

Efecto del sesgo de desclasificación no diferencial e independiente sobre el trastorno, la exposición y una tercera variable

Maribel Peró Cebollero y Joan Guàrdia Olmos
Universidad de Barcelona

El objetivo del presente trabajo consiste en establecer la influencia del sesgo de desclasificación no diferencial e independiente para el trastorno, la exposición y una tercera variable en la medida de asociación a obtener, además del posible patrón que se puede generar. Para conseguir este objetivo se realizaron una serie de simulaciones en las que se generó el sesgo de desclasificación en las tres variables antes mencionadas a partir de la manipulación de la sensibilidad y especificidad de los instrumentos para clasificar a los sujetos en cada una de las tres variables (valores de 0.7 a 1 con incrementos de 0.1). Los resultados muestran que el sesgo de desclasificación generado para el trastorno y la exposición lleva a un sesgo hacia la hipótesis nula en el riesgo relativo, en tanto que en el caso de la tercera variable el patrón que se establece es más complicado, ya que el sesgo puede ser tanto hacia la hipótesis nula como lejos de ella.

The effect of non differential and independent declassification bias towards the disease, the exposure and third variable. The objective of this study was the determination of the influence of the non-differential and independent declassification bias for the disease, the exposure and another variable in the association measure, and also, the obtained pattern. In order to do this, we did some simulations, manipulating the sensibility and specificity of the disease, the exposure and the other variable (0.7 to 1 with increases of 0.1). The results show that the generated declassification bias for the disease and the exposure implies a bias towards the null value in the relative risk, and for the third variable the pattern is more complicated, this bias could be towards or away the null value.

El sesgo de información o de desclasificación consiste en la distorsión en la estimación de la medida de asociación o de efecto debida a una incorrecta clasificación de los sujetos en alguna de las variables a estudiar. Si la desclasificación es igual en los dos grupos a comparar (expuestos-no expuestos o casos-controles) el sesgo es no diferencial, en tanto que si es diferente en los dos grupos a comparar el sesgo es diferencial.

Diferentes autores han puesto de manifiesto la importancia que un sesgo de desclasificación no diferencial puede tener en la determinación de las distintas medidas de relación (Kleinbaum, Kupper y Morgenstern, 1982, Sackett, Haynes y Tugwell, 1989, Rothman, 1986...), e incluso algunos, han estudiado la influencia de la desclasificación no diferencial de la exposición o del trastorno en el caso de la obtención de medidas del efecto, ya sea tanto para la proporción atribuible como para la fracción prevenida (Hsieh y Walter, 1988 y Hsieh, 1991, respectivamente).

Así pues algunos autores como Kleinbaum, Kupper y Morgenstern, (1982) señalan que la desclasificación no diferencial ya sea del trastorno o bien de la exposición, en los casos en que estas

variables son dicotómicas, lleva siempre a un sesgo de la medida de asociación hacia la hipótesis nula, en tanto que una desclasificación diferencial, del trastorno o de la exposición puede llevar tanto a un sesgo hacia la hipótesis nula o lejos de la hipótesis nula; otros autores que aportan evidencias en esta línea son Flegal, Brownie y Haas (1986) en el caso de que el sesgo sea no diferencial tanto para el trastorno como para la exposición (dicotómicas) a excepción de que el sumatorio de la sensibilidad y especificidad del instrumento diagnóstico sea inferior a 1, ya que en este caso se producirá una inversión en la dirección de la relación, o Greenland (1980) para la desclasificación no diferencial de la exposición (dicotómica). Ahora bien, cuando la exposición es politómica o bien continua y el sesgo de desclasificación para esta variable es no diferencial, las conclusiones a las que llegan los autores no son tan claras, ya que si bien algunos apuntan al hecho de que el sesgo en estos casos puede ser tanto hacia la hipótesis nula como lejos de la hipótesis nula, e incluso provocar un cambio en la tendencia de la medida de asociación a través de los diferentes estratos establecidos (Dosemeci, Wacholder y Lubin, 1990, Weinberg, Umbach y Greenland, 1994, Thomas, 1995), otros autores ponen de manifiesto que la desclasificación no diferencial de la exposición siempre provoca un sesgo hacia la hipótesis nula en la medida de asociación entre el trastorno y la exposición (Yuen Fung y Howe, 1984). Algunos autores (Brenner, 1991 y Myers y Ehrlich, 1992) remarcan que si Dosemeci et al (1990) muestran en sus resultados que el sesgo no es siempre hacia la hipótesis nula es debido a que

al obtener la medida de asociación no utilizan el grupo de no expuestos como grupo de referencia o bien a que han creado una situación excesivamente artificial, a lo que los autores (Dosemeci et al 1991 y Dosemeci et al 1992) replican que la inversión de la tendencia es posible pero que ellos no propugnan que tenga que darse siempre cuando existe desclasificación no diferencial de la exposición siendo ésta politómica. Por otra parte, Birkett (1995) argumenta que el marco estadístico de referencia en el que se basan Weinberg et al (1994) es consistente en el caso de que la exposición esté medida en escala de intervalo pero no que esté basada en categorías de exposición como utiliza en el desarrollo de su trabajo, a lo cual Weinber et al (1995) replican que sólo tienen en cuenta el signo de la covarianza para comprobar si se mantiene la tendencia. Además, Delpizzo y Borghesi (1995) ponen de manifiesto que en el caso de que la variable de exposición sea continua un sesgo de desclasificación no diferencial puede resultar diferencial al dicomizar la variable con la consecuente influencia que tiene en la estimación de la medida de asociación a obtener (el sesgo puede ser tanto hacia la hipótesis nula como lejos de ella).

Por otra parte Flegal, Brownie y Haas (1986) remarcan que un sesgo de desclasificación no diferencial en la exposición si además se tiene en cuenta en el análisis una tercera variable que sea dicotómica puede enmascarar las diferencias existentes entre las medidas de asociación para los dos estratos cuando el riesgo relativo es mayor en el grupo con menor prevalencia de la exposición y el test diagnóstico tiene una alta sensibilidad, o cuando el riesgo relativo es mayor en el grupo con mayor prevalencia de la exposición y el test diagnóstico tiene una alta especificidad, o bien si la sensibilidad y la especificidad del test diagnóstico son bajas. Y también puede generar diferencias espúreas en el caso de que la prevalencia de la exposición sea diferente en los dos grupos.

Lo dicho en el párrafo anterior evidencia la necesidad de tener en cuenta la influencia de terceras variables en el estudio del efecto del sesgo de desclasificación, pero ello no se ha de entender única y exclusivamente bajo la perspectiva de incluir a la variable en el análisis sino que también se debería estudiar el efecto que tiene el sesgo que se puede cometer al determinar el estatus para cada uno de los sujetos del estudio en las diferentes variables que pueden afectar a la relación entre el factor de exposición y el trastorno bajo estudio, tal y como señala Greenland (1980). Ahora bien, si en el estudio de la desclasificación para la exposición hemos podido apreciar que existen diferentes puntos de vista y no hay un claro acuerdo en la sistematización del efecto que puede causar en la medida de asociación a obtener, en el caso de incluir una tercera variable los trabajos realizados ponen de manifiesto la no existencia de una patrón claro del efecto de la desclasificación de la tercera variable. (Greenland, 1980, Yuen Fung y Howe, 1984, Savitz y Baron, 1989, Cox y Elwood, 1991).

Greenland (1980) señala que una incorrecta clasificación de la tercera variable puede llegar a enmascarar o exagerar la heterogeneidad de las medidas de asociación entre los dos estratos de la tercera variable, de tal modo que la exageración de la heterogeneidad depende del grado de confusión para cada estrato de la variable a controlar.

Yuen Fung y Howe (1984) ponen de manifiesto que si la variable de confusión está positivamente relacionada con la exposición un sesgo en la clasificación del confusor provocará un sesgo de la medida de asociación lejos de la hipótesis nula, en tanto que si estas dos variables están negativamente relacionadas el sesgo en la medida de asociación será hacia la hipótesis nula y en el caso de

que en realidad esta tercera variable no afecte a la relación entre el factor de exposición y el trastorno un sesgo en esta tercera variable no afectará a la medida de asociación a obtener.

Kupper (1984) citado en Savitz y Baron (1989) puntualiza que el efecto de la desclasificación no diferencial del factor de confusión puede ser hacia o lejos de la hipótesis nula dependiendo del patrón original de la confusión.

Cox y Elwood (1991) para el caso de que la tercera variable sea dicotómica remarcan que el sesgo en la «odds ratio» para la desclasificación no diferencial del confusor depende de la sensibilidad del confusor en uno de los niveles y de la especificidad en el otro nivel, así pues, la «odds ratio» en el primer nivel del confusor no quedará sesgada si la sensibilidad es igual a 1 independientemente de la especificidad y la «odds ratio» del segundo nivel del confusor no quedará sesgada si la especificidad es igual a 1 independientemente de la sensibilidad del confusor. Por otra parte estos autores remarcan que los resultados obtenidos son válidos tanto para el caso de que la tercera variable a tener en cuenta sea un confusor como para el caso de que sea un modificador del efecto.

De todo lo dicho hasta el momento se desprende la importancia de tener en cuenta la adecuación de los instrumentos diagnósticos utilizados en la determinación del estatus de los sujetos que participan en el estudio en los diferentes factores implicados en el mismo, pero por otra parte también se puede apreciar la disparidad de resultados obtenidos en las diferentes investigaciones citadas cuando el sesgo de desclasificación es no diferencial, lo que nos llevó a plantearnos la necesidad de la realización de un estudio empírico en el cual se tuviera en cuenta la desclasificación no diferencial por separado, del trastorno bajo estudio, del factor de exposición y de una tercera variable que pueda afectar a la relación entre las dos primeras, bien porque sea un confusor o bien porque sea un modificador del efecto, con el fin de establecer su influencia en la medida de asociación a obtener y estudiar el patrón que se pueda generar.

Método

Procedimiento

Para poder llevar a cabo el estudio del sesgo de desclasificación en el trastorno, la exposición y una tercera variable por separado se plantearon diferentes fases:

- Derivación del algoritmo a partir del cual se podía generar un sesgo de desclasificación no diferencial e independiente.
- Generación de las matrices de datos a partir de las que se evaluaría el sesgo de desclasificación en las tres variables, así como de las medidas de asociación.
- Generación de las matrices de datos incorrectamente clasificadas, y en consecuencia, las medidas de asociación obtenidas a partir de los datos no correctamente clasificados.

Para la derivación del algoritmo a partir del cual se generaron las tablas de datos incorrectamente clasificados se partió del supuesto de que las tres variables tenían exclusivamente dos niveles, por lo que los parámetros que definen el test diagnóstico bajo esta condición son la sensibilidad y la especificidad. Por otra parte, se partió también del supuesto de que el sesgo a generar es no diferencial e independiente. Así pues, si se produce una clasificación incorrecta es porque cualquier sujeto puede ser clasificado en una de las ocho casillas posibles de la tabla resumen de datos (ver tabla 1) y ello depende de la sensibilidad y la especificidad del ins-

trumento diagnóstico utilizado en cada una de las tres variables a evaluar.

Así pues, que un sujeto sea clasificado como expuesto, que tiene el trastorno y que está en el primer nivel de la tercera variable (A_1) depende de la probabilidad de que un sujeto que verdaderamente está expuesto, tiene el trastorno y está en el primer nivel de la tercera variable (A_1) sea clasificado como tal, más la probabilidad de que un sujeto que no está expuesto, tiene el trastorno y está en el primer nivel de la tercera variable (B_1), sea clasificado como que está expuesto, que tiene el trastorno y que está en el primer nivel de la tercera variable (A'_1), y así sucesivamente hasta la última casilla, lo que algebraicamente quedaría expresado como:

$$A_1' = A_1 \cdot p(T'E'C/TB) + B_1 \cdot p(T'E'C/\bar{T}\bar{B}) + C_1 \cdot p(T'E'C/\bar{T}E\bar{C}) + D_1 \cdot p(T'E'C/\bar{T}\bar{E}\bar{C}) + A_0 \cdot p(T'E'C/T\bar{B}) + B_0 \cdot p(T'E'C/T\bar{E}) + C_0 \cdot p(T'E'C/T\bar{E}\bar{C}) + D_0 \cdot p(T'E'C/T\bar{E}\bar{C}) = A_1 \cdot S_T \cdot S_E \cdot S_C + B_1 \cdot S_T \cdot (1 - E_E) \cdot S_C + C_1 \cdot (1 - E_T) \cdot S_E \cdot S_C + D_1 \cdot (1 - E_T) \cdot (1 - E_E) \cdot S_C + A_0 \cdot S_T \cdot S_E \cdot (1 - E_C) + B_0 \cdot S_T \cdot (1 - E_E) \cdot (1 - E_C) + C_0 \cdot (1 - E_T) \cdot S_E \cdot (1 - E_C) + D_0 \cdot (1 - E_T) \cdot (1 - E_E) \cdot (1 - E_C)$$

Donde:

- T'E' y C' son los sujetos que se clasifican como que tienen el trastorno, están expuestos y están en el primer nivel de la tercera variable.

- T, E y C sujeto que tiene el trastorno, está expuesto y está en el primer nivel de la tercera variable.

- \bar{T} , \bar{E} y \bar{C} sujeto que se clasifica como que no tiene el trastorno, no está expuesto y está en el segundo nivel de la tercera variable.

- $\bar{T}\bar{E}$ y \bar{C} sujeto que no tiene el trastorno, no está expuesto y está en el segundo nivel de la tercera variable.

- S_T y E_T son la sensibilidad y la especificidad para el trastorno.

- S_E y E_E son la sensibilidad y la especificidad para la exposición.

- S_C y E_C son la sensibilidad y la especificidad para la tercera variable.

En el caso que nos ocupa, la desclasificación para el trastorno, la exposición y la tercera variable se evalúa por separado y en consecuencia, la expresión anterior quedaría reducida a:

- Desclasificación del trastorno: $A_1' = A_1 \cdot S_T + C_1 \cdot (1 - E_T)$

- Desclasificación de la exposición: $A_1' = A_1 \cdot S_E + B_1 \cdot (1 - E_E)$

- Desclasificación de la tercera variable: $A_1' = A_1 \cdot S_C + A_0 \cdot (1 - E_C)$

Para las siete casillas restantes la lógica de la derivación es exactamente igual por lo que ya no se reproduce aquí.

Una vez derivadas las fórmulas que nos permitirían generar la desclasificación en los datos correctamente clasificados se proce-

		Sí exposición (1)	No exposición (0)
Tercera Variable (1)	Si trastorno (1)	A ₁	B ₁
	No trastorno (0)	C ₁	D ₁
Tercera Variable (0)	Si trastorno (1)	A ₀	B ₀
	No trastorno (0)	C ₀	D ₀

dió a generar las matrices en las que generaríamos el sesgo de desclasificación no diferencial. Ello se hizo teniendo en cuenta el valor del verdadero riesgo relativo para la tabla de datos brutos (niveles: no relación, RR = 1, la exposición es un factor de riesgo, RR ≅ 3 y RR ≅ 9, y la exposición es un factor de protección, RR ≅ 0.5 y RR ≅ 0.25) la prevalencia del trastorno (niveles: 0.01, 0.05, 0.1, 0.15 y 0.2, no se consideraron niveles superiores porque en el ámbito de la psicología es muy rara la existencia de trastornos con una prevalencia superior, aunque al introducir la tercera variable estas prevalencias no se mantenían en los estratos a excepción de las situaciones en las que en realidad esta tercera variable no afectaba en la que se mantenía la prevalencia del trastorno de la tabla 2 x 2 en los dos estratos) y como podía afectar la tercera variable a la relación entre el trastorno y el factor de exposición (niveles: no afecta, es un modificador del efecto o es un confusor). El tamaño de muestra utilizado fue de 20.000 en todas las matrices y la prevalencia de la exposición se fijó en 0.2 para las tablas de datos brutos, aunque al introducir la tercera variable en realidad la prevalencia de la exposición fue de 0.06 para el primer nivel de esta variable y de 0.6 para el segundo nivel de esta variable. La combinación de los distintos niveles para los tres parámetros manipulados, valor del verdadero riesgo relativo, valor de la prevalencia del trastorno y efecto de la tercera variable en la relación entre el trastorno y el factor de confusión, dio lugar a un total de 75 matrices de datos diferentes a partir de las que se estudió el sesgo de desclasificación no diferencial e independiente (ver tabla 2).

Finalmente, se generaron las matrices de datos incorrectamente clasificados a partir de la implementación del algoritmo derivado para obtener el número de sujetos incorrectamente clasificados en cada una de las ocho casillas de la tabla resumen de datos. Para ello, se realizó un programa elaborado 'ad hoc' en lenguaje de

Tabla 2
Combinaciones de las situaciones bajo las que se han generado las 75 matrices de datos a partir de las que se ha evaluado el sesgo de desclasificación

Tipo de relación entre el trastorno y la exposición		Tipo de efecto de la tercera variable en la relación entre trastorno y exposición	Niveles p(t)
No relación	RR ¹ = 1	No Afecta: RR ₁ = 1 y RR ₂ = 1 M.E. ³ : RR ₁ = 2.72 y RR ₂ = 1.11 Confusor: RR ₁ = 2.6327 y RR ₂ = 2.66	
Factor de Riesgo	RR = 3.017	No Afecta: RR ₁ = 3.017 y RR ₂ = 3.017 M.E.: RR ₁ = 7.45 y RR ₂ = 1.12 Confusor: RR ₁ = 1.2 y RR ₂ = 1.21	0.05 ²
	RR = 9.33	No Afecta: RR ₁ = 9.33 y RR ₂ = 9.33 M.E.: RR ₁ = 14 y RR ₂ = 3.33 Confusor: RR ₁ = 3.783 y RR ₂ = 3.768	0.01 0.1 0.15
Factor de Protección	RR = 0.494	No Afecta: RR ₁ = 0.494 y RR ₂ = 0.494 M.E.: RR ₁ = 0.795 y RR ₂ = 0.126 Confusor: RR ₁ = 0.231 y RR ₂ = 0.234	0.2
	RR = 0.255	No Afecta: RR ₁ = 0.255 y RR ₂ = 0.255 M.E.: RR ₁ = 0.583 y RR ₂ = 0.068 Confusor: RR ₁ = 0.469 y RR ₂ = 0.44	

¹ RR es la nomenclatura utilizada para expresar la medida de asociación riesgo relativo.

² Para la prevalencia del trastorno de 0.05 en el caso de no relación para la situación en que la tercera variable es un confusor en realidad el RR₁ es igual a 2.33 y el RR₂ es igual a 2.33.

³ M.E. quiere decir modificador del efecto.

programación Turbo Pascal 7.0. El programa construido, además de proporcionar las matrices de datos con los sujetos incorrectamente clasificados, calculaba el riesgo relativo de la tabla de datos brutos y de cada uno de los dos estratos para los datos mal clasificados. Los niveles de sensibilidad y especificidad utilizados tanto para el trastorno, como para la exposición como para la tercera variable oscilaron de 0.7 a 1 con un incremento de 0.1, lo cual dio lugar a un total de 16 condiciones para cada una de las 75 matrices de datos correctamente clasificados. El límite inferior para la sensibilidad y la especificidad se fijó en 0.7 ya que si estos valores fueran inferiores el test diagnóstico utilizado no sería adecuado ni para la exclusión del diagnóstico ni para la detección del trastorno, así por ejemplo Peeters (1991) apunta que si se da una desclasificación del 60% sería mejor utilizar una moneda para clasificar a los sujetos en una de las dos categorías del factor a evaluar, por otra parte en un trabajo anterior sobre la valoración de los valores predictivos se utilizó ya como valor mínimo de sensibilidad y especificidad del test diagnóstico el 0.6 (Peró y Guàrdia, 1995).

Resultados

Los resultados obtenidos se comentaran diferenciando en tres apartados, el efecto de la desclasificación no diferencial del trastorno en el riesgo relativo, el efecto de la desclasificación no diferencial de la exposición en el riesgo relativo y el efecto de la desclasificación no diferencial de la tercera variable en el riesgo relativo.

Trastorno

En la situación en la que no hay relación entre el factor de exposición y el trastorno y la tercera variable no afecta a esta relación no se produce sesgo en el riesgo relativo de la tabla de datos brutos ni en los de los estratos. Ahora bien, si el factor de exposición es un factor de riesgo o de protección, el sesgo es siempre hacia la hipótesis nula, a excepción de que la especificidad del instrumento diagnóstico para el trastorno sea igual a 1, situación en la que no se produce sesgo en la estimación del riesgo relativo. Además, cabe destacar que en las situaciones en las que se produce sesgo, éste disminuye al aumentar el valor de la prevalencia del trastorno y los de sensibilidad y especificidad del instrumento diagnóstico, siendo mayor la influencia de esta última.

Por otra parte, cuando la tercera variable es en realidad un modificador del efecto o un confusor el sesgo en el riesgo relativo es siempre hacia la hipótesis nula, a excepción de en las tablas de datos brutos cuando no existe relación entre el trastorno y el factor de exposición y en cualquier situación en la que la especificidad del instrumento diagnóstico para el trastorno es igual a 1. Nuevamente, cuando se produce sesgo, éste disminuye al aumentar el valor de la prevalencia del trastorno y los valores de sensibilidad y especificidad del instrumento diagnóstico, siendo superior la influencia de ésta última.

Factor de Exposición

Para empezar este subapartado se ha de destacar el hecho de que la prevalencia del trastorno no influye en el sesgo en el riesgo relativo cuando se clasifica incorrectamente la exposición, ya que, sea cual sea la prevalencia del trastorno el valor del riesgo relativo se mantiene constante bajo unos valores determinados de sensibilidad y especificidad del test diagnóstico.

En las condiciones en las que no existe relación entre el trastorno y el factor de exposición y la tercera variable no afecta a esta relación no se produce sesgo en el riesgo relativo. Ahora bien, si el factor de exposición es un factor de riesgo o de protección el sesgo es siempre hacia la hipótesis nula, disminuyendo al aumentar los valores de sensibilidad y especificidad del instrumento diagnóstico para la exposición, siendo mayor la influencia de esta última, al igual que pasaba para el trastorno.

Finalmente, cuando la tercera variable es un confusor o un modificador del efecto, si no existe relación entre trastorno y exposición no se produce sesgo en la estimación del riesgo relativo, pero para el resto de condiciones sí (hacia la hipótesis nula), y al igual que en el caso anterior, éste disminuye al aumentar la sensibilidad y especificidad de la prueba diagnóstica. Además el sesgo es superior en el primer estrato, y ello es debido al hecho de que la prevalencia de la exposición es inferior ($p_{e1} = 0.06$ y $p_{e2} = 0.6$).

Tercera Variable

En el caso del sesgo de desclasificación producido en la tercera variable, y al igual que sucedía para el factor de exposición, debe remarcarse el hecho de que el sesgo en la medida de asociación calculada es igual independientemente de la prevalencia del trastorno de las tablas de datos brutos.

No se produce sesgo en el riesgo relativo al generar un sesgo de desclasificación en la tercera variable cuando ésta no afecta a la relación entre el trastorno y la exposición, en las tablas de datos brutos cuando la tercera variable es un factor de confusión o un modificador del efecto, y finalmente cuando la especificidad es igual a 1 para el riesgo relativo del primer estrato o cuando la sensibilidad es igual a 1 para el riesgo relativo del segundo estrato. En el resto de condiciones el sesgo puede ser hacia la hipótesis nula o lejos de la hipótesis nula.

Así pues, cuando no existe relación entre trastorno y exposición en la tabla de datos brutos (2×2), pero la tercera variable es un factor de confusión o un modificador del efecto se produce un sesgo hacia la hipótesis nula en el riesgo relativo del primer estrato, disminuyendo al aumentar la sensibilidad y la especificidad del instrumento diagnóstico de la tercera variable (mayor influencia de especificidad), en tanto que se produce una inversión del efecto en el riesgo relativo del segundo estrato. Lo cual nos llevaría a concluir que la tercera variable está actuando siempre como un modificador del efecto.

Cuando la exposición es un factor de riesgo y la tercera variable es un factor de confusión, en las condiciones en las que se produce sesgo, éste es siempre hacia la hipótesis nula para el riesgo relativo del primer estrato y lejos de la hipótesis nula para el riesgo relativo del segundo estrato, lo cual nos lleva a la conclusión errónea de que la variable es un modificador del efecto. Nuevamente, el sesgo es inferior al aumentar los valores de sensibilidad y especificidad, aunque, para el riesgo relativo del primer estrato tiene mayor influencia la sensibilidad, en tanto que para el riesgo relativo del segundo estrato tiene mayor influencia la especificidad.

Si la tercera variable es un modificador del efecto y la exposición un factor de riesgo, se produce una sobreestimación del riesgo relativo, tanto el del primer estrato como el del segundo, siendo mayor el sesgo en el primero. Este sesgo disminuye al aumentar los valores de sensibilidad y especificidad.

En aquellas situaciones en las que la exposición es un factor de protección y la tercera variable un modificador del efecto el sesgo

es lejos de la hipótesis nula para el riesgo relativo del primer estrato y hacia la hipótesis nula para el riesgo relativo del segundo estrato. Ahora bien, si la tercera variable es un confusor, el sesgo es hacia la hipótesis nula en las condiciones en las que el riesgo relativo de la tabla de datos brutos es de 0.494 y lejos de la hipótesis nula en las condiciones en las que el riesgo relativo de la tabla de datos brutos es de 0.225. En ambas situaciones la especificidad tiene mayor influencia en el primer estrato, y la sensibilidad en el segundo estrato.

Discusión

A partir de los resultados observados, se puede concluir que un sesgo de desclasificación no diferencial en el trastorno cuando éste es dicotómico, lleva a un sesgo hacia la hipótesis nula en el riesgo relativo, conclusión que coincide con las obtenidas en otros trabajos (Kleinbaum, Kupper y Morgenstern, 1982, Flegal, Brownie y Hass, 1986). Ahora bien, en el presente trabajo se ha detectado que cuando el valor de la especificidad es igual a uno la estimación de la medida de asociación es perfecta. Por otra parte, también cabe destacar el hecho de que la prevalencia del trastorno tiene un papel muy importante en la cantidad de sesgo generado en la medida de asociación, así como la sensibilidad y la especificidad del test diagnóstico para determinar el estatus de trastorno; así pues, cuanto mayor sea la prevalencia del trastorno menor será el sesgo, y a medida que aumentan los valores de sensibilidad y de especificidad el sesgo en el riesgo relativo disminuye, siendo mayor el efecto de la especificidad.

Para el caso de la desclasificación no diferencial del factor de exposición y dejando de lado la polémica generada cuando éste es politémico o continuo (Dosemeci, Wacholder y Lubin, 1990, 1991, 1992, Weinberg, Umbach y Greenland, 1994, Thomas, 1995, Yuen Fung y Howe, 1984, Brenner, 1991 y Myers y Ehrlich, 1992), al igual que han mostrado otros autores (Kleinbaum, Kupper y Morgenstern, 1982 y Flegal, Brownie y Hass, 1986) cuando el factor de exposición es dicotómico el sesgo en el riesgo relativo es hacia la hipótesis nula. Además como se desprende de los resultados obtenidos el sesgo en la medida de asociación disminuye a medida que aumentan los valores de sensibilidad y especificidad teniendo mayor influencia la especificidad tal y como sucedía en el sesgo de desclasificación generado para el trastorno. Por otra parte, independientemente del verdadero valor del riesgo relativo y de que la tercera variable sea un modificador del efecto o un factor de confusión, cuando en la tabla de datos brutos se detecta relación entre el trastorno y el factor de exposición el sesgo generado en la medida de asociación es mayor en el primer estrato que en el segundo y ello es debido al hecho de que la prevalencia de la exposición es menor en el primer estrato, tal y como remarcaban Flegal, Brownie y Hass (1986).

En la situación en la que se ha generado el sesgo de desclasificación no diferencial en la tercera variable, el patrón en el sesgo producido en el riesgo relativo es menos claro tal y como apuntan los diferentes trabajos realizados en este ámbito (Greenland, 1980, Yuen Fung y Howe, 1984, Kupper, 1984 citado en Sawitz y Baron, 1989 y Cox y Elwood, 1991). Así pues, tal y como señalan estos autores, el sesgo en el riesgo relativo puede ser tanto hacia la hipótesis nula como lejos de ella. De todos modos, nuestros resultados muestran que el riesgo relativo no queda sesgado cuando la tercera variable no afecta; el mismo resultado se obtiene cuando en realidad la tercera variable es un modificador del efecto o un fac-

tor de confusión para las tablas de datos brutos, así como para el riesgo relativo del primer estrato cuando el valor de especificidad es igual a uno independientemente del valor de la sensibilidad o para el riesgo relativo del segundo estrato cuando la sensibilidad es igual a uno independientemente del valor de la especificidad, resultado similar al obtenido por Cox y Elwood (1991) para la «odds ratio» aunque en la situación inversa ya que ellos codifican como 0 el primer estrato y como 1 el segundo, al revés de cómo se ha realizado en el presente trabajo. Para el resto de situaciones, se produce un sesgo en el riesgo relativo que puede ser tanto hacia la hipótesis nula como lejos de ella. Cuando la tercera variable no afecta a la relación entre el trastorno y el factor de exposición o bien cuando en realidad es un modificador del efecto, tanto si no existe relación como si el factor de exposición es un factor de riesgo, así como cuando no existe relación entre trastorno y exposición y la tercera variable es en realidad un factor de confusión, el riesgo relativo del primer estrato queda sesgado hacia la hipótesis nula siendo mayor la influencia de la especificidad en tanto que se produce una inversión en la estimación del efecto en el segundo nivel de la tercera variable, llegando a la conclusión de que el factor de exposición actúa como un factor de protección cuando en realidad es de riesgo, además para esta situación la sensibilidad tiene mayor influencia en el sesgo en la medida de asociación que la especificidad, por lo que en el caso de que la tercera variable sea en realidad un factor de confusión se llega a la conclusión incorrecta de que es un modificador del efecto. Ahora bien, si el factor de exposición es factor protector y la tercera variable es un modificador del efecto el riesgo relativo del primer estrato queda sobreestimado en tanto que el riesgo relativo del segundo estrato va hacia la hipótesis nula, siendo mayor la influencia de la especificidad en el primer estrato y de la sensibilidad en el segundo estrato. Por otra parte, cuando el factor de exposición es un factor de riesgo y la tercera variable es en realidad un factor de confusión los riesgos relativos de los dos estratos quedan sobreestimados pero el sesgo generado no es igual de grande en los dos estratos, lo que nos lleva a la falsa conclusión de que la variable es un modificador del efecto. Finalmente, cuando el factor de exposición es un factor protector y la tercera variable es en realidad un factor de confusión, en el caso de que el riesgo relativo de la tabla 2×2 sea igual a 0.494, se produce un sesgo en los riesgos relativos de los estratos hacia la hipótesis nula, en tanto que cuando el riesgo relativo de la tabla 2×2 es de 0.255 el sesgo en los riesgos relativos de los estratos provoca una sobreestimación de la asociación, este comportamiento diferencial puede ser debido al hecho de que en la primera situación existe confusión porque los riesgos relativos de los estratos son inferiores al de la tabla 2×2 , en tanto que en el segundo caso es la situación contraria. En consecuencia, la afirmación realizada por Cox y Elwood (1991) de que el sesgo producido en la medida de asociación tiene el mismo patrón tanto si la tercera variable es un modificador del efecto como si es un factor de confusión, sólo es parcialmente válido, especialmente en las situaciones en las que la estimación del riesgo relativo no está sesgada.

A partir de los resultados observados para la desclasificación no diferencial en las tres variables, se puede llegar a la conclusión que cuando el sesgo generado implica una tendencia hacia la hipótesis nula del riesgo relativo existe una mayor influencia del valor de la especificidad del test diagnóstico que no del valor de la sensibilidad, sea cual sea la variable en la que se ha generado inicialmente la desclasificación, trastorno, factor de exposición o tercera variable.

Este estudio nos ha permitido aclarar de que modo afecta el sesgo de desclasificación no diferencial por separado para el trastorno, el factor de exposición y una tercera variable, ahora bien, se ha hecho el supuesto de que cuando una de las tres variables queda desclasificada en las otras dos no ocurre sesgo de

desclasificación no diferencial, aspecto bastante difícil de creer en la realidad por lo que posteriores estudios tendrían que ir en la línea de determinar como afecta el sesgo de desclasificación no diferencial independiente generado en las tres variables a la vez.

Referencias

- Brenner, Hermann (1991). «Re: Does nondifferential misclassification of exposure always bias a true effect toward the null value?». *American Journal of Epidemiology*, 1991 134 (4), 438-439.
- Birkett, Nicholas J. (1995). «Re: When will nondifferential misclassification of an exposure preserve the direction of a trend?» (Letter). *American Journal of Epidemiology*, 142 (7), 783-784.
- Cox, Brian, y Elwood, J. Mark. (1991). The effect on the stratum-specific odds ratios of nondifferential misclassification of a dichotomous covariate. *American Journal of Epidemiology*, 133 (2), 202-207.
- Delpizzo, Vincent, y Borghesi, Joanna. (1995). Exposure measurement errors, risk estimate and statistical power in case-control studies using dichotomous analysis of a continuous exposure variable. *International Journal of Epidemiology*, 24 (4), 851-862.
- Dosemeci, Mustafa, Wacholder, Sholom, y Lubin, Jay H. (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value?. *American Journal of Epidemiology*, 132 (4), 746-748.
- Dosemeci, Mustafa, Wacholder, Sholom, y Lubin, Jay H. (1991). «Re: Does nondifferential misclassification of exposure always bias a true effect toward the null value?». *American Journal of Epidemiology*, 1991 134 (4), 441-442.
- Dosemeci, Mustafa, Wacholder, Sholom, y Lubin, Jay H. (1992). «Re: Does nondifferential misclassification of exposure always bias a true effect toward the null value?». *American Journal of Epidemiology*, 1991 135 (12), 1.430-1.431.
- Flegal, Katherine, Brownie, Clavel, y Haas, Jere D. (1986). The effects of exposure misclassification on estimates of relative risk. *American Journal of Epidemiology*, 123 (4), 736-751.
- Greenland, Sander (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology*, 112 (4), 564-569.
- Hsieh, Chung-Cheng, y Walter, Stephen D. (1988). The effect of non-differential exposure misclassification on estimates of the attributable and prevented fraction. *Statistics in Medicine*, 7, 1.073-1.085.
- Hsieh, Chung-Cheng (1991). The effect on non-differential outcome misclassification on estimates of the attributable and prevented fraction. *Statistics in Medicine*, 10, 361-373.
- Kleinbaum, D.G., Kupper, L.L., y Morgenstern, H. (1982). *Epidemiologic Research. Principles and quantitative methods*. London: Lifetime Learning Publications.
- Myers, J. Y Ehrlich, R. (1992). «Re: Does nondifferential misclassification of exposure always bias a true effect toward the null value?». *American Journal of Epidemiology*, 1991 135 (12), 1.429-1.430.
- Peeters, Petra H.M. (1991). Re: «Does nondifferential misclassification of exposure always bias a true effect toward the null value?» (Letter) *American Journal of Epidemiology*, 134 (4), 439-440.
- Peró, Maribel, y Guàrdia, Joan. (1995, Abril). *Efecto de la prevalencia, la sensibilidad y la especificidad en los valores predictivos de los tests diagnósticos*. Comunicación presentada en el 'IV simposium de Metodología de las Ciencias del Comportamiento', Murcia, España.
- Rothman, K.J. (1986). *Epidemiología moderna*. Madrid: Díaz de Santos.
- Sackett, D.L., Haynes, R.B., y Tugwell, R. (1989). *Epidemiología clínica. Una ciencia básica para la medicina clínica*. Madrid: Ediciones Díaz de Santos, S.A.
- Savitz, David A., y Barón, Anna E. (1989). Estimating and correcting for confounder misclassification. *American Journal of Epidemiology*, 129 (5), 1.062-1.071.
- Thomas, Duncan C. (1995). «Re: When will nondifferential misclassification of an exposure preserve the direction of a trend?» (Letter). *American Journal of Epidemiology*, 142 (7), 782-783.
- Weinberg, Clarice R., Umbach, David M., y Greenland, Sander (1994). When will nondifferential misclassification of an exposure preserve the direction of a trend? *American Journal of Epidemiology*, 140 (6), 565-571.
- Weinberg, Clarice R., Umbach, David M., y Greenland, Sander. (1995). «Re: When will nondifferential misclassification of an exposure preserve the direction of a trend?» (Letter). *American Journal of Epidemiology*, 142 (7), 784.
- Yuen Fung, Karen, y Howe, Geoffrey R. (1984). Methodological issues in case-control studies III: - The effect of joint misclassification of risk factors and confounding factors upon estimation and power. *International Journal of Epidemiology*, 13 (3), 366-370.