

Una concepción topológica del "bootstrap" permite la demostración del sesgo de Berkson en epidemiología nutricional

A topological conception of "bootstrap" proofs Berkson bias in nutritional epidemiology

Gómez González, Carmen¹; Peña Rodríguez, Amelia²; Salas Díaz, Inmaculada³; Praena Fernández, Juan Manuel⁴; Gálvez Acebal, Juan⁵; Lozano Rodríguez, Jesús⁶; Vilches Arenas, Ángel⁷; Ortega Calvo, Manuel⁸

1 Unidad de Cuidados Intensivos. Ciudad Sanitaria Virgen del Rocío. Servicio Andaluz de Salud. Sevilla.

2 Unidad Clínica de Cardiología. Área del Corazón. Hospital Virgen del Rocío. Sevilla.

3 Centro de Salud San Pablo. Unidad de Investigación Distrito Sanitario de Atención Primaria. Sevilla.

4 Unidad de Estadística, Metodología y Evaluación de Investigación. Hospital Virgen del Rocío/IBIS/FISEVI. Sevilla.

5 Unidad Clínica de Enfermedades Infecciosas, Microbiología y Medicina Preventiva. Hospital Universitario "Virgen Macarena". Departamento de Medicina. Universidad de Sevilla.

6 Grupo de Investigación AL-ANDALUS. Sevilla.

7 Departamento de Medicina Preventiva. Universidad de Sevilla.

8 Unidad de Investigación Distrito Sanitario de Atención Primaria. Sevilla. CIBER de la Fisiopatología de la Obesidad y de la Nutrición. Instituto de Salud Carlos III. Madrid. España.

Recibido: 29/10/2015. Aceptado: 24/agosto/2016.

RESUMEN

Introducción: El núcleo semántico común para todos los usos del término "bootstrap" es la realización de una tarea compleja mediante la práctica de un gesto sencillo (un individuo y su caballo pueden dar un gran salto después de que tan sólo el jinete se haya tirado de los cordones de las botas). El "bootstrapping" es un método estadístico diseñado para la estimación de la distribución muestral de un estimador mediante remuestreo con reemplazamiento.

Metodología: Intentando compensar las debilidades epistemológicas del cálculo del tamaño muestral, se deben obtener por parte de los investigadores, los valores más pequeños posibles del error relativo muestral o de efecto diseño. Por otro lado, nosotros podemos también crear un universo virtual (U_V) ubicando topológicamente las muestras obtenidas mediante "bootstrap".

Resultados: El tamaño de U_V será aproximadamente igual al número de repeticiones multiplicado por el tamaño de la muestra original. En términos frecuentistas podemos emitir una hipótesis de igualdad (H_0) y otra de desigualdad (H_1) entre nuestro U_V y la población real (P_R) de donde proviene la muestra primitiva. Para sustentar estas hipótesis hemos desarrollado un ejercicio práctico de demostración del sesgo de Berkson en un diseño de casos y controles mediante bootstrap.

Conclusión: Nosotros defendemos una concepción topológica del remuestreo con "bootstrap" que permite ampliar el esquema jerárquico de validación externa propuesta por Justice y cols. a un nivel 0.1 tan sólo con la realización del efecto simulador en el paquete de datos del estudio primitivo. Este concepto permite la demostración del sesgo de Berkson en epidemiología nutricional.

PALABRAS CLAVE

Estudios de muestreo, estudios de validación, estadística no paramétrica, epidemiología, bioestadística, sesgo.

Correspondencia:
Manuel Ortega Calvo
106mayorque104@gmail.com

ABSTRACT

Introduction: The common semantic core for all uses of "bootstrapping" is the realization of a complex task by practicing a simple gesture (an individual and his horse can take a big leap after only rider has been thrown the bootlaces). The "bootstrapping" is a statistical method designed to estimate the sampling distribution of an estimator by re-sampling with replacement.

Methodology: Trying to compensate for epistemological weaknesses of sample size calculations should be obtained by the researchers the smallest possible values of the sampling relative error or design effect. On the other hand, we can also create a virtual universe (\sqrt{U}) by a topological placing of samples obtained by "bootstrap".

Results: \sqrt{U} size will be approximately equal to the number of repeats multiplied by the size of the original sample. In frequentist terms we can issue an equality hypothesis (H_0) and another of inequality (H_1) between our \sqrt{U} and the actual population (${}_A P$) from which comes the sample. To support these hypotheses we have developed a practical demonstration of Berkson bias in a case-control design by bootstrap resampling.

Conclusion: We stand for a topological concept of resampling with "the bootstrap" that can extend the hierarchic external validation scheme proposed by Justice et al. to a 0.1 level just to the embodiment of the simulator effect on the original data package study.

This notion allows the Berkson bias display in nutritional epidemiology.

KEYWORDS

Sampling studies, validation studies, non parametric statistics, epidemiology, statistics, bias.

ABREVIATURAS Y ACRÓNIMOS

AUC: Área under the curve (área por debajo de curva).

deff: Efecto diseño (design effect).

Erm: Error relativo muestral.

F_V : Fuerza de validación externa.

H_0 : Hipótesis nula.

H_1 : Hipótesis alternativa.

P_R : Población real.

ROC: Receiver operating characteristic (Característica operativa del receptor).

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis²⁴.

U_V : Universo virtual.

INTRODUCCIÓN

El escritor alemán Rudolph Raspe publicó en el año 1785 la obra titulada; "Las narraciones del Barón de Münchhausen y sus maravillosos viajes y campañas en Rusia". En ella refería un episodio en el que el protagonista lograba salir de un lodazal en donde se hallaba retenido con su caballo al "apretarse los cordones de sus botas y saltar hacia fuera..." ("pulling his bootstraps up ..."). El significado exacto de "bootstraps" es el de unos cordones enlazados una o dos veces en la parte superior de la bota de montar a caballo¹. Esa imagen literaria sustenta el significado del término "bootstrapping" que no sólo se utiliza en estadística sino también en informática, en economía, en biología, en electrónica e incluso en lingüística¹⁻².

El núcleo semántico común para todos estos usos es el de obtener una acción compleja mediante la realización de un gesto sencillo (un individuo y su caballo pueden dar un gran salto después de que tan sólo el jinete se haya tirado de los cordones de las botas).

Remuestreo y bootstrap

El "bootstrapping" es un método estadístico diseñado para la estimación de la distribución muestral de un estimador mediante remuestreo con reemplazamiento². Fue "inventado" por Bradley Efron²⁻⁶ y desarrollado posteriormente por él mismo junto a Tibshirani⁷. Realizar un "bootstrap" es crear una serie de muestras "posibles" a partir de la real.

Llevando a cabo el análisis estadístico de todas las muestras posibles ("población virtual") se pueden extraer conclusiones de cómo se distribuye el estimador obtenido a partir de la muestra original. Efron diseñó el "bootstrap" con el objetivo de realizar inferencias.

Otra diferencia estriba en que el "bootstrapping" emplea remuestreo con reemplazamiento que es más exacto en términos de probabilidad de simulación. En la validación cruzada y en el "jackknife", el tamaño de la submuestra es menor que el de la muestra original, mientras que cuando utilizamos "bootstrap" cada submuestra tiene el mismo tamaño que la muestra original. Por lo tanto el bootstrap posee la ventaja de "modelizar los impactos" que se pueden producir sobre el tamaño de muestra verdadero⁸⁻⁹.

El "jackknife" es una forma de hacer remuestreo creando estimaciones de la muestra primitiva pero siempre con un elemento menos. Se calcula determinando el promedio de todas esas estimaciones. Es especialmente útil para la evaluación de la varianza y del sesgo del estimador.

Ventajas de las técnicas de remuestreo

En primer lugar están las de tipo empírico. Los procedimientos basados en el muestreo convencional se fundamentan en distribuciones teóricas que necesitan fuertes asuncio-

nes previas al análisis tanto en la muestra como en la población. Existen también ventajas de tipo conceptual, al ser el remuestreo un método limpio y claro que no necesita de un soporte matemático sofisticado⁸.

Las asunciones previas que necesitan las distribuciones estadísticas teóricas se suelen cumplir sólo en los tamaños de muestra grandes. Cuando el tamaño de muestra es pequeño y no satisface las asunciones paramétricas, el remuestreo es un recurso bastante útil⁸. Sin embargo, Good¹⁰ estableció que el test de permutación está sujeto al problema de Behrens-Fisher, en el cual la estimación se considera problemática cuando la varianza poblacional se desconoce. Para ser correctos, los tests de permutación asumen todavía varianzas iguales que es también una premisa requerida en las pruebas clásicas de muestreo⁸.

Los procedimientos tradicionales requieren de muestreo aleatorio para validar las inferencias desde la muestra a la población general. Edgington¹¹ demostró que el remuestreo es válido para cualquier tipo de datos, tanto aleatorios como no aleatorios. Lunneborg¹² sugirió que aunque la utilización de muestras no aleatorias en el remuestreo puede que no conduzca a una conclusión de tipo inferencial, por lo menos el ocuparse de muestras no aleatorias nos puede informar más acerca de las características locales de los datos y de la estabilidad de los resultados.

Si disponemos de una muestra pequeña, aunque su estructura cumpla los requerimientos paramétricos, puede que carezca del suficiente poder. El "bootstrapping" puede tratar una muestra pequeña como una población virtual para generar así más observaciones.

Los procedimientos clásicos de muestreo no informan a los investigadores de la probabilidad que tienen los resultados de replicarse. Los estudios de remuestreo con validación cruzada o con "bootstrapping" pueden utilizarse como repeticiones internas¹³.

Tamaño de muestra, dogma o falacia

La inferencia estadística ha ido obteniendo un papel relevante en el contexto de la investigación clínica y epidemiológica a lo largo de los últimos 50 años. En el proceso de transición ha encontrado más de un escollo, principalmente en relación con su aplicación correcta. Como todo proceso de cimentación, ha experimentado fortalecimientos, ha exigido correcciones y presenta dificultades¹⁴.

Quizás el ejemplo más elocuente de esta realidad sea la aplicación de técnicas y procedimientos de manera ritual. Con frecuencia se emplean con la convicción de que proveerán la respuesta adecuada a las preguntas formuladas más en virtud de una tradición o de una expectativa externa que como un recurso necesario para solucionar el proyecto de investigación.

Estos problemas se han estudiado desde hace muchos años, pero más recientemente se han agregado evidencias acerca de su perjudicial impacto sobre la popularidad de la estadística en particular y los intereses de la ciencia en general.

En este contexto surge la dificultad de decidir el tamaño muestral, un asunto que desde el punto de vista económico es muy relevante y que el investigador necesariamente tiene que afrontar (de manera implícita o explícita). Las guías de publicación a las que se adhieren las revistas más prominentes en investigación sanitaria, exigen que los artículos contengan una justificación detallada acerca de cómo fue determinado el tamaño de muestra empleado. Pero dan por sentada la conveniencia de ofrecer tal justificación sin que se discuta o explique por qué pudiera ser útil. Es interesante que en el propio CONSORT¹⁵, un documento muy explícito en diversas áreas, no figure un solo argumento que respalde tal demanda.

La racionalidad de la exigencia de exponer la forma en que se calculó el tamaño muestral ofrece dudas. Sirva como ejemplo nada más, la regla práctica para el cálculo del tamaño muestral en la construcción y desarrollo de un modelo predictivo con regresión logística binaria, la de diez eventos de interés por variable por la probabilidad del evento que estamos estudiando¹⁶.

Representatividad de la muestra

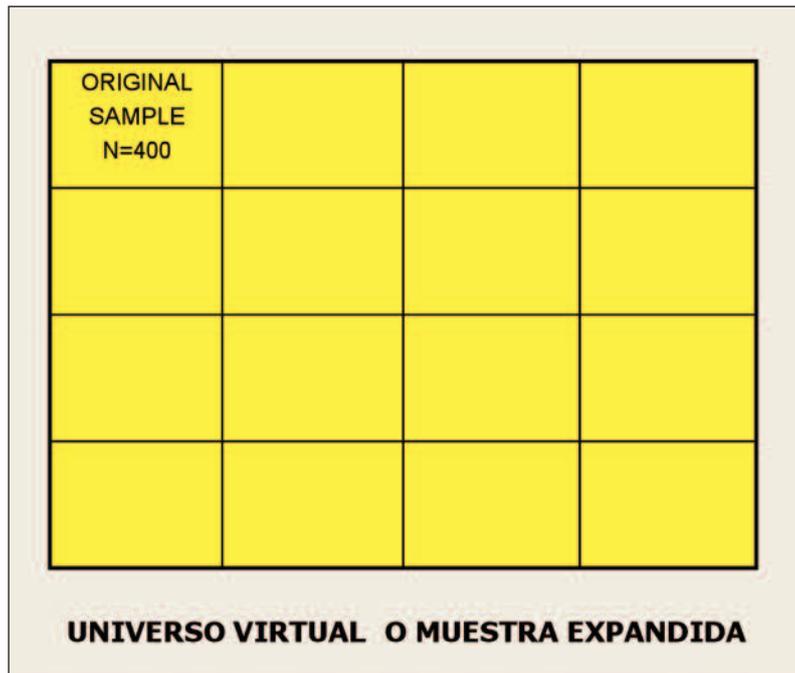
En la medida en la que el investigador cumpla con mayor rigor las diferentes premisas de los diversos tipos de muestreo, mayor será la eficacia representativa de la muestra con respecto a su población-universo de origen¹⁷. Aunque Edgington demostró¹¹ la utilidad del remuestreo en datos no aleatorios, nosotros defendemos el muestreo aleatorio simple como el más adecuado en estos casos. Se debe obtener el menor error relativo muestral posible.

En el caso de que no se pudiera realizar un muestreo aleatorio simple, deberíamos obtener el menor valor posible del denominado efecto diseño (*deff*). El valor de "*deff*" es siempre superior a 1, pero no debe ser excesivo¹⁸.

La creación de un universo virtual

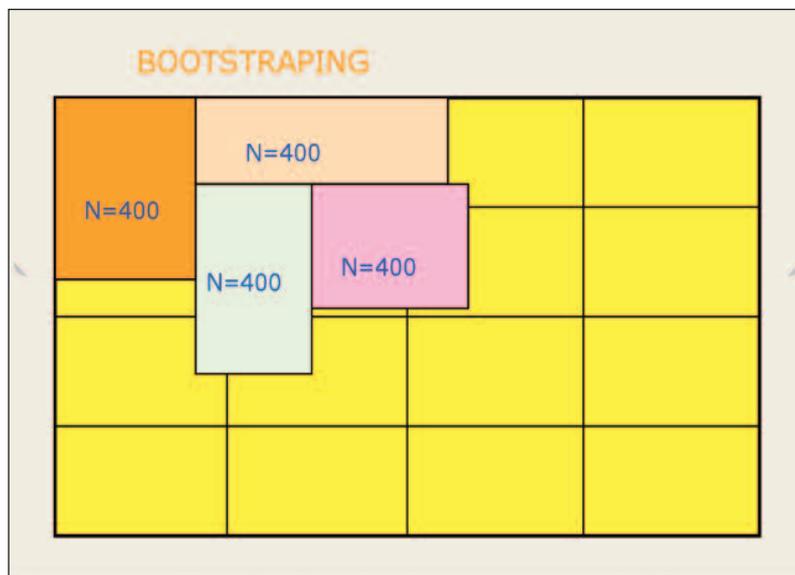
Como se puede apreciar en la Figuras n° 1 y 2, cuando realizamos un procedimiento de bootstrap, estamos creando miles de muestras "nuevas" pero muy parecidas a la primitiva ("original sample") y con todas ellas podríamos confeccionar un Universo Virtual (U_v). El tamaño de U_v viene dado aproximadamente por el número de remuestreos que podamos o queramos realizar multiplicado por el tamaño de la muestra original. Se trata pues de una "**concepción topológica**" del bootstrap, siguiendo el contenido semántico del término en lengua inglesa.

Figura nº 1. Muestra original ("original simple") de 400 elementos en el espacio de un Universo Virtual (U_V).



Tamaño de $U_V \approx N^\circ$ de muestras bootstrap.
 * tamaño de muestra original.

Figura nº 2. Concepción topológica.



Se representan las cuatro primeras muestras bootstrap (de las n posibles) ocupando el espacio del Universo virtual (U_V). Las submuestras pueden estar solapadas parcialmente entre sí.

En términos de estadística frecuentista podemos emitir una hipótesis nula (H_0) de igualdad entre ese universo virtual (U_V) creado por nosotros mediante miles de remue-

stros bootstrap y la población real (P_R) de donde proviene nuestra muestra.

$$H_0 U_V \sim P_R$$

La hipótesis alternativa (H_1) sería la desigualdad parcial o total de U_V con respecto a P_R .

$$H_1 U_V \neq P_R$$

Para nosotros está claro que no podremos refutar H_0 cuanto más rigurosa haya sido la técnica de selección muestral en nuestro diseño. Para un contraste de hipótesis numérico habremos de utilizar una prueba no paramétrica, entre otras razones porque el "bootstrap" es en sí mismo una prueba no paramétrica.

El bootstrap como validación externa

Con la no refutación de H_0 , las miles de repeticiones internas que supone la técnica del bootstrap se pueden transformar en un tipo de validación externa epidemiológica. Sobre la premisa de que una validez perfecta es imposible de obtener¹⁹, Justice y colaboradores²⁰ propusieron una estructura jerárquica de cinco niveles para la validación externa de un modelo predictivo. Cada nivel refleja los tipos de generabilidad que se han analizado para el modelo en cuestión de una forma acumulada y también el grado de exactitud (calibración y discriminación) manifestada en estas pruebas.

Un modelo predictivo jamás podrá ser totalmente validado de manera que el investigador nunca podrá estar absolutamente seguro de poder aplicarlo al siguiente paciente que se nos presente¹⁹⁻²⁰. Si muestra niveles de exactitud aceptables en escenarios diferentes al primitivo, es más probable que se comporte en de la misma forma en cualquier entorno nuevo en donde se le analice. El esquema jerárquico será por lo tanto un proceso acumulativo²⁰.

En el nivel 0 de validación interna se emplean técnicas de exclusión de datos (muestra obtenida mediante corte aleatorio) o técnicas de remuestreo ("bootstrapping") mediante las cuales se estudia el grado de sobreoptimización del modelo. En cierta forma se evalúa nada más que la reproducibilidad. La validez interna se considera un prerrequisito para la validez externa.

Como hemos referido líneas arriba, nosotros defendemos una concepción topológica del "bootstrap" que permite definir un nivel 0.1 de validación externa siguiendo las ideas de Justice y colaboradores²⁰ (Tabla nº 1). De tal

Tabla nº 1. Jerarquía de la validación de un modelo predictivo modificado de Justice y colaboradores²⁰.

Jerarquía de Validación Epidemiológica
Nivel 0: Validación interna.
Nivel 0.1: Remuestreo con <i>bootstrap</i> .
Nivel 1: Validación prospectiva.
Nivel 2: Validación independiente.
Nivel 3: Validación en varios lugares.
Nivel 4: Validaciones independientes múltiples.
Nivel 5: Validaciones independientes múltiples con diferentes períodos de seguimiento.

forma que cuando un investigador aplique una técnica de remuestreo bootstrap esté realizando ya una validación mixta ("interna" y "externa") de su modelo estadístico – epidemiológico. Interna por la concepción clásica del remuestreo y externa por la conceptualización topológica.

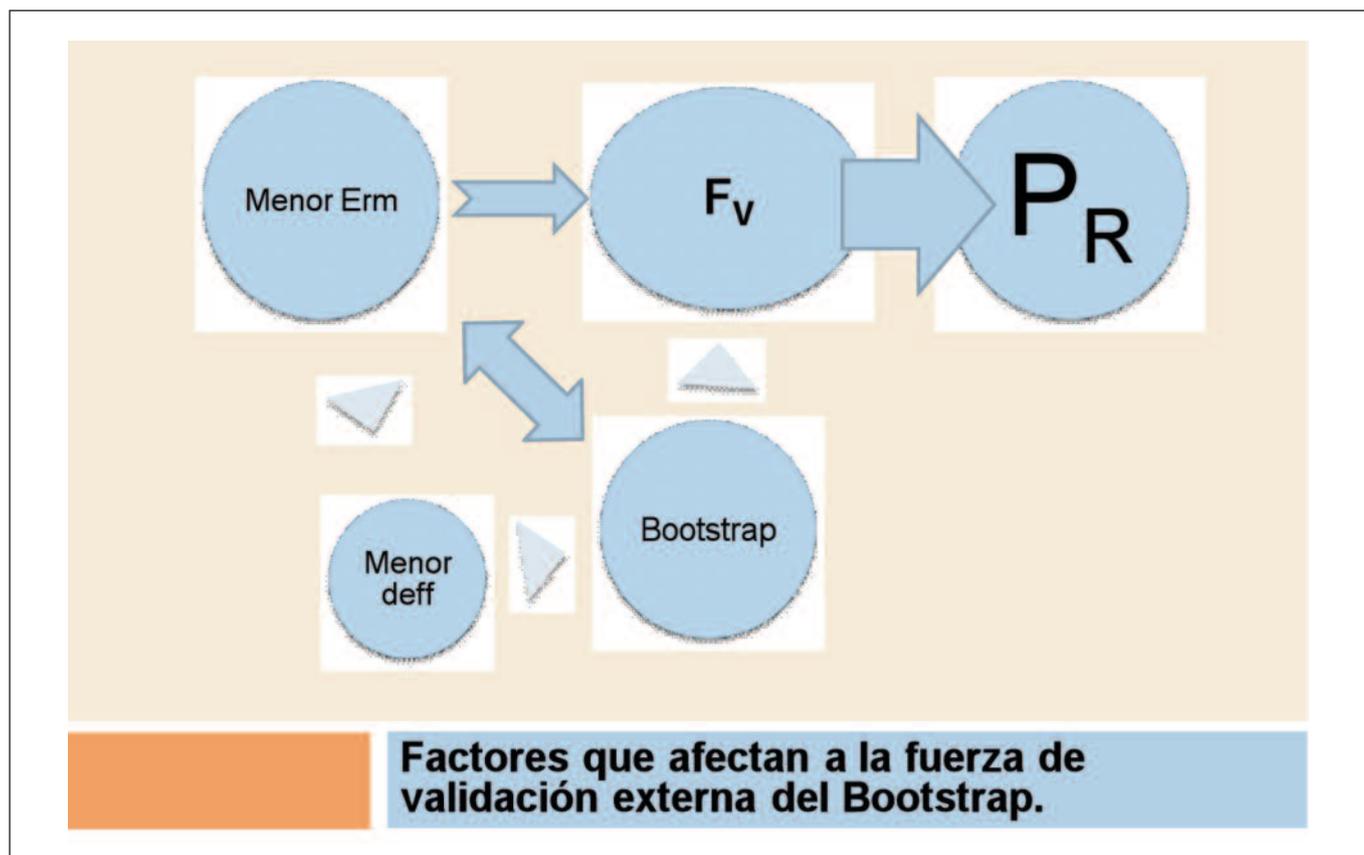
Con esta fusión conceptual entre la validación interna y la externa mediante la aplicación de remuestreo con "bootstrap" entendemos que se intentan compensar en parte las

debilidades epistemológicas del cálculo del tamaño muestral y las dificultades de la representatividad en la obtención de las muestras. La fuerza de la validación externa (F_v) con respecto a la población real (P_R) será mayor cuanto menor sea el error relativo muestral (E_{RM}) y/o el efecto diseño (*deff*) (Figura nº 3).

Whittemore y cols. publicaron un diseño de muestreo en dos pasos para la validación externa de modelos de riesgo personalizados²¹ mediante estimaciones bootstrap de las varianzas. La solución informática del análisis la realizaron confeccionando un script para el programa estadístico R (RMAP) de acceso libre en Internet. El uso de la simulación es cada vez más frecuente en la investigación clínica²² tanto en el campo de los estudios observacionales como en los ensayos clínicos²³.

La iniciativa TRIPOD nos parece un instrumento muy útil para la mejora de la calidad de los trabajos de investigación epidemiológicos sobre modelos predictivos, entre sus items sí que incluyen la explicación de los tamaños muestrales²⁴. Las ideas expuestas en este artículo quizás sean argumentos en favor de la lenta fusión de la bioestadística y la epidemiología en una disciplina común, la ciencia de los datos bíosanitarios²⁵.

Figura nº 3. Efectos del bootstrap sobre la fuerza de validación externa epidemiológica.



Erm = error relativo muestral; deff = efecto diseño; F_v = fuerza de validación externa; P_R = población real.

Aplicaciones del bootstrap en la investigación epidemiológica

Sobre todo han sido notables en epidemiología genética. Guy y cols.²⁶ describieron un método basado en remuestreo bootstrap de árboles de decisión alternantes para el estudio de la interacción entre factores ambientales y la herencia del individuo en trastornos genéticos complejos.

Salganik²⁷ diseñó un método basado en el remuestreo bootstrap para el cálculo del tamaño muestral de poblaciones ocultas como por ejemplo los adictos a drogas intravenosas. Este investigador recomienda un tamaño de al menos el doble del considerado necesario en un supuesto de muestreo aleatorio puro. Otro campo de extraordinaria aplicación del bootstrap es la Filogenómica²⁸. Al realizar inferencias filogenéticas se manejan paquetes de datos muy grandes que permiten extraer conclusiones evolutivas con varianzas extremadamente pequeñas y con significaciones estadísticas elevadas. Se necesita en estos casos calcular también la magnitud de las diferencias. Una aplicación muy utilizada en este campo es un test bootstrap para determinar la fortaleza de un árbol filogenético inferido por alineación de secuencias homólogas.

Biomarcadores

En los estudios de epidemiología nutricional uno de los principales problemas es conocer la ingestión de alimentos y sus componentes de manera válida y precisa. Para ayudar en este proceso se ha planteado repetidas veces la necesidad de contar con buenos biomarcadores, que de manera objetiva nos permitan conocer la dieta consumida. Existen varias definiciones de biomarcador y también distintas clasificaciones de los mismos.

En general un biomarcador es una característica que se puede medir objetivamente en distintas muestras biológicas y que puede evaluarse como indicador de exposiciones, de procesos biológicos normales o patogénicos o de respuestas a una intervención determinada. Las muestras biológicas más utilizadas en epidemiología nutricional son sangre total, eritrocitos, plasma, suero, orina, uñas, saliva, heces y muestras de distintos tejidos. En estas muestras se pueden determinar biomarcadores de exposición (ingesta dietética), biomarcadores de efectos y biomarcadores de estado de enfermedad. A su vez los biomarcadores de exposición pueden categorizarse temporalmente en biomarcadores de efectos agudos, a medio plazo y crónicos²⁹.

OBJETIVO

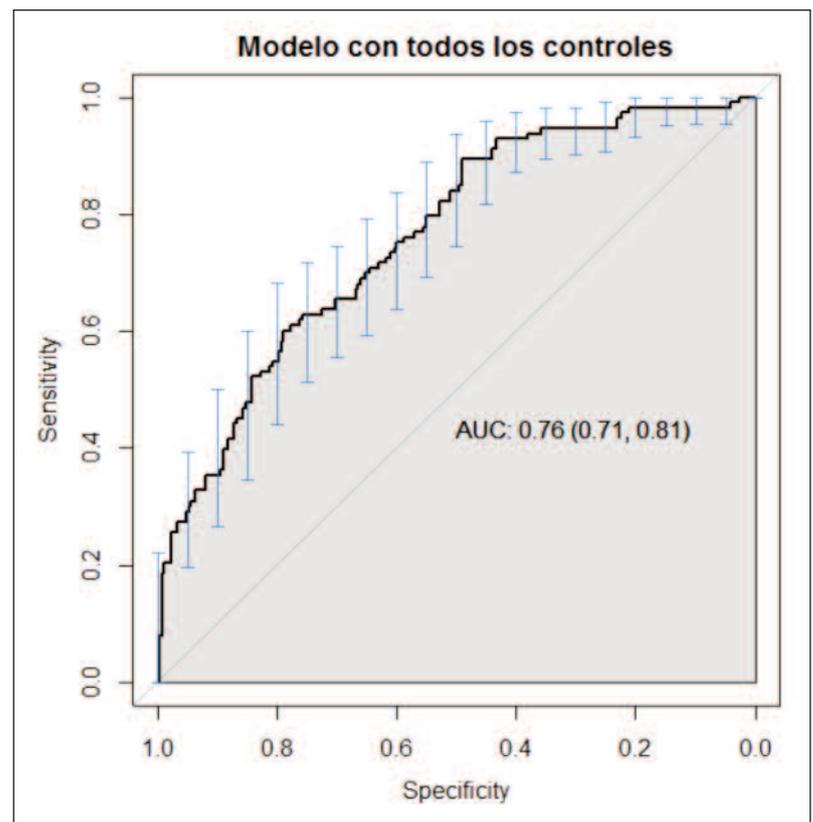
Nuestro objetivo en este artículo ha sido desarrollar un método demostrativo de la existencia de sesgo selección de Berkson en los diseños de casos – control de epidemiología nutricional ba-

sado en la concepción topológica del remuestreo bootstrap. El sesgo de selección de Berkson es inherente al diseño de casos – control siempre que se hayan seleccionado controles en atención primaria y en los hospitales.

MÉTODOS

Entendiendo que la disminución del colesterol y de sus fracciones sean biomarcadores de exposición a medio plazo del carcinoma colorrectal, Villadiego y cols. publicaron en el año 2009 un artículo que contenía un modelo predictivo de neoplasia colorrectal esporádica en fases iniciales sobre un diseño de casos y controles^{30 31}. Los controles tenían origen en varios hospitales y en varios centros de atención primaria. El sesgo de Berkson³²⁻³³ es un sesgo de selección típico de los diseños de casos y controles que indica que los controles que se recogen en atención primaria suelen estar más sanos que los recogidos en instituciones hospitalarias aunque ambos cumplan la condición metodológica de control (criterios de inclusión y exclusión). Nosotros vamos a exponer aquí tres curvas ROC no suavizadas y confeccionadas con los modelos logísticos del mencionado trabajo original. El primero elaborado con todos los controles (Figura nº 4),

Figura nº 4. Curva ROC del modelo con todos los controles de atención primaria y hospitalarios (modelo crudo).



AUC: Área under the curve (área por debajo de curva). Los dígitos entre paréntesis son los intervalos de confianza. Sensitivity: Sensibilidad. Specificity: Especificidad.

Figura nº 5. Curva ROC del modelo con todos los controles de atención primaria (modelo con controles de primaria).

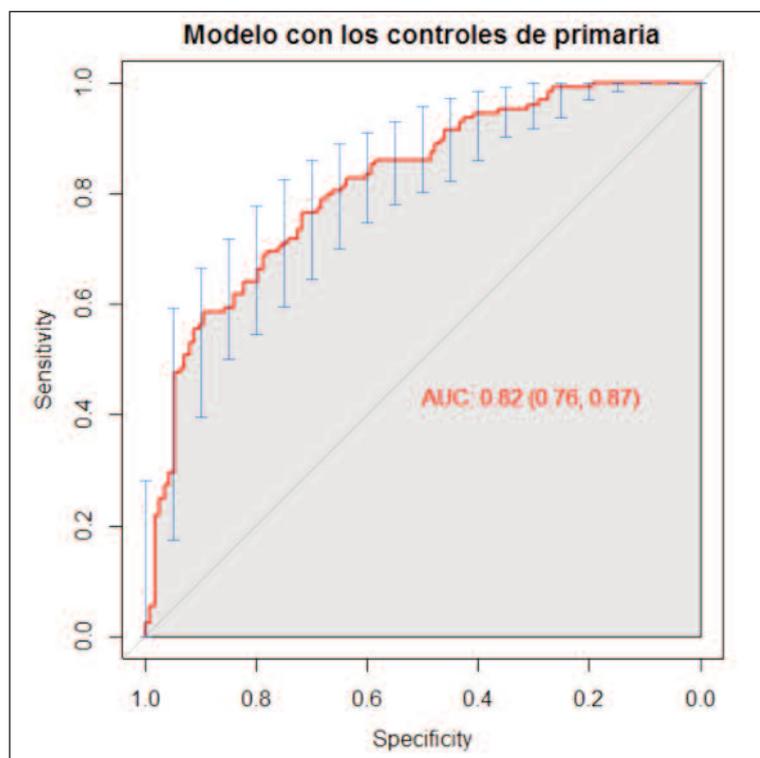
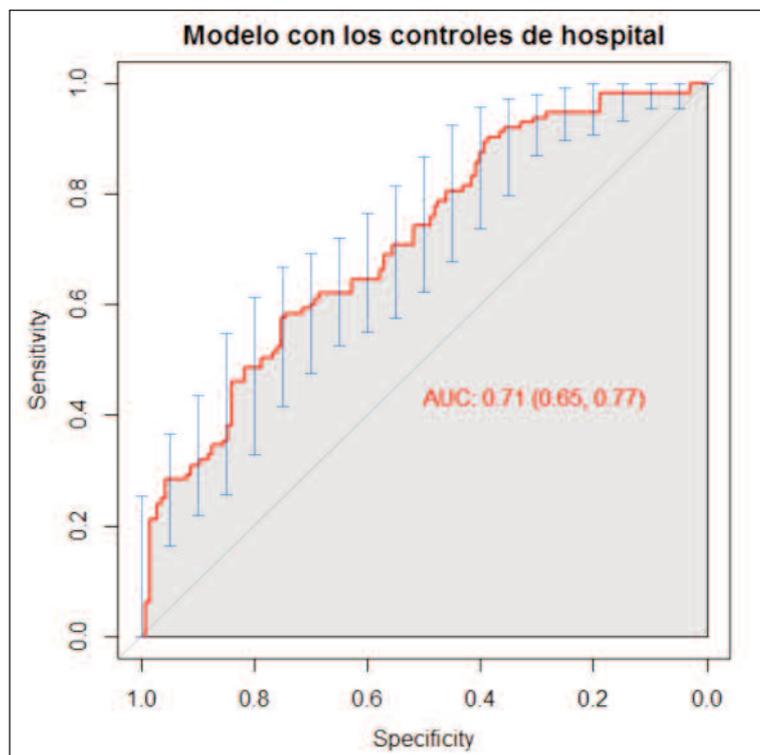


Figura nº 6. Curva ROC del modelo con todos los controles de hospital (modelo con controles de hospital).



AUC: Área under the curve (área por debajo de curva). Los dígitos entre paréntesis son los intervalos de confianza. Sensitivity: Sensibilidad. Specificity: Especificidad.

el segundo realizado con los controles de atención primaria (Figura nº 5) y el tercero creado con los controles de origen hospitalario (Figura nº 6).

Tantos las áreas de las curvas ROC como sus respectivos intervalos de confianza están confeccionados mediante remuestreo bootstrap (2000 repeticiones no estratificadas). El software empleado fué el programa R y el pluginROC de RCommander, que permite estas procedimientos estadísticos³⁴.

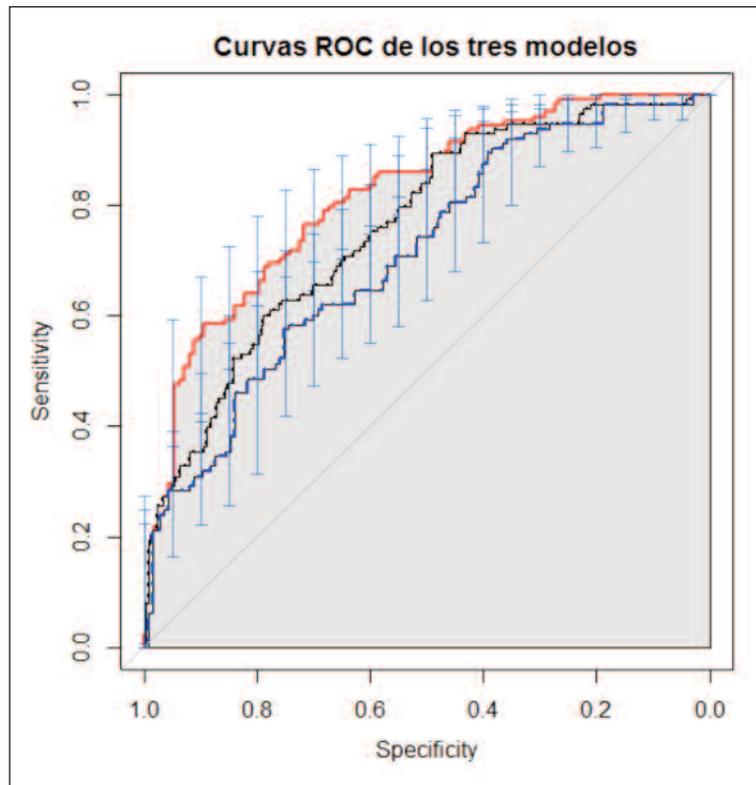
RESULTADOS Y DISCUSIÓN

Como puede observarse por un análisis gráfico y numérico muy simple de las Figuras 4 a 6, el modelo que mejor permite discriminar es el que muestra la Figura nº 5 elaborado con los controles de atención primaria (Área bajo la curva ROC=0.82; IC al 95% 0.76 - 0.87). Y sucede así porque existe sesgo de Berkson. Este hecho se puede visualizar mejor en la Figura nº 7 que recoge las tres curvas ROC en colores diferentes (los controles de primaria en rojo, todos los controles en negro y los controles de hospital en azul).

Creemos que este método, basado en remuestreo bootstrap no estratificado, permite demostrar gráficamente la existencia de sesgo de Berkson, los controles de atención primaria enfrentados con los mismos casos que los otros tipos (de origen hospitalario y totales) muestran una mejor discriminación binaria al estar menos enfermos que los controles de hospital. El tamaño original de la muestra era de 401 elementos. Pero el modelo con más alto poder de discriminación, el de los controles de primaria, se construyó con 254 individuos (126 casos y 128 controles) que al ser remuestreados unas 2000 veces, pueden condensarse en un universo virtual (U_v) máximo de 500000 individuos, que a nuestro entender permitiría realizar una primera validación externa de los resultados²⁰ a partir de la concepción topológica expuesta en la introducción. Este ejercicio práctico expone otra utilidad del remuestreo bootstrap³⁵, la demostración gráfica del sesgo de Berkson en epidemiología nutricional y refuerza algunos instrumentos ya conocidos de lectura crítica en los estudios observacionales^{24,36-37}.

CONCLUSIÓN

El cálculo del área bajo la curva ROC y de sus respectivos intervalos de confianza mediante bootstrap en modelos de regresión logística binaria contruídos con controles de primaria y de hospital, permite la demostración gráfica del sesgo de Berkson en los diseños de casos y controles.

Figura nº 7. Curvas ROC de los tres modelos.

Curva con los controles de primaria en rojo, con todos los controles en negro y con los controles de hospital en azul. Los dígitos entre paréntesis son los intervalos de confianza. Sensitivity: Sensibilidad. Specificity: Especificidad.

BIBLIOGRAFÍA

- Página Web. Palabras comunes. <http://www.worldwidewords.org/qa/qa-boo2.htm>. Visitada el 12 de noviembre de 2015.
- Término de Wikipedia. Bootstrapping. <http://en.wikipedia.org/wiki/Bootstrapping>. Visitada el 12 de noviembre de 2015.
- Diaconis, P., Efron, B: Métodos estadísticos intensivos por ordenador. *Investigación y Ciencia*.1983;82: 70- 83.
- Efron B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*.1979; 7: 1-26.
- Efron B. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*.1981; 68: 589-599.
- Efron B. The jackknife, the bootstrap, and other resampling plans. *Society of Industrial and Applied Mathematics. CBMS - NSF*. 1982. Monographs. n 38.
- Efron, B, Tibshirani R J. *An introduction to the bootstrap*. New York: Chapman & Hall. 1993.
- Yu, Chong Ho. *Resampling methods: concepts, applications, and justification*. Practical Assessment, Research & Evaluation. 2003; 8.
- Fan X, Wang L. Comparability of jackknife and bootstrap results: An investigation for a case of canonical correlation analysis. *Journal of Experimental Education*.1996; 64: 173-189.
- Good P. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York: Springer-Verlag.2000.
- Edgington E S. *Randomization tests*. New York: M. Dekker.1995.
- Lunneborg C E. *Data analysis by resampling: Concepts and Applications*. Pacific Grove,CA. Duxbury.2000.
- Ludbrook J, Dudley H. Why permutation tests are superior to t and F tests in biomedical research. *American Statistician*, 1998; 52: 127-132.
- Silva Ayçaguer LC, Alonso Galbán P. Explicación del tamaño muestral empleado: una exigencia irracional de las revistas biomédicas. *Gac Sanit*. 2013; 27:53-7.
- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010 Mar 24.
- Ortega Calvo M, Cayuela Domínguez A. Regresión Logística no condicionada y tamaño de muestra: una revisión bibliográfica. *Rev Esp Salud Pública*. 2002; 76: 85-93.
- Silva Ayçaguer LC. *Diseño razonado de muestras y captación de datos para la investigación sanitaria*. Madrid: Díaz de Santos; 2000.
- Guillén M, Ayuso M. La importancia del efecto del diseño. *Med Clin (Barc)*. 2004; 122 Suppl 1:35-8.
- Altman DG, Royston P. What do we mean by validating a prognostic model ? *Statistics in Medicine*. 2000; 19: 453-473.
- Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann. Intern. Med*. 1999; 130: 515 - 524.
- Whittemore AS, Halpern J. Two-stage sampling designs for external validation of personal risk models. *Stat Methods Med Res*. 2013 Apr 16. 1-17.
- Monleón-Getino T. Uso de la simulación en la investigación clínica. *Med Clin (Barc)*. 2013; 141:550-5.
- Monleón Getino T, Ocaña Rebull J. Simulación como herramienta de optimización de la investigación clínica. *Med Clin (Barc)*. 2006; 127:702-4.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): theTRIPOD statement. *Ann Intern Med*. 2015; 162:55-63.
- Khoury MJ, Lam TK, Ioannidis JP, Hartge P, Spitz MR, Buring JE et al. Transforming epidemiology for 21st century medicine and public health. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:508-16.
- Guy RT, Santago P, Langefeld CD. Bootstrap aggregating of alternating decision trees to detect sets of SNPs that associate with disease. *Genet Epidemiol*. 2012; 36: 99-106.

27. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health*. 2006; 83(6 Suppl):i98-112.
28. Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. Statistics and truth in phylogenomics. *Mol Biol Evol*. 2012; 29:457-72.
29. Corella D, Ordovás JM. Biomarkers: background, classification and guidelines for applications in nutritional epidemiology. *Nutr Hosp*. 2015; 31 Suppl 3: 177-88.
30. Villadiego-Sánchez JM, Ortega-Calvo M, Pino-Mejías R, Cayuela A, Iglesias-Bonilla P, García-de la Corte F, Santos-Lozano JM, Lapetra-Peralta J. Multivariate explanatory model for sporadic carcinoma of the colon in Dukes' stages I and IIa. *Int J Med Sci*. 2009;6:43-50.
31. Méndez-Mora, JL, Ortega-Calvo M, Cayuela-Domínguez, A, Villadiego-Sánchez, JM, Barros-Pérez, MM, Cantillana-Martínez J: Comportamiento del CA 19.9 y de la fracción HDL-colesterol en una muestra de pacientes con carcinoma colorrectal esporádico. *An. Med.Interna*. 2004; 21: 227-230.
32. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chronic Dis*. 1986;39:495-504.
33. Regeer EJ, Krabbendam L, De Graaf R, Have MT, Nolen WA, Van Os J: Berkson's bias and the mood dimensions of bipolar disorder. *Int J Methods Psychiatr Res*. 2009; 18: 279-86.
34. Proyecto CRAN. The Comprehensive R Archive Network. <https://cran.r-project.org/web/packages/RcmdrPlugin.ROC/index.html> Visitado el 21 de Julio de 2015.
35. Linden, A, Adams, J.L., Roberts, N. Evaluating disease management program effectiveness. An introduction to the bootstrap technique. *Dis Manage & Health Outcomes*. 2005; 13: 159-167.
36. González-García L, Chemello C, García-Sánchez F, Serpa-Anaya DC et al. Aphorisms and short phrases as pieces of knowledge in the pedagogical framework of the andalusian school of public health. *Int J Prev Med*. 2012; 3:197-210.
37. Ortega-Calvo, M., Santos-Lozano, J., Lapetra, J., Salas-Salvadó, J., Martínez-González, M., Lamuela-Raventós, R., Estruch, R: The Metaphor of Patina. *Open Journal of Philosophy*. 2014; 4: 623-627.