

El control de sesgos en la investigación cuantitativa enfermera

¹Rafael del Pino Casado

²Antonio Frías Osuna

³Pedro A. Palomino Moral

¹ Enfermero. Licenciado en Antropología y Doctor por la Universidad de Jaén. Profesor Contratado Doctor. Departamento de Enfermería, Universidad de Jaén.

² Enfermero. Licenciado en Antropología y Doctor por la Universidad de Granada. Profesor Titular de Universidad. Departamento de Enfermería, Universidad de Jaén.

³ Enfermero. Licenciado en Psicopedagogía y Doctor por la UNED. Profesor Titular de Universidad. Departamento de Enfermería, Universidad de Jaén.

Dirección de correspondencia:
Rafael del Pino Casado. Departamento de Enfermería.
Universidad de Jaén. Campus Las Lagunillas, s/n. 23071 Jaén.

E-mail: rdelpino@ujaen.es

Resumen

El presente artículo tiene como finalidad la exposición y análisis de un tema importante en la investigación cuantitativa enfermera: el control de los sesgos. Para ello se realiza un recorrido conceptual por los distintos tipos de sesgos y se analiza en profundidad las distintas formas de evitarlos al diseñar, ejecutar y analizar los datos de una determinada investigación.

Con este documento se pretende generar una herramienta de trabajo para el investigador enfermero novel que maneja los conceptos básicos de investigación, pero que necesita ir más allá de dichos conceptos y adentrarse en el terreno de la aplicación práctica.

Interesa destacar como puntos fuertes de este artículo: el reunir y actualizar en un solo documento aspectos que están dispersos en la literatura, la orientación práctica de los contenidos abordados y la ejemplificación de situaciones que, por un lado, ayudan a la comprensión de los mencionados contenidos y por otro, están encuadradas en situaciones de investigación de Enfermería Comunitaria.

Palabras clave: investigación cualitativa; sesgos; Enfermería Comunitaria.

Abstract

Bias control in quantitative nursing research

This article aims to expose and to analyze an important subject in quantitative nursing research: bias control. To this end, a conceptual review has been made according to the type of bias as well as an in-depth analysis of the different ways to avoid these biases when designing, implementing and analyzing the data of a given research.

This document is intended to generate a working tool for the novel nursing researcher that handles basic research concepts but who needs to go beyond such concepts and gain further insight into practical applications.

It is of interest for us to highlight as strong points of this article: the gathering and updating of aspects that are dispersed in the literature into one single document, the practical orientation of the contents approached, and the exemplification of situations that, on the one side, help to understand the said contents and, on the other, are framed within research situations in Community Nursing.

Key words: quantitative research; bias; Community Nursing.

Introducción

Que el desarrollo de la investigación enfermera de calidad pasa inevitablemente por una adecuada formación es una proposición que podría catalogarse de tautológica. La actual proliferación de manuales y recursos web sobre metodología de investigación cuantitativa podría hacer pensar que la necesidad de apoyo documental a dicha formación en investigación cuantitativa está más que cubierta. La experiencia docente de los autores de este artículo apunta en sentido contrario, ya que, en la práctica, muchos de los manuales y recursos sobre este tema adolecen de dos déficits en lo que respecta a la formación básica: a) hablan de qué hay que hacer, pero no de por qué hacerlo y cómo hacerlo; b) los niveles de dificultad se sitúan en extremos opuestos: o tan básicos que apenas aportan nada o de un nivel que hace difícil su comprensión a iniciados. Por otro lado, de todos los aspectos que se necesitan conocer en profundidad sobre investigación, el apartado de sesgos adquiere una importancia capital, ya que de su conocimiento depende su prevención y control y por tanto, la veracidad de las conclusiones de un estudio. Es por ello que hemos decidido escribir el presente artículo como contribución a una formación de calidad en investigación enfermera, con la finalidad de analizar los posibles sesgos de una investigación cuantitativa y proponer recomendaciones de cómo evitarlos. Para ello, se recordará en primer lugar algunos conceptos básicos relacionados con la temática abordada, para pasar a analizar a continuación los distintos tipos de sesgos y su control. Para una mejor comprensión de los distintos sesgos se basarán en varios ejemplos, la mayoría de los cuales pertenecen a una interesante línea de investigación enfermera: la investigación sobre el cuidado no retribuido de los mayores dependientes.

Conceptos básicos

Se denomina "error sistemático" a aquel que está relacionado con el diseño del estudio y su ejecución (la selección de la población de estudio, la medición de las variables de interés y el análisis de los datos). A los errores sistemáticos se les denomina también "sesgos". Al ser debidos al diseño y ejecución del estudio, los sesgos son controlables (prevenibles) mediante la mejora de dichos procesos, a diferencia del error aleatorio que no puede prevenirse, ya que se debe a la variabilidad de los fenómenos bio-psico-sociales, unida al hecho de que siempre que se investiga se trabaja con muestras de individuos y no con toda la población posible.

La validez interna de un estudio es el grado de ausencia de errores sistemáticos o sesgos y se relaciona con la capacidad del estudio para observar y medir lo que realmente se

quiere observar y medir y por tanto, con el grado de confianza que se puede tener en los resultados de una investigación.

Cuando un estudio de investigación cuantitativa posee validez interna y precisión*, se dice que tiene validez externa. Dicha validez externa hace alusión a la capacidad de generalizar o extrapolar los resultados obtenidos a diferentes personas, contextos y momentos temporales.

En general, la validez (tanto externa como interna) hace referencia al grado de seguridad o confianza que se puede tener sobre de las conclusiones de una investigación y de su aplicación a otros contextos.

Existen tres tipos de sesgos: de selección, de clasificación o información y de confusión. En términos generales, el sesgo de selección está relacionado con la falta representatividad de la muestra estudiada, el sesgo de clasificación con errores en la medición y el sesgo de confusión se relaciona con la distorsión de la asociación estudiada provocada por la influencia de una tercera variable. En la Figura 1 se resumen las consecuencias de dichos sesgos. Se puede observar seguidamente con mayor detalle cada uno de estos sesgos y sus consecuencias, así como las distintas formas de controlarlos.

Sesgo de selección

Concepto y clasificación

Se denomina así a un tipo de error sistemático que tiene su origen en la falta de representatividad de la muestra estudiada y que tiene como consecuencias la limitación de la extrapolación de las conclusiones del estudio y/o la aparición de asociaciones espurias (inexistentes) y/o la alteración de las asociaciones existentes (ocultamiento de asociaciones reales o distorsión de éstas por infra o sobreestimación del efecto).

Una muestra es representativa de la población de referencia de donde se ha extraído cuando las variables que influyen en el problema de estudio se distribuyen en la muestra de igual forma que en la población de referencia. El sesgo de selección aparece, por tanto, cuando las personas incluidas en un estudio difieren en algunas características o variables relevantes, respecto de la población a la que se quiere aplicar las conclusiones del estudio.

El sesgo de selección puede producirse en la fase de planificación del estudio por errores en la formación de la muestra (sesgos relacionados con la formación de la muestra), y/o en la fase de ejecución al recoger y analizar los datos del mismo (sesgos relacionados con la integridad de la mues-

*La precisión mide el grado de aproximación de los valores obtenidos en un estudio a los valores reales existentes en la población de referencia y por tanto, estima el grado de ausencia de error aleatorio. La precisión se mide mediante la amplitud de los intervalos de confianza de las medidas del estudio.

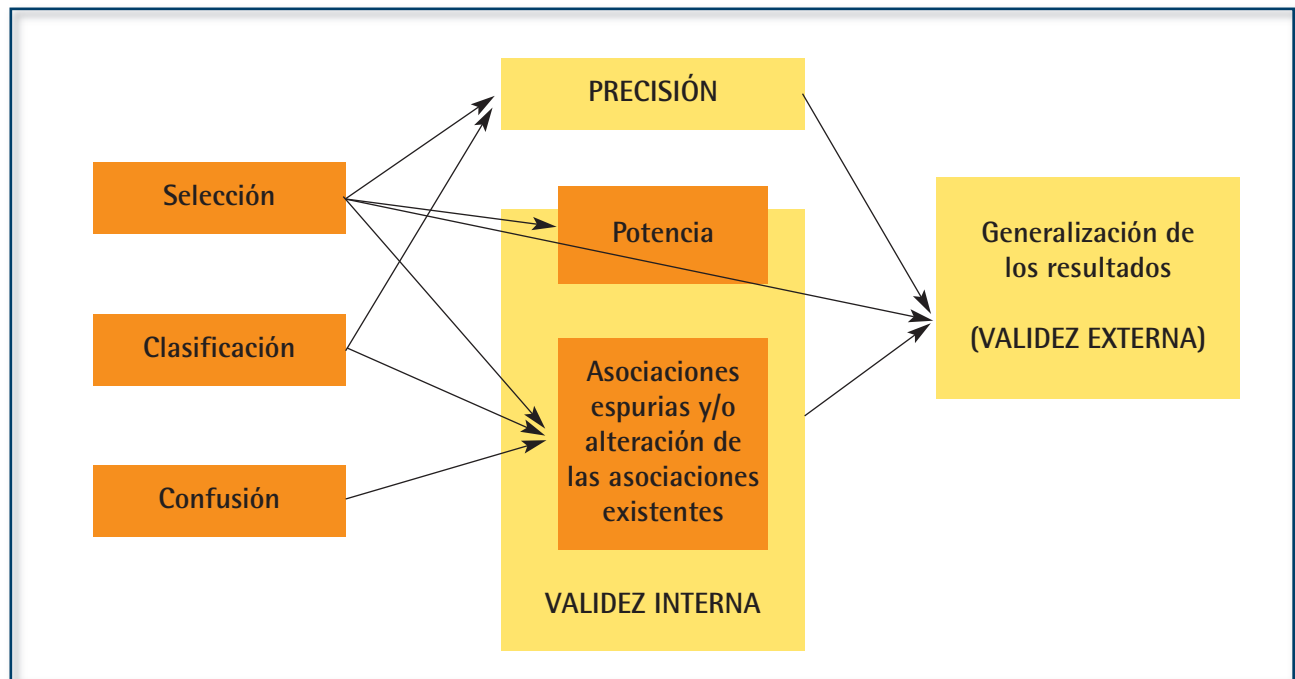


Figura 1. Consecuencias de los sesgos

tra). En la Figura 2 aparecen los distintos tipos de sesgos de selección que se explican a continuación.

Los sesgos relacionados con la formación de la muestra pueden deberse a errores en la definición de la población de

referencia y a errores en la selección de la muestra (1). Dentro de los posibles errores en la definición de la población de referencia, destaca la utilización de casos del problema de estudio ingresados en el hospital (sesgo de Berkson), los cuales se toman como referente de la zona de influencia del hospital. Los errores en la selección pueden a su vez estar relacionados con la utilización de procedimientos de muestreo no probabilísticos (sesgo de muestreo) o con la sobrerrepresentación de casos de larga duración. Algunos casos concretos de sesgos de selección relacionados con el muestreo no probabilístico serían (1-3): estudios con voluntarios (sesgo del voluntariado o de autoselección), la selección diferencial (cuando la forma de seleccionar los sujetos condiciona la distribución de la exposición o el efecto en los grupos que se comparan) y la detección diferencial (cuando la presencia del factor de exposición favorece el diagnóstico del efecto). Dentro de la sobrerrepresentación de casos de larga duración estaría como caso particular la Falacia de Neyman o supervivencia selectiva, que se produce cuando el factor de exposición es a su vez factor pronóstico y condiciona una mayor mortalidad, con lo cual las personas estudiadas no son representativas de la población de referencia.

A continuación se observan algunos ejemplos de los anteriores tipos de sesgos. Respecto de la selección diferencial, se imagina que se quiere comprobar la hipótesis de que en personas cuidadoras existe mayor prevalencia de ansiedad que en población general, desarrollándose para ello un estudio descriptivo transversal, en una ciudad de 60.000 habitantes. Para ello se elige a las personas cuidadoras de ma-

Relacionados con la formación de la muestra:

- Inadecuada definición de la población de referencia:
 - Casos hospitalarios (sesgo de Berkson)
- Selección inadecuada de la muestra:
 - Muestreo no probabilístico:
 - Voluntarios (sesgo del voluntariado)
 - Selección diferencial
 - Detección diferencial
 - Sobrerrepresentación de casos de larga duración:
 - Supervivencia selectiva o falacia de Neyman

Relacionados con la integridad de la muestra:

- No respuesta (sesgo de no respuesta)
- Pérdidas (desgaste de la muestra)
- Datos perdidos
- Rechazo de datos

Figura 2. Clasificación de los sesgos de selección

yores dependientes a partir de las que acuden a un centro de día y a las personas no cuidadoras a partir de las personas de la misma edad y sexo que acuden a consulta de medicina general. Es este estudio se cometería un sesgo de selección que no solo afectaría a la generalización de los datos (ya que los grupos que se comparan no son representativos de toda la ciudad) sino que también conllevaría asociaciones espurias por selección diferencial, ya que la ansiedad en personas cuidadoras puede estar infraestimada en las que tienen posibilidad de respiro (al disponer de centro de día) y la ansiedad en población no cuidadora puede estar sobreestimada en aquellas personas que acuden a consultas de medicina general.

Respecto de la detección diferencial, se supone que se quiere estudiar la relación entre depresión y edad en una determinada localidad. Se podría diseñar un estudio descriptivo transversal en el que la muestra se seleccionara a partir de las historias clínicas del centro de salud. Si el hecho de ser mayor conllevara la inclusión en un examen de salud en el que estuviese incluida la detección sistemática de depresión, se encontraría una mayor frecuencia de diagnóstico de depresión en los mayores respecto de los adultos, lo que traería a la aparición de una asociación artificial entre edad y depresión. Así, no solo se tendría un problema de representatividad de la muestra (los usuarios del centro de salud no tienen por qué ser representativos de la población de referencia) sino también la identificación de una asociación espuria.

En relación con la sobrerrepresentación, ésta podría aparecer en un estudio descriptivo transversal donde se intentase determinar el nivel de ansiedad y depresión de las personas cuidadoras. En dicho estudio estarían infrarepresentadas aquellas personas cuidadoras más frágiles (y por tanto con mayor nivel de depresión y ansiedad), ya que en ellas el riesgo de abandono es mayor.

Los sesgos relacionados con la integridad de la muestra se pueden deber a pérdidas en el seguimiento (desgaste de la muestra), a la no respuesta en un estudio tipo encuesta (sesgo de no respuesta), a la presencia de datos perdidos y al rechazo de datos en la fase de análisis (datos anómalos [puntos extremos**] o rechazados de forma arbitraria) (3).

Como se ha comentado anteriormente, el sesgo de selección tiene consecuencias en dos planos de análisis: la posibilidad de extrapolar los resultados y la posibilidad de aparición de asociaciones espurias y/o de alteración de las asociaciones existentes. La posibilidad de extrapolar los resultados se ve condicionada por la falta de representatividad de la muestra estudiada, que impide la generalización de los resulta-

dos más allá de la propia muestra. El segundo aspecto se produce cuando la selección de los grupos a comparar o los errores en la recogida y/o procesamiento de la información condicionan la distribución de las variables del estudio, de manera que se alteran las relaciones entre dichas variables. En el caso concreto de diseños epidemiológicos, se produce una sobre o infrarepresentación del efecto estudiado en el grupo de las personas expuestas al factor estudiado y por tanto, una estimación del efecto en los sujetos estudiados distinta a la que se obtendría en la población de referencia. Además, los sesgos relacionados con errores en la recogida y/ análisis de la información tienen también consecuencias en la potencia estadística*** y la precisión del estudio, que van a disminuir al reducirse el tamaño muestral (Figura 1).

La falta de representatividad lleva necesariamente a la imposibilidad de generalizar los resultados de la investigación, si bien no es condición suficiente para generar asociaciones espurias o enmascaramiento de asociaciones causales. Por ejemplo, un estudio con voluntarios con un determinado problema no es representativo de la población general, pero si la condición de exposición no está ligada al hecho de ser voluntario, no tiene por qué haber consecuencias en la determinación de la asociación estadísticas. Éstas solo se darían cuando el motivo para presentarse voluntario al estudio estuviese asociado, por ejemplo, a una mayor exposición en los que tienen el problema e estudio. Cuando la autoselección conlleva una distribución diferencial de la variable de interés en los grupos que se comparan, se habla de participación diferencial (2). Morabia (4) hace referencia a este aspecto en un artículo metodológico específico de estudios de casos y controles, donde deja constancia que la falta de representatividad no tiene porque llevar necesariamente a una estimación sesgada de la odds ratio****.

Para entender mejor estas asociaciones espurias, se puede tomar en consideración a modo de símil el siguiente episodio ocurrido en la Segunda Guerra Mundial (5): los militares norteamericanos inspeccionaban las partes dañadas de los bombarderos a la vuelta de las misiones, para reforzarlas. En principio parece una decisión acertada, ya que mejoraría la resistencia de los aparatos, pero si se tiene en cuenta que una parte de los bombarderos no volvían (porque habían sido derribados) y que eran precisamente los que volvían los que presentaban daños en zonas menos vitales, se producía una decisión errónea, ya que lo único para lo que servía la inspección era para marcar zonas del aparato que no eran esenciales para su correcto funcionamiento. Así, la selección de la muestra de estudio (partes dañadas de los aviones) estaba afectada por la exposición (al fuego enemi-

** Valores de una variable cuantitativa que están muy por encima o por debajo del conjunto de los datos.

*** Capacidad del estudio de detectar la ausencia de asociación estadística.

**** Medida de la magnitud de la asociación o tamaño del efecto en los estudios de casos y controles.

go), de manera que dicha exposición es un factor pronóstico que origina una disminución de la exposición en las unidades que presentan el efecto estudiado. A este sesgo se le conoce, como ya se ha visto, como Falacia de Neyman.

En el terreno concreto de la investigación en el cuidado informal son frecuentes los estudios con muestras de conveniencia (6), siendo varios los autores que han advertido de las limitaciones a la generalización de los resultados cuando se emplean muestras no probabilísticas (7,8). Dichas limitaciones se basan en las características diferenciales de las personas que participan en estudios de muestras de conveniencia (autoselección), frente a los que no participan. Así, Dura y Kiecolt-Glaser (7) encontraron diferencias entre personas cuidadoras en domicilio frente a las que estaban en una institución y entre personas cuidadoras que participaban en un estudio con muestra de conveniencia frente a las que no participaban. En el primer caso, se tiene que las cuidadoras en el hogar invertían más tiempo de cuidado, afrontaban más problemas de comportamiento, cuidaban a más personas, tenían más problemas de depresión y las personas cuidadas tenían mayor nivel de dependencia. En el segundo caso, las personas cuidadoras que no participaban atendían a pacientes con mayor nivel de dependencia. De la misma manera, Pruchno y cols (8) compararon una muestra de personas cuidadoras obtenida mediante muestreo aleatorio con otra de conveniencia, encontrando diferencias entre ambas que pueden justificar el pensar que se alcanzan resultados distintos en cada caso.

Salvo los errores en la fase de análisis de datos y la inadecuada definición de la población de referencia, que son independientes del tipo de diseño empleado, el resto de sesgos de selección están relacionados con el tipo de diseño de investigación (Figura 3).

La aparición de asociaciones espurias o enmascaramiento de asociaciones reales como consecuencia de la inadecuada selección de la muestra están relacionadas con diseños transversales y de casos y controles. En los primeros predomina la sobrerrepresentación de casos de larga duración (con independencia del tipo de muestreo) y en los segundos la selección o detección diferencial por muestreo no probabilístico. Esto último se debe a que, en los estudios de casos y controles, el efecto ya se ha producido al iniciar la investigación y al seleccionar los casos y los controles se puede alterar la distribución de la exposición, es decir, que los casos y/o controles seleccionados estén más o menos expuestos que el total del casos y/o controles de la población de referencia. Este problema se da con bastante frecuencia ya que es habitual que en los estudios de casos y controles trabajen con controles elegidos de forma no probabilística. La selección y detección diferenciales pueden aparecer también en estudios descriptivos transversales con muestreo no probabilístico, cuando el criterio de muestreo sea tener o no el problema de estudio.

En lo que respecta a los errores relacionados con la integridad de la muestra, están relacionados con los diseños longitudinales prospectivos (las pérdidas) o investigaciones mediante encuestas (la no respuesta).

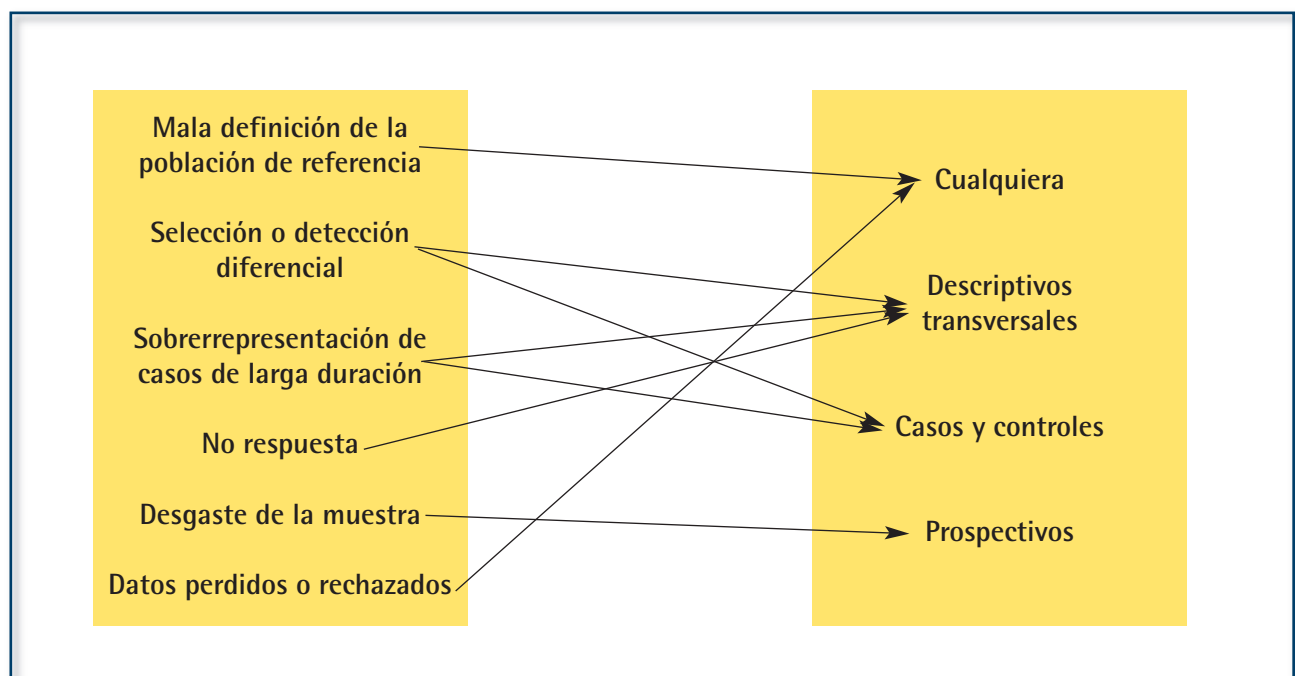


Figura 3. Sesgo de selección y tipo de diseño

El conjunto de limitaciones de un estudio relacionadas con la formación e integridad de la muestra se completa con el tamaño muestral insuficiente y con el empleo de restricciones en la definición de la muestra. Dicho conjunto de limitaciones y sesgos relacionados con la formación e integridad de la muestra, así como sus consecuencias, se muestran en la Figura 4.

Control del sesgo de selección

En términos generales, la mejor prevención del sesgo de selección es fijar adecuadamente la población de referencia y trabajar con muestras representativas de dicha población. Para lo primero, lo más adecuado es que el marco de muestreo (conjunto de elementos de donde se obtiene la muestra) coincida con la población de referencia. Por ejemplo, si se quiere pasar una encuesta sobre una zona geográfica determinada y se emplea como marco de muestreo la base de datos del sistema sanitario público, nuestra población de referencia no es la población del área geográfica, sino las personas de dicha área con cobertura del sistema sanitario público. Para lo segundo (muestras representativas), la opción de mínimos estaría representada por la utilización de una muestra probabilística de suficiente tamaño muestral. No obstante, hay situaciones en las que esto no es posible al no tener identificada la población de referencia. Un ejemplo de esto lo constituyen los estudios de casos y controles donde no son una muestra representativa del área de estudio (bien por falta de recursos o de forma intencionada para

que casos y controles sean comparables). En estos casos, la prevención del sesgo de selección, por lo menos en lo que a asociaciones espurias o enmascaramiento de asociaciones reales se refiere (ya que la extrapolación es imposible), pasa por (1,3): utilizar casos incidentes, igualar los criterios de selección y diagnóstico en los grupos a comparar, evitar que los controles conozcan la hipótesis del estudio para que su participación no dependa de la misma y confirmar que la exposición no influye en el diagnóstico del efecto.

Cuando la muestra no es representativa, con independencia del tipo de estudio, es necesario describir las características sociodemográficas de la muestra estudiada para mejorar en lo posible la escasa validez externa del estudio en cuestión. Esta medida es, no obstante, conveniente en cualquier situación, ya que favorece la validez externa y por tanto, la utilidad clínica de la investigación. Otro tema interesante en el caso de muestras no representativas, aunque a veces solo sea una cuestión de "higiene conceptual", es circunscribir las conclusiones del estudio a la muestra estudiada. Y con independencia de la representatividad de la muestra, siempre es una buena medida diagnosticar el efecto de manera enmascarada en relación con el nivel de exposición.

Por otro lado, cuando el problema no es la representatividad inicial de la muestra sino las posibles pérdidas en su seguimiento o la baja respuesta en un estudio tipo encuesta, es conveniente la sobreestimación del tamaño muestral en la fase de planificación del estudio, si bien esta medida, aunque puede evitar las repercusiones en la potencia y la precisión del estudio, no previene necesariamente el sesgo de se-

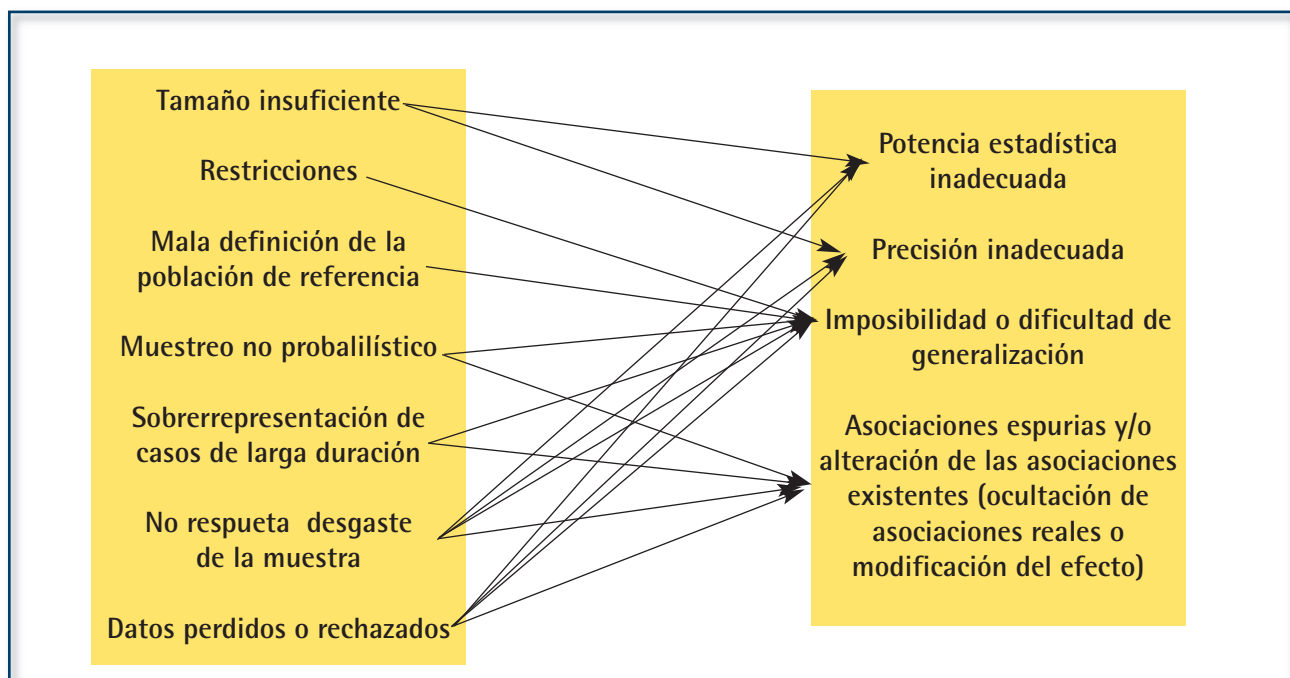


Figura 4. Limitaciones y sesgos relacionados con la forma e integridad de la muestra y sus consecuencias

lección, ya que las personas que abandonan o no responden pueden tener características diferenciales de las que continúan o responden. La sustitución muestral (reposición de elementos que no contestan) en estudios tipo encuesta conlleva similares reflexiones. En estudios con seguimiento es importante contar con información detallada de contacto con los participantes, así como utilizar sistemas potentes de seguimiento y control de los participantes (que permitan identificar constantemente su localización y su estado), contactos frecuentes y mecanismos de recordatorio (3,9).

Cuando las pérdidas o la baja tasa de respuesta ya se han producido, la estrategia de control depende de si se conocen o no las características de la población de referencia (en caso de muestreo no probabilístico) o de las pérdidas (en caso de muestreo probabilístico). En el caso de ser desconocidas, el sesgo de selección no se podrá controlar en estudios de seguimiento, mientras que en estudios tipo encuesta la única posibilidad es la sustitución muestral, ya comentada. En el caso de conocer las características de las personas perdidas en muestras representativas, lo primero que hay que hacer es ver si hay diferencias entre los que participan y los que no, respecto de aquellas variables que influyen en el efecto estudiado. La existencia de diferencias pone de manifiesto la presencia de sesgo de selección. En el caso de conocer las características de la población de referencia, se puede intentar controlar el sesgo de selección (con independencia de la representatividad de la muestra original) demostrando que la muestra final y la población de referencia son similares en aquellas variables que influyen en el efecto estudiado o bien mediante técnicas de ajuste estadístico****.

El sesgo de selección debido a pérdidas de datos solo es controlable en la fase de planificación y ejecución, intentando evitar al máximo la pérdida de información. En lo que respecta a la existencia de datos o puntos extremos en variables cuantitativas, cuando son verdaderamente puntos extremos y no errores en la recogida de datos, la mejor solución es analizar los datos del estudio mediante técnicas estadísticas basadas en estadísticos robustos (como por ejemplo, la mediana o la media recortada), que evitan el rechazo de datos.

Sesgo de clasificación e información

Concepto y clasificación

El sesgo de clasificación consiste en la presencia de errores en la medición, los cuales pueden tener su origen en el investigador o en la persona investigada (Figura 5). El sesgo

de clasificación puede afectar a la precisión y puede llevar a asociaciones espurias o a la alteración de las asociaciones existentes. Los sesgos o errores que tienen su origen en el investigador (sesgo del observador) pueden estar relacionados con la inadecuada definición y/o aplicación de criterios diagnósticos o de medición, con la utilización de métodos y/o instrumentos de medida no pertinentes para lo que se quiere medir o inadecuados (aparatos mal calibrados o con fallos, escalas no validadas, etc.) y con errores en el procesamiento de los datos. Los sesgos que tiene su origen en la persona estudiada (sesgo del observado) pueden ser de varios tipos: imprecisiones, confusiones o temas tabú, errores al recordar un hecho pretérito (sesgo de memoria o de recuerdo), cansancio, variación de comportamiento al sentirse observadas (efecto Hawthorne) o como consecuencia de la búsqueda deliberada de un determinado resultado por parte del investigador (efecto del investigador).

Control del sesgo de clasificación

Los errores en la medición no pueden controlarse en la fase de análisis de datos. Las medidas orientadas a su control

Sesgo que tienen su origen en el observador (sesgo del observador)

- Derivados de la definición o aplicación de criterios diagnósticos
- Derivados de los métodos y/o instrumentos de medida
- Derivados de errores en el procesamiento de los datos

Sesgos que tienen su origen en el individuo estudiado (sesgo del observado)

- Sesgo de memoria o de recuerdo
- Confusiones, imprecisiones o temas tabú. Efecto Hawthorne
- Efecto investigador (parcialidad)

Figura 5. Sesgo de clasificación o información

**** Entre los distintos tipos de ajuste, podemos poner como ejemplo uno diseñado para encuestas, consistente en la asignación a los participantes de un factor de ajuste basado en la suposición de que las respuestas de los no respondedores hubiesen sido similares a los respondedores en cada estrato homogéneo de la muestra, construido a partir de la información disponible de dicha muestra (edad, sexo, etc.) (10). Para otros tipos de ajustes puede consultarse Miller y Hollist (9) y Cuddeback y cols (11).

hacen referencia a la fase de planificación y a la fase de recogida de datos. Dichas medidas son:

- Elaboración de un protocolo preciso para la recogida de datos (donde se detallan los procesos, personas e instrumentos relacionados con la obtención de datos) y entrenamiento adecuado de las personas encargadas de la recogida de datos.
- Utilización de definiciones operativas inequívocas de los eventos a estudiar (con categorías excluyentes y exhaustivas para las variables cualitativas o categóricas y precisando las escalas y/o unidades de medida para las cuantitativas).
- Establecimiento de codificaciones en la recogida de datos para mejorar su posterior procesamiento.
- Empleo de normas estandarizadas de medida para todos los sujetos así como instrumentos estructurados y calibrados/validados.
- Mantenimiento de controles a lo largo de todo el proceso (evaluaciones intermedias de la información recogida, reuniones con las personas encargadas de la recogida de datos, etc.).
- Aplicación de medidas que eviten la influencia de la persona que recoge los datos sobre la persona estudiada (efecto del investigador y/o Hawthorne).
- En caso de cuestionarios o escalas recogidos mediante entrevista, fraccionamiento de la recogida en más de una entrevista si fuese necesario.
- Realización de una prueba piloto para valorar la idoneidad y viabilidad de los métodos e instrumentos.

En relación con el sexto punto se han desarrollado diversas estrategias para su control en estudios experimentales, tales como el enmascaramiento y el diseño de cuatro grupos de Salomón. El enmascaramiento consiste en ocultar en que grupo de tratamiento se encuentra cada sujeto. Esta información se puede ocultar a la persona estudiada (simple ciego), a ésta y a la persona que recoge los datos (doble ciego) y a éstas y a la persona que realiza el análisis de los datos (triple ciego). El simple ciego previene el efecto Hawthorne. El doble y triple ciego previenen este tipo de sesgo más el del efecto del investigador. En estudios de interven-

ciones educativas, que suelen ser frecuentes en investigación enfermera, no es posible la prevención del efecto Hawthorne mediante el enmascaramiento a simple ciego, ya que la persona investigada es consciente del tipo de intervención que recibe, pero sí se puede prevenir el efecto del investigador cegando la recogida y análisis de datos y minimizar el efecto Hawthorne ocultando el efecto concreto que se persigue (si ello es posible desde el punto de vista ético). El diseño de cuatro grupos de Salomón permite analizar si la medición preintervención que se realiza en un ensayo clínico aleatorio tiene algún efecto sobre el resultado perseguido con la intervención (efecto Hawthorne). Para ello, se forman aleatoriamente cuatro grupos, aplicando la intervención que se quiere evaluar en dos de ellos y dejando los otros dos como grupo control (sin intervención o con otra intervención distinta), de manera que dentro de cada pareja se hace medición preintervención un grupo y en otro no. Si no hay efecto Hawthorne, los resultados (medición postintervención) en los dos grupos de intervención, por un lado, y en los dos grupos control, por otro, han de ser estadísticamente iguales.

En los estudios observacionales no hay procedimientos tan estructurados, pero sí se puede aplicar la misma lógica que en los experimentos. Así, para la prevención del efecto Hawthorne se pueden emplear las siguientes medidas: ocultar la hipótesis del estudio a las personas estudiadas (3) (si ello es posible desde el punto de vista ético) y evitar que exista algún vínculo entre las personas investigadas y las encargadas de recoger la información; mientras que para la prevención del efecto del investigador se puede ocultar la hipótesis del estudio a la personas encargadas de recoger los datos.

En la Tabla 1 aparecen resumidas las distintas medidas para el control del efecto Hawthorne y del efecto del investigador, en función del tipo de estudio.

Sesgo de confusión

Concepto y clasificación

El sesgo de confusión se produce cuando la asociación entre dos variables está distorsionada por el efecto de una tercera variable, la cual influye en la variable dependiente

Tabla 1. Medidas que previenen los errores derivados de la interactividad (efectos Hawthorn y del investigador)

	Experimentales	Observacionales
Efecto Hawthorne	Simple ciego No vínculo entre sujetos estudiados y recogedores de datos	Ocultar hipótesis a sujetos estudiados No vínculo
Efecto del investigador	Doble y triple ciego	Ocultar hipótesis a personas encargadas de recoger datos

y está asociada con la variable independiente (12). A ésta se le denomina variable de confusión o confundidora. La distorsión puede consistir en una sobre o infraestimación del efecto, en el cambio del sentido de la asociación, en crear una asociación que no existe (espuria) e incluso en ocultar una asociación real. En el caso de diseños epidemiológicos, donde siempre se comparan dos grupos, esta correlación con la variable independiente se traduce en una distribución desigual de la variable de confusión en los grupos que se comparan. La distorsión de la asociación estadística debida a una variable de confusión suele diferenciarse de la debida a un sesgo de selección o clasificación, ya que en estos últimos son relaciones que se crean artificialmente al seleccionar o medir (artefactos), mientras que en el caso del sesgo de confusión, la asociación entre la variable de confusión y el efecto analizado o variable dependiente es real.

Se comprenderá mejor todo esto con un ejemplo. Se quiere estudiar el efecto de la utilización del andador en ancianos con debilidad en miembros inferiores sobre el diagnóstico enfermero riesgo de caídas (variable dependiente). Para ello se comparan dos grupos de personas: aquéllas que han sufrido una caída en el último año y las que no han sufrido en dicho período de tiempo, de manera que se ve si existen diferencias en la utilización de andadores. Si el grupo de los que caen tiene mayor media de edad que los que no caen, las diferencias encontradas pueden no deberse al uso de andadores sino a la edad, ya que dicha edad está relacionada con las caídas. En este ejemplo, a la variable edad se le denomina variable de confusión.

Así, en el caso de un diseño epidemiológico, donde las variables independiente y dependiente son cualitativas o categóricas dicotómicas, para que una tercera variable sea potencialmente confundidora tienen que cumplirse tres condiciones (13):

- La variable de confusión está asociada a la variable independiente y a la variable dependiente.
- La asociación entre las variables independiente y dependiente es la misma en cada estrato o categoría de la variable de confusión, pero dichas asociaciones son distintas a la del análisis crudo (sin estratificar).
- La asociación cruda entre las variables independiente y dependiente es diferente de la asociación ajustada o controlando por la variable de confusión. A continuación se va a hacer algunas consideraciones al respecto de estas condiciones.

En lo que respecta a la asociación entre la variable de confusión y la variable dependiente, todas aquellas variables que la literatura muestra que tienen asociación con dicha variable dependiente han de ser consideradas potencialmente confundidoras y deben ser controladas por los procedimientos que más adelante describirán, aunque en los datos del estudio no exista dicha asociación. Esto es así por dos razones:

por un lado, no siempre se dispone de la potencia estadística suficiente para encontrar asociación estadística entre la variable de confusión y la variable independiente y por otro, al analizar la asociación entre dichas variables, el efecto de una tercera variable puede ocultar dicha asociación. Un ejemplo concreto del error que supondría no considerar variable de confusión a una variable descrita como tal en la literatura, pero que en los datos no se asocia a la variable dependiente, se puede ver en el artículo de Artaso y cols (14). En dicho artículo, se estudian variables relacionadas con la sobrecarga subjetiva en cuidadoras de mayores con demencia, una de las cuales es el afrontamiento basado en emociones. Existen diversas investigaciones donde se relaciona dicha variable con la sobrecarga subjetiva, pero en el artículo comentado, al existir un tamaño muestral insuficiente y por tanto, poca potencia estadística, los investigadores no encuentran asociación entre el mencionado tipo de afrontamiento y la sobrecarga subjetiva, por lo que cometen el error de no tener en cuenta dicho tipo de afrontamiento como variable potencial de confusión.

Ahora bien, si dicha variable de confusión no está asociada con la variable independiente, no tiene que ser tenida en cuenta como variable de confusión, ya que aunque sea realmente una variable de confusión, no actúa como tal en nuestros datos.

A diferencia de los otros dos sesgos, que pueden darse en cualquier tipo de estudio, el sesgo de confusión, por su naturaleza, está relacionado con estudios de correlación, es decir, estudios que miden asociación entre variables, estando libres de dicho sesgo los trabajos cuya finalidad es la determinación o descripción de uno o varios parámetros poblacionales, también llamados univariantes en la terminología de Denise Polit (15).

Control del sesgo de confusión

El sesgo de confusión es el único que puede controlarse (prevenirse) con garantías, además de en la fase de diseño (como el resto de sesgos), en la fase de análisis. Sin embargo, en cualquier caso, las variables potenciales de confusión han de ser identificadas mediante búsqueda bibliográfica y han de ser definidas en la fase de diseño y recogidas junto con el resto de variables del estudio.

Así, si se une la anterior idea a lo que se vió en el apartado de sesgo de selección, donde una de las recomendaciones era que siempre es conveniente describir la muestra estudiada para favorecer su validez externa, queda que en cualquier estudio donde se analicen hipótesis de correlación entre variables, hay como mínimo tres grupos de variables:

- Variables sociodemográficas: de descripción de la muestra.
- Variables de la/s hipótesis a estudiar: que pueden clasificarse en independiente/s y dependiente/s.

- Variables potenciales de confusión. En la Tabla 2 se desarrolla esta idea, ampliándose con el caso de estudios univariantes y mixtos (con hipótesis univariantes y de correlación).

El control de una variable de confusión puede realizarse mediante las siguientes técnicas: a) en la fase de diseño del estudio: intentando eliminar del estudio la variable o las variables de confusión mediante restricción (es decir, mediante criterios de inclusión o exclusión) o bien consiguiendo que éstas estén igualmente distribuidas en los grupos que se comparan (expuestos vs. no expuestos, casos vs. controles, intervención vs. control, etc.), utilizando para ello: apareamiento, muestreo aleatorio estratificado proporcional o asignación aleatoria de las personas del estudio a los grupos a comparar (en caso de experimentos); b) en la fase de análisis: mediante técnicas estadísticas tales como: análisis multivariante, análisis estratificado o estandarización o ajuste. Es posible que algunos de estos métodos necesiten ser explicados. A continuación se pasará a ello.

El apareamiento suele utilizarse en los estudios de casos y controles y consiste en la asignación a cada caso de uno o varios controles con igual valor en la/s variable/s potenciales de confusión, con lo que quedan igualadas dichas variables en los grupos a comparar. Por ejemplo, si la edad y el sexo son variables potenciales de confusión y la relación entre casos y controles es de 1 a 1, si se selecciona un caso que sea mujer y tenga 54 años, se elegirá un control que sea mujer y tenga la misma edad.

El muestreo aleatorio estratificado proporcional consiste en dividir a la población de referencia en subconjuntos (estratos) en base a uno o varios criterios (edad, sexo, etc.), de manera que en cada estrato se selecciona una muestra de forma aleatoria y cada estrato contribuye a la muestra total en función de su peso en la población de referencia. Cuando el criterio o criterios de estratificación son variables potenciales de confusión y los grupos a comparar se eligen de forma separada en cada estrato, se garantiza que dichas variables estén igualmente distribuidas entre los grupos a comparar.

El análisis multivariante permite analizar la relación entre las variable de interés (independiente y dependiente), eliminando los efectos de terceras variables (las potenciales de confusión), que son introducidas en el modelo estadístico como variables independientes. Entre los métodos de análisis multivariante con una sola variable dependiente más utilizados en Ciencias de la Salud están: a) regresión lineal múltiple: cuando la variable dependiente es cuantitativa y las independientes cuantitativas y/o dicotómicas; b) regresión logística: cuando la variable dependiente es cualitativa o categórica y las independientes cuantitativas y/o dicotómicas; y c) análisis de la covarianza (ANCOVA): cuando la variable dependiente es cuantitativa, la variable independiente principal es cualitativa y las potenciales variables de confusión (llamadas aquí covariables) son cuantitativas.

El análisis estratificado consiste en examinar la relación entre la variable independiente y la dependiente en cada uno de los estratos o categorías de una variable potencial de confusión (que, por tanto, ha de ser cualitativa o categórica). Cuando dicha variable es realmente confundidora, los resultados del análisis estadístico tienen que ser iguales en cada estrato de la variable de confusión pero distintos de los del análisis sin estratificar.

Por último, la estandarización o ajuste consiste en comparar el resultado de una determinada variable en dos grupos haciendo que dichos grupos tengan la misma distribución respecto de una tercera variable que sería la de confusión. El ejemplo clásico es de comparar la tasa bruta de mortalidad (variable dependiente) en dos poblaciones (variable independiente o grupos a comparar). Dado que el envejecimiento influye en la mortalidad (es decir, la edad es una variable de confusión en este caso), para poder comparar la mortalidad hay que ajustarla por edad. Para ello se multiplican, en cada población, las tasas de mortalidad específica de cada grupo de edad por los efectivos de dichos grupos de edad en una tercera población que se toma como referencia, obteniéndose con ello los sucesos esperados, es decir, las defunciones que tendrían las poblaciones que se comparan si tuviesen la misma estructura de edad que la población de referencia. Dicho

Tabla 2. Variables de un estudio en función de las hipótesis planteadas

	Estudio univariante (caracterización)	Estudio de correlación (asociación)	Estudio mixto (caracterización y asociación)
Variables que describen la muestra	X	X	X
Variables relacionadas con las hipótesis/objetivos del estudio	X	X	X
Variables potenciales de confusión a controlar en la fase de análisis		X	X

de otro modo, con este procedimiento se elimina el efecto de la edad sobre la mortalidad.

La restricción, como ya se ha visto, tiene el inconveniente de que limita la validez externa del estudio. Respecto de la asignación aleatoria a los grupos que se comparan (solo es posible en estudios experimentales), hay que tener en cuenta que únicamente garantiza que los grupos sean comparables cuando el tamaño de cada grupo es grande. Así, siempre se aconseja, a pesar de emplear este método, que se demuestre que no hay diferencias estadísticamente significativas entre los grupos a comparar respecto de las variables potenciales de confusión. El apareamiento, la estandarización o ajuste, el muestreo estratificado y análisis estratificado, tienen la limitación del número de variables por las que se ajusta, que no puede ser mayor de una (en el caso de la estandarización o ajuste) o dos (en el resto). El análisis estratificado, además, tiene el requisito de necesitar potencia suficiente en cada estrato, lo que aumenta el tamaño muestral necesario. Todo ello hace que la mejor opción de control sea el análisis multivariante.

Para terminar con la exposición del sesgo de confusión y su control, hay que ver una cuestión interesante cuando se emplean métodos multivariantes para el control del sesgo de confusión o se realiza análisis estratificado. Dicha cuestión consiste en decidir si la asociación en cada estrato (en análisis estratificado) o la asociación ajustada o controlada mediante análisis multivariante, son sustancialmente distintas de la asociación cruda o sin controlar (es decir, bivariante). Esta cuestión es necesaria tanto para dilucidar si la variable en cuestión es de confusión (ver condiciones para que una variable sea considerada de confusión descritas más arriba) como para interpretar los resultados finales del análisis estadístico multivariante, cuando se introducen todas las variables de confusión como variables independientes. Cuando la asociación aparece, desaparece o cambia de sentido al realizar el análisis estratificado o multivariante, la cuestión es fácil. Sin embargo, cuando lo que varía es la fuerza de la asociación, ¿cuál es el límite? Por convención, se recomienda un umbral de un 10% de variación, es decir, se considera un cambio significativo en el efecto cuando varía más de un 10% respecto del efecto medido mediante análisis bivariante. No obstante, se cree que la mejor opción es realizar un contraste de hipótesis de los efectos, es decir, ver si la diferencia entre efectos es estadísticamente significativa o no. Una forma relativamente sencilla para ello es el cálculo de los intervalos de confianza de la medidas de efecto o asociación. Si no hay solapamiento entre el intervalo de confianza del efecto calculado mediante análisis bivariante y el cal-

culado mediante análisis multivariante o estratificado, se puede concluir que los efectos son distintos.

Bibliografía

- [1] Delgado M, Llorca J. Estudios de observación. En: Piedrola G. Medicina preventiva y salud pública. Barcelona: Masson; 2000. p. 133-41.
- [2] Hernández-Ávila M, Garrido F, Salazar-Martínez E. Sesgos en estudios epidemiológicos. Salud Pub Mex 2000; 42:438-46.
- [3] De Irala Estévez J, Martínez González M, Toledo E, Basterra Gortari F. Errores en la estimación de medidas de asociación en estudios epidemiológicos. En: de Irala Estévez J, Martínez-González M, Seguí Gómez M. Epidemiología aplicada. 2ª ed. Barcelona: Ariel; 2008. p. 273-366.
- [1] Morabia A. Case-control studies in clinical research: mechanism and prevention of selection bias. Preventive medicine 1997; 26(5): 674-7.
- [5] Vermeulen F. Beware the dangers of selection bias. Blog Freek Vermeulen; 2009 [En línea] [fecha de acceso: 20 de mayo de 2011]. URL disponible en: <http://blogs.hbr.org/vermeulen/2009/03/beware-the-danger-of-selection.html>
- [6] Houde SC. Methodological issues in male caregiver research: an integrative review of the literature. J Adv Nurs 2002; 40(6): 626-40.
- [7] Dura JR, Kiecolt-Glaser JK. Sample bias in caregiving research. J Gerontol. 1990; 45(5): 200-4.
- [8] Pruchno RA, Brill JE, Shands Y, Gordon JR, Genderson MW, Rose M et ál. Convenience samples and caregiving research: how generalizable are the findings? Gerontologist 2008; 48(6): 820-7.
- [9] Miller R, Hollist C. Attrition bias. En: Salkind N, editor. Encyclopedia of Measurement and Statistics. Thousand Oaks: Sage; 2007. p. 57-60.
- [10] Vives A, Ferreccio C, Marshall G. Comparación de dos métodos para corregir el sesgo de no respuesta a una encuesta: sustitución muestral y ajuste según propensión a responder. Gacet Sanit 2009; 23(04): 266-71.
- [11] Cuddeback G, Wilson E, Orme J, Combs-Orme T. Detecting and statistically correcting sample selection bias. Journal of Social Service Research 2004; 30(3): 19-33.
- [12] Rothman K, Greenland S, Lash T. Modern epidemiology: Lippincott Williams & Wilkins; 2008.
- [13] Szklo M, Nieto F. Epidemiology Beyond the Basics. 2nd ed. Boston, MA: Jones and Bartlett Publishers; 2007.
- [14] Artaso B, Goñi A, Biurrún A. Cuidados informales en la demencia: predicción de sobrecarga en cuidadoras familiares. Rev Esp Geriatr Gerontol 2003; 38(4): 212-8.
- [15] Polit D, Beck C. Nursing research: Principles and methods. 7th ed. Philadelphia PA: Lippincott Williams & Wilkins; 2004.