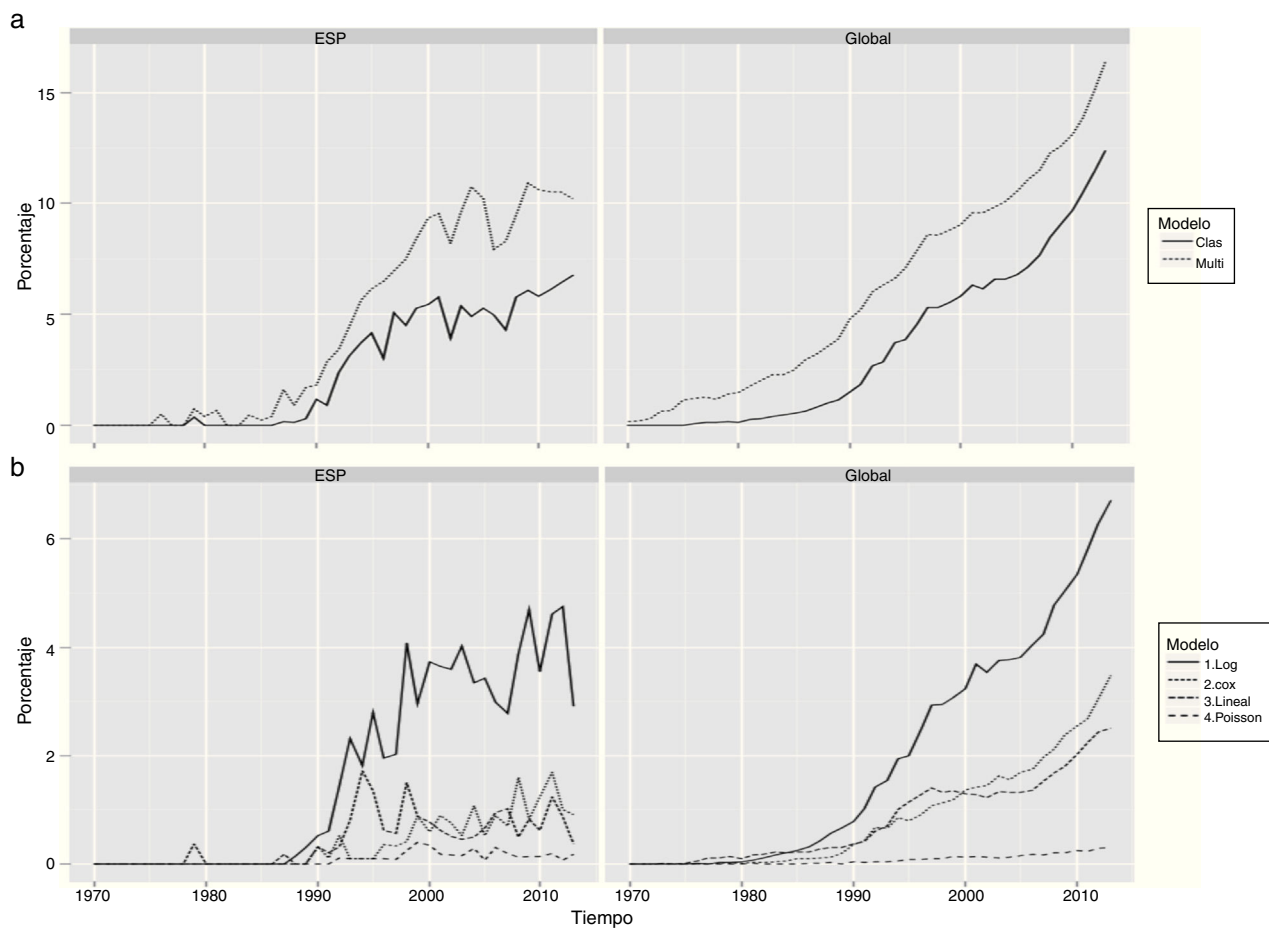
MeSH: términos MeSH; ptyp: tipo de publicación; text: texto completo; tiab: título o resumen; +: término incluido; -: término excluido; (+): incluido en todas las búsquedas; (-): excluido en todas las búsquedas.

Fecha de publicación entre el 1 de enero de 1970 hasta el 31 de diciembre de 2013. Estrategia lanzada en PubMed el 16 de abril de 2014.

**Tabla 2** Evolución porcentual de la utilización de los modelos de regresión logística, lineal, Cox y Poisson en los artículos observacionales indexados en PubMed a nivel global y en lengua española (1970-2013)

Periodo	Número de citas	Logístico, %	Lineal, %	Cox, %	Poisson, %	Término «multivariante», %
<i>Global</i>						
1970-1979	134.839	0,01	0,05	0,00	0,00	0,96
1980-1989	300.325	0,32	0,23	0,09	0,01	2,75
1990-1999	574.794	2,11	1,01	0,84	0,07	7,16
2000-2009	1.013.344	4,07	1,44	1,78	0,15	10,78
2010-2013	536.601	6,02	2,30	2,92	0,27	14,55
1970-2013	2.559.903	3,39	1,31	1,51	0,13	9,30
PCA		18,47	15,37	20,09	17,73	11,15
<i>Lengua española</i>						
1970-1979	3.132	0,00	0,03	0,00	0,00	0,13
1980-1989	4.950	0,08	0,02	0,00	0,00	0,83
1990-1999	10.809	2,13	0,84	0,33	0,13	5,55
2000-2009	14.086	3,59	0,69	0,85	0,19	9,41
2010-2013	4.937	4,15	0,85	1,28	0,14	10,51
1970-2013	37.914	2,49	0,61	0,58	0,13	6,57
PCA		13,27	7,42	4,70	2,72	8,56

PCA: porcentaje de cambio anual de todo el periodo respecto al primer año con porcentaje superior a 0. Número de citas de estudios observacionales según la estrategia establecida en PubMed el 16/04/2014.



**Figura 1** Evolución del porcentaje de los términos referentes a modelos estadísticos en artículos observacionales indexados en PubMed en lengua española (ESP) y a nivel internacional (Global): (a) término multivariante (Multi) respecto a término modelos clásicos (Clas: Logístico, Cox, Lineal o Poisson) y (b) términos específicos relativos a cada modelo.

pacientes. En ese sentido, las guías de recomendaciones son herramientas desarrolladas para facilitar información más exacta y completa de los aspectos clave de los estudios de investigación<sup>13</sup>. De hecho, un estudio realizado en España demostró que su requerimiento mejoraba la calidad de los manuscritos publicados<sup>14</sup>. En este sentido, que los editores y revisores dispongan y requieran guías o herramientas estandarizadas para evaluar la calidad metodológica es clave para mejorar la presentación de los trabajos científicos.

El control de la confusión es uno de los aspectos esenciales incluidos en la guía STROBE<sup>7</sup>. Sin embargo, no incluye explícitamente la validación del método estadístico. Otras guías, como la SAMPL<sup>6</sup>, de carácter más metodológico, sí que incluyen como recomendación la validación de la metodología estadística empleada, aunque también es incompleta en lo que se refiere a metodología multivariante.

Una de las limitaciones de esta revisión deriva en que la fuente analítica principal se basa en el motor de búsqueda de PubMed. Otra limitación del estudio es la estrategia de búsqueda utilizada, ya que los resultados de la búsqueda dependen de que los autores hayan mencionado las técnicas de regresión en el resumen. En este sentido, nuestros resultados globales podrían estar infraestimados. Por otro lado, también puede haber varios artículos derivados del mismo proyecto, que puede ocasionar cierta sobreestimación del resultado. Este tipo de limitación es común en los estudios bibliométricos basados en motores de búsqueda por palabras como PubMed<sup>10</sup>. Sin embargo, teniendo en cuenta que a partir del año 1990 todos los descriptores ya estaban incorporados en PubMed y que nos hemos basado en todo su repositorio, en términos de evolución, el efecto de esta limitación probablemente sea reducido.

Por otro lado, en relación con el requerimiento sobre el cumplimiento de STROBE, se han revisado todas las revistas españolas que en la última actualización del año 2012 estaban indexadas en la Web Of Science y PubMed. De todos modos, este análisis no está directamente vinculado con el estudio bibliométrico, y no tiene una referencia temporal, ya que la guía STROBE se publica por primera vez en 2007, más tarde que la CONSORT, por lo que las revistas han tenido más tiempo en adaptar las normas de publicación a esta última.

En conclusión, los modelos de regresión multivariantes (logística, lineal, Cox y Poisson) en estudios observacionales publicados e indexados en PubMed son cada vez más utilizados tanto a nivel global como en revistas publicadas en lengua española. Debido al aumento de la utilización de los métodos multivariantes parece necesario establecer filtros que garanticen el correcto uso de estos métodos. Además, un porcentaje bajo de las revistas científicas españolas indexadas en PubMed incluyen en las normas de autoría el requerimiento de la guía STROBE.

## Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

## Financiación

Los autores no han recibido financiación específica para realizar este estudio.

## Autoría

JMMS concibió el trabajo. JR realizó todos los análisis. JR y JMMS escribieron el primer borrador del manuscrito y todos los autores contribuyeron significativamente en sus versiones posteriores. Todos los autores han aprobado la versión final del manuscrito.

## Conflicto de intereses

Los autores declaran no tener conflictos de intereses.

## Bibliografía

1. Dobson AJ. An introduction to generalized linear models. 2nd ed. United States of America: Chapman and Hall; 2001.
2. Bender R. Introduction to the use of regression models in epidemiology. En: Mukesh V, editor. *Methods in Molecular Biology, Cancer Epidemiology*. Totowa, NJ: Springer Science; 2009. p. 179–95.
3. Szklo M, Nieto FJ. *Epidemiología intermedia. Conceptos y aplicaciones*. Madrid: Díaz de Santos; 2003.
4. Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Stat Med*. 1995;14:1707–23.
5. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152:726–32.
6. Lang TA, Altman DG. *Statistical analyses and methods in the published literature: The SAMPL Guidelines*. John Wiley & Sons; 2014.
7. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *Gac Sanit*. 2009;23:158, e1–28.
8. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. *ICMJE*; 2011 [consultado 15 Jun 2014]. Disponible en: [www.icmje.org](http://www.icmje.org)
9. Wickham H. *ggplot2: Elegant graphics for data analysis*. New York: Springer; 2009.
10. Scotch M, Duggal M, Brandt C, Lin Z, Shiffman R. Use of statistical analysis in the biomedical informatics literature. *J Am Med Inform Assoc*. 2010;17:3–5.
11. Yergens DW, Dutton DJ, Patten SB. An overview of the statistical methods reported by studies using the Canadian community health survey. *BMC Med Res Methodol*. 2014;14:15.
12. Liang W, Zhao Y, Lee AH. An investigation of the significance of residual confounding effect. *Biomed Res Int*. 2014;2014:658056.

13. Altman D, Hoey J, Marušić A, Moher D, Schulz KF. EQUATOR Network. Enhancing the QUALity and Transparency Of health Research. 2014 [consultado 5 Nov 2014]. Disponible en: <http://www.equator-network.org>
14. Cobo E, Cortés J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: Masked randomised trial. *BMJ*. 2011;343:d6783.