



## FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

### Prueba de Chi cuadrado de homogeneidad en estudios clínicos: una herramienta para analizar diferencias entre tratamientos



R.M. Montaña<sup>a</sup>, Á. Roco-Videla<sup>b</sup>, A.R. Nieves<sup>c</sup> y S.V. Flores<sup>d,\*</sup>

<sup>a</sup> Universidad Arturo Prat, Iquique, Chile

<sup>b</sup> Programa de magíster en ciencias químico-biológicas, Facultad de Salud, Universidad Bernardo O'Higgins, Santiago de Chile, Chile

<sup>c</sup> Universidad Central de Venezuela, Caracas, Venezuela

<sup>d</sup> Instituto de Geografía, Pontificia Universidad Católica de Chile, Valparaíso, Chile

Recibido el 24 de mayo de 2024; aceptado el 5 de agosto de 2024

Disponible en Internet el 5 de diciembre de 2024

#### PALABRAS CLAVE

Prueba de Chi cuadrado;  
Estadística;  
Formación científica en salud

**Resumen** Este artículo explora la aplicación de la prueba de Chi cuadrado de homogeneidad en el contexto de estudios clínicos, utilizando como ejemplo un estudio previo sobre la eficacia de 2 analgésicos en el tratamiento de la cefalea. La muestra incluyó 215 pacientes, cuyas respuestas al tratamiento fueron evaluadas y registradas en 4 categorías de intensidad de dolor. La prueba permitió identificar diferencias significativas entre los tratamientos evaluados, demostrando ser una herramienta estadística valiosa para la comparación de grupos en investigaciones clínicas. Este enfoque proporciona claridad y precisión en la interpretación de datos categóricos, esencial para la toma de decisiones informadas en la práctica médica.

© 2024 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

#### KEYWORDS

Chi square test;  
Statistics;  
Scientific health education

**Chi-square test of homogeneity in clinical studies: A tool for analyzing differences between treatments**

**Abstract** This article explores the application of the chi-squared test of homogeneity in the context of clinical studies, using a previous study on the efficacy of two analgesics for treating headache as an example. The sample included 215 patients, whose responses to treatment were assessed and recorded across four pain intensity categories. The test allowed for the identification of significant differences between the treatments evaluated, proving to be a

\* Autor para correspondencia.

Correo electrónico: [sergio.flores.c@pucv.cl](mailto:sergio.flores.c@pucv.cl) (S.V. Flores).

valuable statistical tool for comparing groups in clinical research. This approach provides clarity and precision in the interpretation of categorical data, essential for informed decision-making in medical practice.

© 2024 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

## Introducción

La prueba de «Chi cuadrado ( $\chi^2$ )» o «Ji al cuadrado», como también se le denomina, fue desarrollada por Karl Pearson en 1900 como una forma de valorar la bondad del ajuste de unos datos a una distribución de probabilidad conocida. El empleo de esta prueba no paramétrica permite verificar si existen diferencias significativas entre las frecuencias de los valores de una variable observada en una muestra, con unas frecuencias esperadas que se postulan como válidas para toda la población a través de una hipótesis. Se considera dentro de la estadística descriptiva aplicada al estudio de 2 variables con énfasis en extraer información sobre la muestra, convirtiéndose en una herramienta de gran utilidad en diversos campos de estudio, incluyendo la salud, donde se trabaja frecuentemente con variables de tipo cualitativo y que requieren de una mejor interpretación de los resultados<sup>1</sup>.

Con relación a esto último, la prueba de Chi cuadrado ha permitido analizar la asociación entre variables categóricas y realizar comparaciones de distribuciones de frecuencia en estudios observacionales o experimentales<sup>2</sup>, en investigaciones de tipo epidemiológico<sup>3</sup>, ensayos clínicos<sup>4</sup>, genética médica<sup>5</sup>, supervivencia<sup>6</sup>, enfermedades infecciosas<sup>7</sup>, comparación de prevalencia<sup>8</sup> y factores de riesgo<sup>9</sup>, entre otros, convirtiéndola en una herramienta estadística con gran versatilidad y facilidad de uso en la investigación científica. Ejemplo de ello son los estudios realizados por Gálvez et al.<sup>10</sup>, quienes aplicaron la prueba de Chi cuadrado para comparar 2 variables cualitativas (tipo de derivaciones vs. prioridad de las derivaciones), Álvarez et al.<sup>11</sup> la utilizaron para comparar las variables categóricas entre los grupos y Herrero et al.<sup>12</sup> analizaron la posible relación entre las características de la migraña y las variables sociodemográficas mediante el uso de esta prueba.

## Características de la prueba de Chi cuadrado

Ahora bien, la aplicación de esta herramienta requiere comprender sus limitaciones y tener en consideración algunas condiciones bajo las cuales es pertinente su uso. En este sentido, la prueba de Chi cuadrado no establece restricciones sobre el número de modalidades por variable, y el número de filas y columnas de la tabla no tienen que ser necesariamente iguales, pues la distribución es positivamente asimétrica. Debe aplicarse en aquellos estudios donde las muestras son independientes y donde todos los

valores esperados son superiores a 5 y, por último, si los grados de libertad resultan ser mayores a 30 se estaría en presencia de una Chi cuadrado normal<sup>13</sup>.

## Prueba de Chi cuadrado de homogeneidad

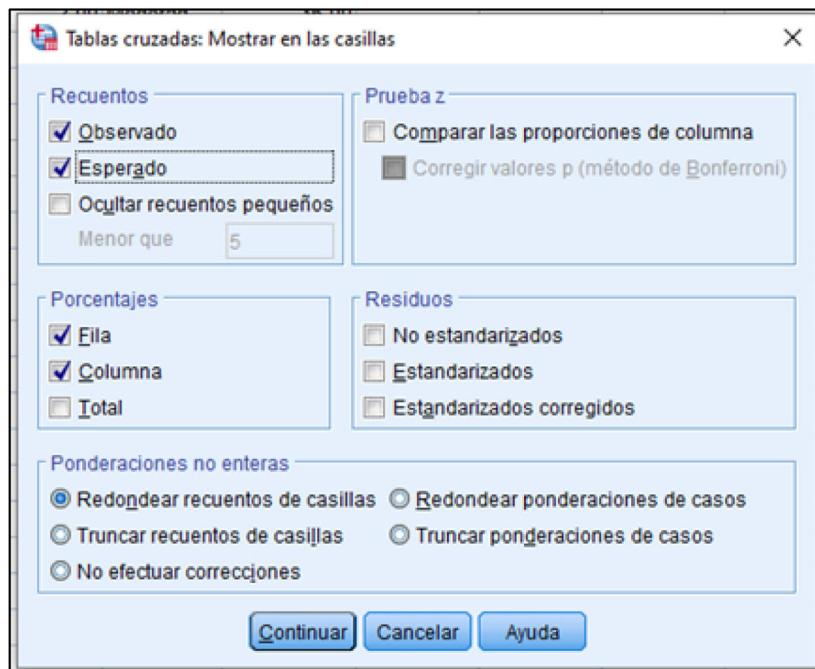
Dentro de los diferentes tipos de pruebas de Chi cuadrado está la prueba de homogeneidad que, a fines prácticos, se utiliza para determinar si existen diferencias significativas entre las distribuciones de frecuencia de 2 o más grupos<sup>14</sup>, resultando conveniente su aplicabilidad en determinados casos clínicos<sup>15</sup>. De ahí que resulta apropiado profundizar en el procedimiento y aplicabilidad de esta herramienta para así lograr una mejor comprensión y claridad en el momento de analizar los datos. Para fines prácticos, este artículo se basará en datos provenientes de un estudio realizado por Chanques et al.<sup>16</sup>, quienes evaluaron el efecto de 2 analgésicos para la reducción del dolor en pacientes con cefalea. La investigación abarcó una muestra de 215 pacientes divididos en 2 grupos, el primero compuesto por 107 pacientes a los cuales se les suministró T1 (analgésico 1) y el segundo conformado por 108 pacientes tratados con T2 (analgésico 2). La intensidad del dolor manifestada por las personas una vez sometidas a los analgésicos evaluados fue registrada a través de 4 escalas: ausente, suave, moderado e intenso.

Se evaluaron 2 características cualitativas: la primera el tipo de tratamiento aplicado (analgésico) y la segunda la intensidad del dolor medida en 4 niveles. Por lo tanto, el uso de la prueba de Chi cuadrado de homogeneidad resulta conveniente para probar si la proporción de pacientes que experimentaron alivio al dolor fue la misma para ambos tratamientos. En función de esto se plantean las siguientes hipótesis a probar para este caso específico:

**Hipótesis nula:** la proporción de pacientes que experimentan alivio al dolor resulta ser la misma para ambos tratamientos (analgésicos) y en los 4 niveles de dolor.

**Hipótesis alternativa:** existen diferencias significativas en la cantidad de pacientes que experimentan un alivio del dolor entre los 2 analgésicos evaluados.

Una vez planteadas las hipótesis se verifica cuál de ellas se ajusta a la realidad, mediante cálculos obtenidos con el programa SPSS versión 25. Primero, se crea una base de datos utilizando el programa de Excel, con la cual se obtienen las frecuencias de ocurrencia de cada una de las variables. Posteriormente, se despliegan 3 columnas. En la primera columna se agrega el factor analgésico, donde «1» indica el primer analgésico y «2» representa el segundo



**Figura 1** Selección de reencuentros observados, esperados, porcentajes de filas y columnas para la construcción de la tabla de contingencia.

analgésico. En la segunda columna se registran los diferentes niveles de dolor medidos (ausente, suave, moderado e intenso). En la tercera columna se registra la frecuencia de ocurrencia de cada combinación entre las 2 características estudiadas. Luego, se carga la matriz de datos en el programa SPSS y se realizan diversos pasos para calcular el estadístico Chi cuadrado. La ruta sugerida a seguir en el programa es la siguiente: a) datos; b) ponderar casos; y c) variable de frecuencia.

Seguidamente, se selecciona la variable con la cual se desea realizar la ponderación, que en este caso es la variable frecuencia. Una vez realizada la selección se procede a aceptar para obtener las frecuencias ponderadas. Es importante señalar que este paso se realiza exclusivamente cuando se tiene contabilizada la frecuencia de cada combinación. Por el contrario, si se dispone de valores individuales para cada combinación de filas y columnas, este paso no sería necesario.

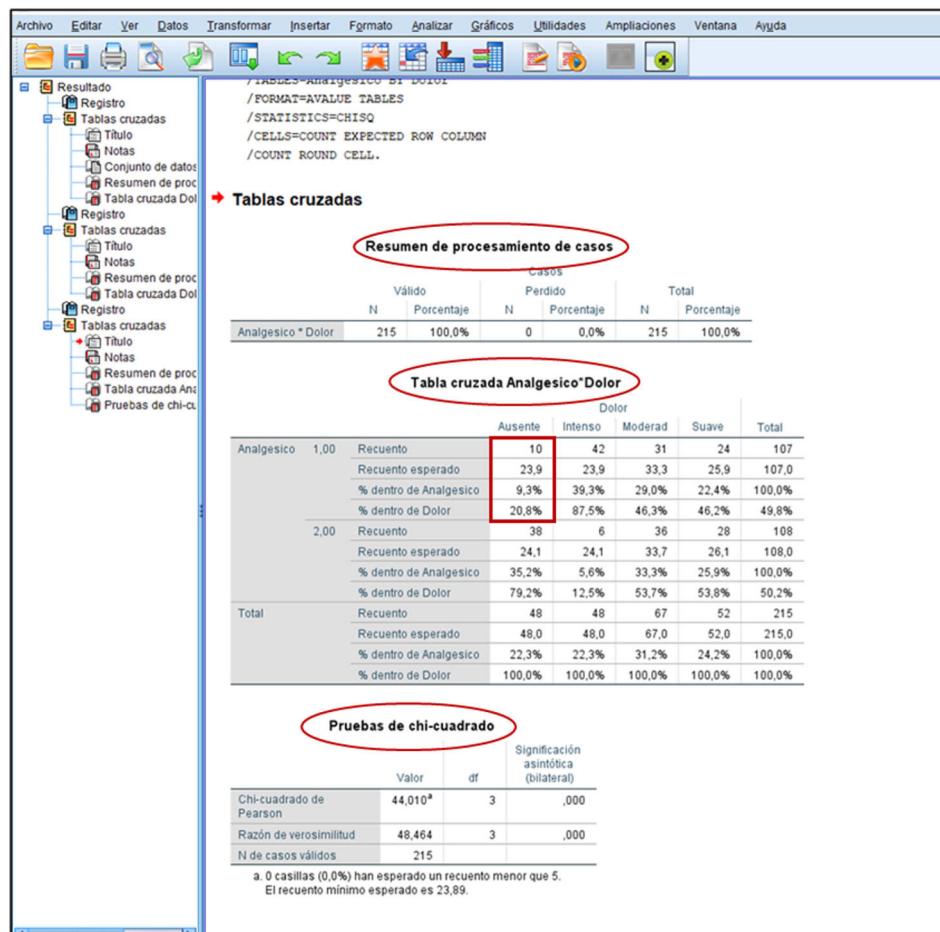
Se solicita la creación de una tabla de doble entrada y el cálculo del estadístico de prueba siguiendo la siguiente ruta: a) analizar; b) estadísticos descriptivos; y c) tablas cruzadas. En el menú se selecciona la opción «filas» para incorporar la calidad analgésica y la opción «columnas» para la calidad de dolor, determinando así la disposición de cada calidad.

Después se elige la opción «casillas» y se marca los recuentos «esperados», así como los porcentajes de «filas» y «columnas», como se muestra en la figura 1. A continuación se procede a clicar en «continuar» en el mismo menú y se selecciona la opción «estadísticos», optando por la alternativa «Chi cuadrado». De esta manera, se obtiene la tabla cruzada con los valores observados y esperados, así como los porcentajes respectivos de cada celda y, finalmente, el estadístico de prueba con su valor de probabilidad (fig. 2).

En la figura 2 se presentan varios aspectos importantes a considerar en el momento de analizar los resultados. En primer lugar, se muestra una tabla titulada «Resumen de procesamiento de casos», donde se puede observar que se utilizaron 215 datos para este caso clínico, lo que indica que no hay datos perdidos y que se está analizando el 100% de la información disponible.

Por otro lado, se presenta una tabla identificada como «Tabla cruzada analgésico», que contiene una gran cantidad de información. En esta tabla se registra inicialmente la cantidad de personas que experimentaron diferentes niveles de intensidad de dolor después de recibir el analgésico. Por ejemplo, el primer valor de la tabla indica el «recuento», que señala que solo 10 personas reportaron ausencia de dolor después de recibir el analgésico uno (tratamiento 1). Asimismo, se proporciona el «recuento esperado» de 23,9, el porcentaje que estas 10 personas representan dentro del total de pacientes que recibieron el analgésico (9,3%) y el porcentaje que representan dentro del total de casos de dolor (20,8%). Por lo tanto, para cada analgésico se proporciona información sobre el valor observado (recuento), el valor esperado (recuento esperado) y los porcentajes que representan en relación con el total de pacientes que recibieron el analgésico y la intensidad de dolor.

Antes de proseguir con el análisis de las combinaciones entre analgésico e intensidad de dolor, es crucial verificar si existe un efecto significativo del factor analgésico. Esto se logra al observar la tabla identificada como «Pruebas de Chi cuadrado» mostrada en la figura 2, donde se encuentra la celda designada como Chi cuadrado de Pearson, con un valor numérico de 44,010 y un valor de p de 0,0001. Al comparar estos datos con el nivel de significación establecido para este estudio (0,05) se rechaza la hipótesis nula ( $0,0001 < 0,05$ ). Por consiguiente, se concluye que existen diferencias signifi-



**Figura 2** Resultados obtenidos de la prueba de Chi cuadro, incluyendo la tabla cruzada mediante el programa SPSS.

cativas en cuanto a la eficacia de los analgésicos para aliviar el dolor experimentado por los pacientes. Esto implica que uno de los analgésicos evaluados presenta una acción superior en el alivio del dolor.

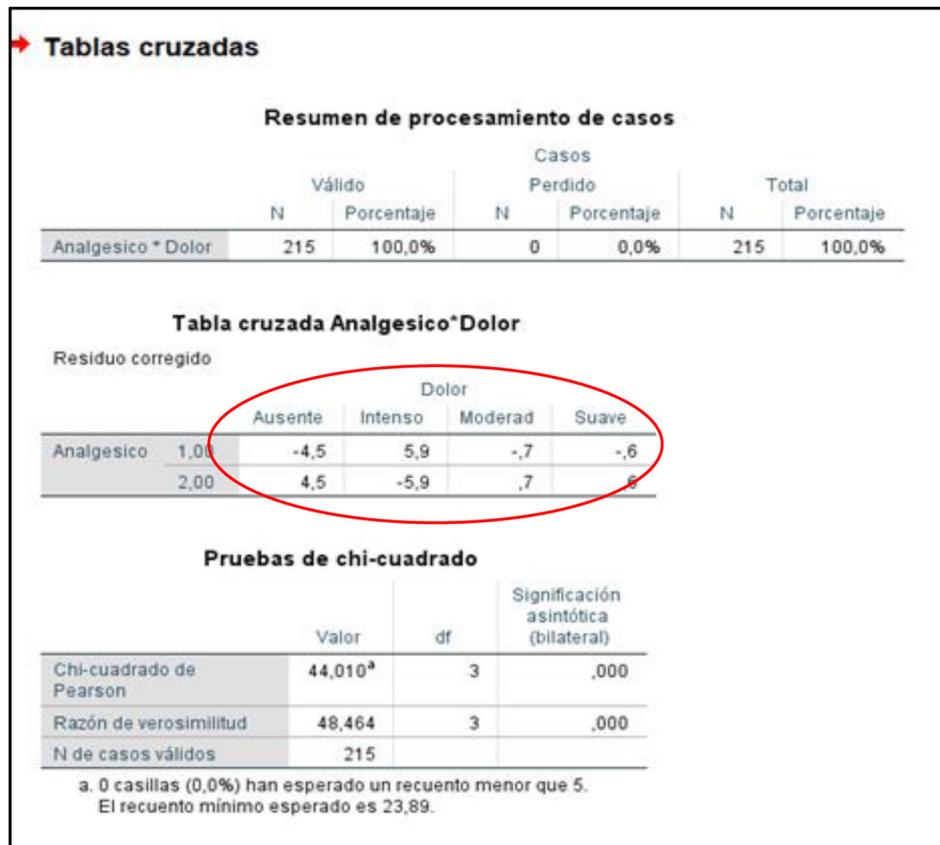
Dada la importancia de la prueba de Chi cuadrado es crucial identificar la combinación más adecuada que ofrezca la mejor respuesta según el objetivo de la investigación. Por lo tanto, se procede a calcular los residuos estandarizados corregidos utilizando el programa SPSS, siguiendo la siguiente ruta: a) analizar; b) estadísticos descriptivos; y c) tablas cruzadas. Una vez en el menú de tablas cruzadas se selecciona la opción «casillas» y luego se eligen los «residuos», específicamente los estandarizados corregidos.

La información generada se presenta en la figura 3, donde se exhiben los residuos estandarizados corregidos para cada una de las combinaciones de la tabla. Estos residuos se incorporan en la base de datos como una nueva variable denominada «Z». Posteriormente, son transformados en valores de Chi cuadrado utilizando el programa SPSS: a) transformar y b) calcular variable. En este menú se incluye la siguiente fórmula: Chi cuadrado:  $Z^2 \cdot Z$ . Luego se determina la probabilidad asociada a estos valores para identificar las combinaciones significativas. Esta probabilidad se determina con el programa SPSS través del siguiente procedimiento: a) transformar; b) calcula variable; c) grupos de funciones; d) significación; e) funciones y variables

especiales; seleccionando Sig.Chisq., de ahí que se genere una probabilidad llamada Sig.Chisq. (Chicuadrado, 1).

Debido a las diferencias significativas observadas entre los analgésicos estudiados y al análisis de los valores de probabilidad, es evidente que el analgésico 2 demostró un rendimiento superior. Esto se refleja en el hecho de que un porcentaje considerablemente mayor de personas no experimentaron cefaleas con este analgésico en comparación con el otro (38% vs. 1%). Por el contrario, el analgésico 1 mostró el menor efecto, ya que un porcentaje significativamente mayor de personas informaron sentir un dolor intenso en comparación con el analgésico 2 (42% vs. 6%). Es importante señalar que estas combinaciones específicas entre analgésicos e intensidades de dolor fueron las que resultaron significativas en la tabla, ya que los valores de probabilidad fueron inferiores al nivel de significación establecido, que en este caso fue del 5%. En cuanto al resto de las combinaciones entre analgésicos e intensidades de dolor, al presentar valores de probabilidad superiores al 5% no se pueden establecer diferencias significativas. Por lo tanto, ambos analgésicos reflejan un comportamiento similar en las escalas de dolor leve y moderado.

Aquí se ha demostrado cómo la prueba de Chi cuadrado de homogeneidad es una herramienta estadística eficaz para evaluar diferencias entre grupos tratados con distintas intervenciones en estudios clínicos. El uso de residuos



**Figura 3** Valores de residuos estandarizados corregidos para la combinación de los niveles de cada variable en estudio.

estandarizados corregidos ayudó a profundizar en el análisis, permitiendo una interpretación más precisa de los datos. Este método puede ser particularmente útil en el campo de la investigación clínica y otros estudios epidemiológicos, donde entender la efectividad de diferentes tratamientos es crucial para la toma de decisiones clínicas y la mejora de las intervenciones de salud.

## Conflictos de intereses

Los autores declaran no tener conflictos de interés.

## Bibliografía

1. Nihan ST. Karl Pearson's Chi-square tests in educational research and reviews. Academic J. 2020;15:575-80, <http://dx.doi.org/10.5897/err2019.3817>.
2. Ibici Akca E, Gökbüllüt N, Cengizhan SO. The effects of MBSR programme on prenatal comfort and fetal health anxiety in pregnant women [Internet]. J Reprod Infant Psychol. 2023;42:449-63, <http://dx.doi.org/10.1080/02646838.2023.2227219>.
3. Wang Q, Xiu S, Yang L, Li L, Yang M, Wang X, et al. Perceptions about respiratory syncytial virus (RSV) and attitudes toward the RSV vaccine among the general public in China: A cross-sectional survey. Human Vaccin Immunother. 2024;20, <http://dx.doi.org/10.1080/21645515.2024.2310916>.
4. Abdelnabi SJ, Munro-Kramer ML, Moyer CA, Williams JEO, Lori JR. Ghanaian women's experience of intimate partner violence (IPV) during group antenatal care: A brief report from a cluster randomised controlled trial. Global Health Action. 2024;17, <http://dx.doi.org/10.1080/16549716.2024.2325250>.
5. Xiang X, Wang D, Leng J, Li N, Wei C. Association of adiponectin and its receptor gene polymorphisms with the risk of coronary heart disease in northern Guangxi. Cytokine. 2024;178, <http://dx.doi.org/10.1016/j.cyto.2024.156567>, 156567.
6. Zhang Y, Li L, Han Q, Wen L. The differential expression of AFF3 in cervical cancer and its correlation with clinicopathological features and prognosis. J Obstet Gynaecol. 2024;44, <http://dx.doi.org/10.1080/01443615.2024.2333784>.
7. Patel K, Twohig P, Peeraphatdit T, Stohs EJ, Samson K, Smith L, et al. Outcomes and factors associated with cryptococcal disease among cirrhotics: A study of the national inpatient sample 2005 to 2014. Clin Res Hepatol Gastroenterol. 2024;48:102337, <http://dx.doi.org/10.1016/j.clinre.2024.102337>.
8. Han L, Liu J, Shataer M, Wu C, Niyazi M. The relationship between long non-coding gene CASC21 polymorphisms and cervical cancer. Cancer Biol Ther. 2024;25, <http://dx.doi.org/10.1080/15384047.2024.2322207>.
9. Xu L, Li D, Song Z, Liu J, Zhou Y, Yang J, et al. The association between monocyte to high-density lipoprotein cholesterol ratio and chronic kidney disease in a Chinese adult population: a cross-sectional study. Renal Failure. 2024;46, <http://dx.doi.org/10.1080/0886022X.2024.2331614>.
10. Gálvez A, Valor M, Pérez M, Ballester I, Canete C, Reig M, et al., Impact on mental healthcare after collaboration between primary healthcare and mental health services. Semergen. 2021;47:385-93, <http://dx.doi.org/10.1016/j.semerg.2021.04.006>.
11. Álvarez J, Díaz A, Méndez E, Crespo N, Cuello C, Seco J. Diagnosis of atherogenic dyslipidaemia by primary

- care physicians in Spain. *Semergen*. 2021;47:99–105, <http://dx.doi.org/10.1016/j.semerg.2020.07.008>.
12. Vicente M, Ramírez M, Ruiz E, Reinoso ML. Migraine medical control and related variables. Results of the European Work and Migraine survey. *Semergen*. 2020;46:225–33, <http://dx.doi.org/10.1016/j.semerg.2019.08.004>.
13. Pavlov G, Shi D, Maydeu-Olivares A. Chi-square difference tests for comparing nested models: An evaluation with non-normal data. *Struct Equ Modeling*. 2020;27:908–17, <http://dx.doi.org/10.1080/10705511.2020.1717957>.
14. Kulinskaya E, Dollinger MB, Bjørkestøl K. Testing for homogeneity in meta-analysis I. The one-parameter case: Standardized mean difference. *Biometrics*. 2010;67:203–12, <http://dx.doi.org/10.1111/j.1541-0420.2010.01442.x>.
15. Schober P, Vetter TR. Chi-square tests in medical research. *Anesth Analg*. 2019;129, <http://dx.doi.org/10.1213/ANE.0000000000004410>, 1193–1193.
16. Chanques G, Viel E, Constantin J, Jung B, de Lattre S, Carr J, et al. The measurement of pain in intensive care unit: Comparison of 5 self-report intensity scales. *Pain*. 2010;151:711–21, <http://dx.doi.org/10.1016/j.pain.2010.08.039>.