

DOCUMENTOS

El proceso de medición con variables cualitativas y su aplicación en cirugía

Dr. CARLOS MANTEROLA D

Departamento de Cirugía, Capacitación, Investigación y Gestión para la Salud Basada en Evidencia (CIGES) Universidad de La Frontera

INTRODUCCIÓN

Lo primero que cabe plantearse antes de seguir desarrollando este artículo es una pregunta elemental, ¿por qué medir? Entonces, se ha de clarificar que el acto de medir es un componente esencial en la investigación científica. Los juicios subjetivos, como "a mí me parece", "el buen criterio indica", "a mí me ha ido bien", etcétera, juegan un papel anecdótico en el proceso de la investigación y debieran desaparecer de la práctica habitual de la medicina, que como sabemos está llena de mitos, algunos de los cuales provienen incluso de siglos anteriores. Debemos pues procurar cada día, actuar más conforme a modelos matemáticos, los que seguramente nos permitirán realizar una medicina más objetiva y, por ende, más predecible. Para ello, el concepto de medición es fundamental y lleva implícito a su vez los conceptos de validez y confiabilidad, los que le otorgan al proceso de medición un peso técnico y objetivo.

En la práctica clínica nos vemos enfrentados permanentemente a la necesidad de medir. Por ejemplo, necesitamos conocer elementos de la exploración física del paciente como la presión arterial o la presión intraocular, y para ello utilizaremos instrumentos de medición específicos, como el esfigmomanómetro y el tonómetro; por otro lado, podemos requerir datos acerca de parámetros bioquímicos del paciente, como glicemia, creatininemia, electrolitos en plasma, etcétera, entonces, requeriremos de instrumentos especiales, ya sean estos analizadores mecánicos o computacionales; pero aun con los mejores instrumentos y la tecnología más avanzada y sofisticada siempre

existirán errores de medición, algunos de los cuales, sobre todo si son sistemáticos, se denominarán sesgos.¹

Está claro, entonces, que siempre se necesita medir, y para ello debe existir el sujeto que mide, el que será medido, y el instrumento con que se realizará la medición; y por ende, se debe tener meridianamente claro que a estos tres niveles podremos cometer errores en la medición.

Pero, uno de los grandes problemas de la práctica clínica es que debemos trabajar con una serie de variables difíciles de objetivar, y por razones obvias muy difíciles de medir. A nadie le causa inconveniente trabajar con variables cuantitativas como el pulso, la presión arterial, los parámetros bioquímicos, hematológicos, etcétera. Pero, es indudable que a la hora de trabajar con variables cualitativas como sensaciones, síntomas, calidad de vida, indicación y/o valoración de terapias; nos vemos en general ante escenarios de gran complejidad, pues la medición de elementos subjetivos es muy difícil; lo que no significa que no debamos hacerlo, que los minimicemos, o incluso que lo reemplacemos por variables cuantitativas.

El objetivo de este manuscrito es dar una visión general de cómo trabajar con variables cualitativas en la práctica clínica y en especial en el ámbito de la cirugía.

Medición de variables cualitativas

La segunda pregunta que resulta en el desarrollo de esta idea es obvia ¿con qué mediremos variables cualitativas? Mediante instrumentos de medición ad-hoc válidos y confiables.^{1,2} Estos ins-

trumentos pueden ser estructurados, semi estructurados, o no estructurados; y pueden generarse con preguntas o ítems abiertos o cerrados dependiendo de la forma de elaboración y orden de éstas, por ejemplo. ¿Cuáles son sus molestias? vs. ¿Ha tenido Ud. vómitos? (la primera pregunta puede generar respuestas variadas en calidad y cantidad, en cambio la segunda puede tener como respuesta un sí o un no). Detrás de esto figura la intención de preguntar a cada participante de la medición y/o estudio las mismas preguntas con una secuencia idéntica, y obtener el mayor nivel de información (a través de preguntas abiertas) o evitar respuestas ambiguas o no atingentes a la investigación (mediante la aplicación de preguntas cerradas).³

Estos instrumentos pueden llamarse encuestas, cuestionarios o escalas, y no son otra cosa que un conjunto de ítems que intentan revelar niveles de variables teóricas difíciles de establecer en forma directa.⁴ Estos instrumentos permiten valorar diferentes tipos de datos, por lo que sus escalas de medición pueden ser de 4 tipos: *nominales* (las observaciones se clasifican en categorías; por ejemplo: sexo, etnia, color de ojos), *ordinales* (en las que se asignan rangos ordenados a las categorías; por ejemplo: leve, moderado y grave), de *intervalos* (en las que la distancia entre dos mediciones ha sido definida; por ejemplo: los escores de las pruebas de inteligencia), y de *razón* (en las que tanto la distancia como la razón entre dos mediciones han sido definidas; por ejemplo: longitud, incidencia de enfermedad, número de hijos, etc.).

Una vez desarrollado el instrumento, el siguiente problema es como administrarlo, y esto puede realizarse a través de entrevistas personales (cara a cara o auto administrado), por teléfono o por correo, y de una manera adicional que es a través del computador, ya sea en forma directa o por correo electrónico.

La administración cara a cara del cuestionario tiene las ventajas de que el investigador puede responder dudas, observar la conducta y otras características del encuestado, permite preguntar ítems más complejos, el entrevistador evita las preguntas sin respuesta o las respuestas inapropiadas y puede controlar el orden de las preguntas,³ pero presenta la desventaja de poder generar un sesgo de respuesta por parte del entrevistador si éste no es independiente a la investigación.

Variables de la modalidad entrevista personal son la aplicación telefónica y en directo por medio del computador. Estas son más económicas pero carecen de la impronta propia de una conversación. Por otra parte, la administración por correo ya sea

éste tradicional o electrónico es absolutamente impersonal, y tiene los inconvenientes del sesgo de respuesta y una pérdida no despreciable de sujetos que no contestarán el cuestionario por diversas razones.

La auto administración del cuestionario tiene la ventaja de reducir los costos del estudio; por otra parte, las instrucciones escritas reducen los sesgos de diferencias de administración del instrumento y de interacción con el entrevistador,³ pero tiene las desventajas de perder información valiosa como observar la conducta y otras características del encuestado, explicar aquellos ítems que el encuestado no comprende en forma adecuada, etc.

Construcción de un instrumento de medición

La tercera pregunta que aparece entonces no es menos obvia ¿cómo construir un instrumento de medición? Hay una gran variedad de detalles metodológicos que se han de tener presentes al momento de construir un instrumento de medición, de modo tal de minimizar errores de medición que posteriormente pueden inducir a resultados equívocos.^{1,5-7} Se ha de decir si se aplicará un cuestionario autoadministrado, una entrevista o un instrumento mixto que contenga ambas modalidades; luego se ha de confeccionar una lista con todas las variables que parecen de interés a la investigación (es decir los ítems); a continuación escribir un borrador del instrumento; someterlo a revisión por pares; realizar un estudio de pretest, aplicando el instrumento a 3 ó 5 personas similares a aquéllas en que se va a utilizar de modo tal de apreciar el nivel de comprensión de los ítems por parte de los sujetos en que se aplicará; luego, si se constata que es muy largo, es el momento de acortar y volver a revisar; para finalmente aplicarlo a la población en estudio.

Los ítems. En términos generales, podríamos definir un ítem como toda aquella pregunta, variable o parámetro que forma parte del instrumento que se está generando.

1. **¿De adónde provienen los ítems?** Existen múltiples fuentes de obtención de ítems para la construcción de un instrumento, cada una de ellas tiene ventajas y desventajas, por lo que el usuario debe saber de dónde provienen y estar atento a los posibles sesgos que a partir de esta fuente se puede introducir en la investigación. Los ítems pueden provenir de escalas ya existentes (algunas escalas existentes pueden no cubrir todos los aspectos de un área, sin embargo, sus ítems suelen seguir utilizándose, entonces, se rescata lo útil y se desecha lo inútil); de observaciones clínicas (siguen existiendo territorios vírgenes a escalas de medición,

por otro lado, cada día aparecen tópicos que no son cubiertos por las herramientas actuales); de la opinión de expertos (esto es útil siempre y cuando sea en forma de un panel de expertos, y que existan diferentes enfoques del problema, de lo contrario, se corre el riesgo que aparezcan los sesgos propios de esta situación); de hallazgos de investigaciones previas (las investigaciones aportan otra fuente de ítems potenciales); e incluso de observaciones hechas por los mismos pacientes acerca de sus propias experiencias (a través del aporte de elementos subjetivos que en oportunidades no son tomados en cuenta por los clínicos). Otras fuentes de obtención de información para la construcción de ítems son: los informantes clave (agentes especiales de acceso a una comunidad de la que se pretende obtener información, normalmente son líderes dentro de su comunidad o personas respetadas por sus pares); las entrevistas por grupos focales (método que consiste en una "discusión", en la que un pequeño grupo de personas guiados por un "facilitador", hablan en forma libre y espontánea acerca de materias importantes para una investigación); y, las denominadas entrevistas cualitativas (método que consiste en acceder a la perspectiva del individuo entrevistado sin influir en ella, especialmente en lo que dice relación con eventos relacionados con el conocimiento, los valores, los sentidos, etc.).¹

2. **Desarrollo de los ítems.** Desarrollar un pool de ítems es el primer paso necesario para construir una escala. Sin embargo, es difícil que todos sean entendidos por los encuestados. Algunos ítems pueden ser mal interpretados, por lo tanto, es necesario eliminar estos "malos ítems" si no cumplen ciertos criterios. Sin embargo, algo que parece sencillo no lo es, pues en la formulación de los ítems existen varios pasos que se deben considerar. Uno de ellos es el nivel de comprensión de los ítems, pues en general, los profesionales de la salud suelen utilizar una jerga denominada "medicalese", entonces, los individuos que contestan un cuestionario no tienen por qué conocer esta terminología y pueden mal interpretar los términos de las preguntas.

3. **Ítems ambiguos u ofensivos.** Cuando un sujeto no responde a una pregunta, puede significar que no está claro cómo se debe contestar. Por otro lado, es posible que algunos ítems puedan ser una intromisión en aspectos íntimos de un individuo, por lo que estos aspectos deben ser omitidos o tratados con mucha cautela. Para evitar ser ambiguo u ofensivo, se debe realizar un estudio piloto, que consiste en aplicar las preguntas a personas similares a las que el cuestionario va dirigido antes

de ponerlo en marcha, para así poder captar errores de este tipo.

Validación de un instrumento de medición

La validez se puede definir como el grado con que los resultados de una medición se corresponden con el estado de veracidad del fenómeno medido. Son sinónimos exactitud y precisión. Existen diversos tipos de validez que se pueden establecer.¹

1. **Validez de contenido.** Una medición que incluya una muestra muy representativa de lo que se desea estudiar permitirá una mayor capacidad de deducción. Si existen aspectos importantes del resultado que han sido obviadas por la escala de medición, entonces será interesante hacer algunas deducciones que permitan probar que existen estos errores. Así, el poder de la validez de contenido de una medición, es la amplitud de las deducciones que se pueden validar acerca de los sujetos bajo una serie de condiciones, y en diferentes situaciones.

2. **Validez de criterio.** Es la correlación de una escala con otras mediciones de la característica del fenómeno en estudio; idealmente, un estándar de referencia que haya sido ya utilizado y aceptado en este campo. La validez de criterio se puede dividir a su vez en validez concurrente y validez predictiva.

3. **Validez concurrente.** Se encuentra inserta en el estudio de validez de criterio, y representa el grado de correlación de la escala generada, con otras mediciones propias del estándar de referencia. Existen estándares de referencia que están compuestos por diversas variables que se resumen finalmente en un score o puntuación, la que puede ser utilizada para establecer validez de criterio. Si esto acontece, se podrá verificar a continuación validez concurrente entre la escala generada con las variables que componen el estándar de referencia.

4. **Validez predictiva.** Se encuentra inserta en el estudio de validez de criterio, y determinando validez predictiva se intenta establecer la capacidad de la nueva escala para predecir resultados o el desarrollo de algún acontecimiento en estudio.

5. **Validez de constructo.** Un constructo se puede definir como una mini-teoría, que permite explicar la relación entre diversas actitudes y conductas. Las razones que inspiran la creación de un constructo son que constituye un nuevo instrumento de medición, que no existen escalas que lo midan; o que las herramientas disponibles omiten algunos aspectos que se encuentran insertos en el constructo.⁵ Por lo tanto, validez de constructo, se

utiliza para valorar variables no medibles (por ejemplo, la "ansiedad"), donde se construye un instrumento de medición a través de conceptos que el paciente relata pero que no son observables, como sensaciones o motivaciones. Un constructo se puede validar utilizando grupos extremos y determinando la validez convergente.

6. Validez de grupos extremos. Como recién se mencionó, es parte del estudio de validez de constructo, y consiste en aplicar la escala a dos grupos, uno que tenga la característica en estudio (puntuación muy alta de la escala) y otro que no (puntuación muy baja, hecho que puede ser inverso y que dependerá del tipo de escala que se utilicen); de esta forma se podrá conocer el comportamiento de la nueva escala en sujetos con características extremas del fenómeno en estudio.

7. Validez convergente. También corresponde al estudio de validez de constructo y se puede explicar como que tan relacionados se encuentran la nueva escala con otras variables y otras mediciones del mismo constructo relacionado.

Confiabilidad de un instrumento de medición

La confiabilidad o reproducibilidad, constituye un aspecto fundamental en medición. Significa que las mediciones realizadas en ocasiones diferentes, o por observadores diferentes, o por pruebas paralelas, determinarán los mismos resultados. Existen varias formas de medir confiabilidad.¹

1. Consistencia interna. La confiabilidad de consistencia interna, como su nombre lo indica, se relaciona con la homogeneidad de los ítems que comprenden una escala. Las escalas basadas en modelos de medición clásicos están pensadas para medir un fenómeno singular. Si los ítems de una escala tienen una relación fuerte con su variable latente, ellos tendrán entonces una relación fuerte entre ellos. Una escala se considera internamente consistente si sus ítems se encuentran altamente interrelacionados. Las correlaciones fuertes entre los diferentes ítems se encuentran altamente interrelacionados. Las correlaciones fuertes entre los diferentes ítems sugieren que todos ellos se encuentran todos midiendo la misma cosa.

2. Confiabilidad test-retest. Una escala debe ser estable permanentemente, y se asume que nada debe cambiar en el intertanto de las mediciones. Si un score es diferente sin haber mediado algún cambio, entonces, es imposible saber que número es una reflexión precisa del estado de una persona en estudio. El valor aceptado para el coeficiente de confiabilidad test-retest depende del intervalo entre las dos administraciones, y de la estabi-

lidad hipotética del efecto que se está midiendo. Si el test ha sido muy largo, el retest no debe ser efectuado antes de 2 semanas del primero. De otra forma los sujetos estarán recordando sus respuestas previas, y tenderán a repetirlas. La confiabilidad test-retest de 0,60 se considera marginal; de 0,70 aceptable; y, sobre 0,80 muy alta.

3. Confiabilidad interobservador o interclasificador. Consiste en que dos o más observadores, y en forma independiente evaluando al mismo sujeto obtengan el mayor grado de acuerdo posible en la medición que realizan. Esto habla del grado de reproducibilidad del instrumento al ser aplicado por diferentes observadores. Tal como en la confiabilidad test re-test, la falta de acuerdo usualmente significa que los resultados no serán verdaderos. Aquí existen 2 potenciales fuentes de error: los clasificadores u observadores son diferentes; y, los sujetos en estudio pueden haber cambiado en el intertanto, o alterado sus respuestas entre ambos evaluadores. La confiabilidad no debe ser inferior de 0,60, idealmente debe ser al menos de 0,70; considerándose como casi perfectos cuando es superior a 0,80.

4. Confiabilidad intraobservador. Consiste en que un observador, evaluando al mismo sujeto dos o más veces, logre puntuaciones similares, es decir sea capaz de reproducir su medición. Esta tiene el inconveniente de que el entrevistador conoce al sujeto entrevistado al momento de aplicar por segunda vez el cuestionario.

Limitaciones metodológicas en la construcción de un instrumento de medición

Las limitaciones del diseño se relacionan fundamentalmente con problemas del estilo y redacción de la fraseología, y la influencia de potenciales sesgos.

1. Estilo y redacción de la fraseología. Es de importancia la consideración de este aspecto, ya que cualquier palabra puede influir en la validez y confiabilidad de la respuesta. Entre los problemas relacionados con la fraseología se encuentra la claridad, la simplicidad y neutralidad de la pregunta; además del uso del "y/o", y la necesidad de establecer tiempos precisos a las preguntas. Este problema se puede minimizar con una cuidadosa observancia de los pasos sugeridos en la confección del cuestionario (confección de una lista de las variables, estudio de instrumentos similares previamente utilizados, escritura de un borrador, revisión de cada pregunta, prueba del borrador, abreviación del borrador, nueva revisión, y preconfirmación).¹

2. **Sesgos potenciales.** El principal tipo de sesgo en el diseño de un instrumento de medición es el denominado sesgo de medición. Este aumenta cuando el investigador que aplica el cuestionario conoce los resultados de los otros exámenes practicados al individuo en estudio. Por ejemplo, si el investigador que aplica el cuestionario tiene conocimiento acerca de los resultados del estándar de referencia del mismo individuo. Este problema se puede atenuar con un diseño adecuado del instrumento de medición, con un sistema apropiado de administración del instrumento, y una revisión periódica de los hallazgos del trabajo en curso. El investigador que aplicará el cuestionario sólo sabrá si el sujeto en estudio tiene síntomas y aplicará por igual el cuestionario a todos los individuos que ingresen en esta etapa del estudio de acuerdo a los criterios de inclusión y de exclusión. El instrumento será guardado y no se entregará al investigador principal hasta que el examen constituyente del estándar de referencia se haya aplicado y finalizado para un sujeto dado. En resumen, el control general del sesgo de medición está dado por la independencia de los investigadores que aplican el cuestionario, el ciego total a los resultados de la aplicación del cuestionario por parte del encargado de aplicar el estándar de referencia, y la dureza de éste en la determinación de una entidad dada.¹

Análisis estadístico y tratamiento de los datos

En este punto se definirá cómo se determina la consistencia interna, la confiabilidad y la validez del instrumento de medición.

1. Determinación de la consistencia interna.

La consistencia interna se puede medir mediante las pruebas de Kuder-Richardson o KR-20 y el alfa de Cronbach o "coeficiente alfa". El KR-20 es apropiado para escalas de ítems cuyas respuestas son dicotómicas, en cambio, el alfa de Cronbach, puede ser utilizado cuando los ítems tienen más de dos alternativas de respuesta. Por otro lado, si el alfa de Cronbach es usado en ítems dicotómicos, los resultados serán idénticos a los obtenidos al aplicar el KR-20. El coeficiente alfa, tiene un rango entre 0,00 y 1,00, y se considera que una escala es adecuada en este sentido cuando tiene un coeficiente mayor a 0,70. De todos modos, éste se puede incrementar con el número de ítems en la escala, así, cuestionarios de 20 ítems suelen tener valores de alrededor de 0,90.^{1,6}

2. **Determinación de confiabilidad.** La confiabilidad interobservador o interclasificador se puede medir a través del coeficiente kappa, que es la proporción de respuestas en las que dos observa-

ciones coinciden. Este método permite estudiar la proporción de respuestas en las dos celdas de acuerdo (sí/sí, no/no), en relación con la proporción de las respuestas que por azar se esperaría encontrar en estas celdas.^{1,6}

3. **Determinación de validez del instrumento.** Para ello se utilizan herramientas como regresión lineal múltiple y regresión logística. Posteriormente, una vez establecido el nuevo instrumento de medición, se determina sensibilidad, especificidad y valores predictivos positivo y negativo. Estos cálculos se realizan según los métodos clásicos basados en la construcción de tablas contingencia o de 2 x 2.^{7,8}

Ejemplos

Antes de analizar algunos ejemplos relacionados con el proceso de medición con variables cualitativas y con la construcción de instrumentos que permitan evaluarlas, parece necesario describir formas de análisis crítico de la literatura biomédica para artículos de esta naturaleza. A modo de ejemplo se expone una pauta de análisis para estudios de prevalencia, que pueden ser utilizados en la práctica quirúrgica habitual para conocer o determinar la prevalencia de fenómenos clínicos difíciles de dimensionar como síntomas, sensaciones, etcétera. Esta pauta fue construida en el curso de una Tesis de Grado, tendiente a determinar prevalencia de enfermedad por reflujo gastroesofágico (ERGE) en población general,⁹ y ha sido aplicada a los artículos que a continuación se discuten (Tablas 1 y 2).

1. *Bohmer C y cols. The prevalence of gastroesophageal reflux disease based on non-specific symptoms in institutionalised, intellectually disabled individuals.*¹⁰ Se trata de un estudio realizado en Holanda con el objeto de determinar la frecuencia de ERGE en individuos con discapacidad mental (coeficiente intelectual inferior a 50). Se investiga la prevalencia de ERGE, esofagitis por reflujo y factores predisponentes. Se aplica una escala para pesquisa de síntomas inespecíficos de reflujo y factores relacionados; por otro lado, se realiza una pHmetría y endoscopia. Se encontró un 65% de esofagitis y en el 55% de los casos la pHmetría fue patológica, cifras que no se relacionan con los resultados de la escala aplicada. Al comparar los resultados del instrumento aplicado en estos individuos con las pruebas diagnósticas, se pudo constatar cierta asociación de ERGE con parálisis cerebral y uso de drogas anticonvulsivantes. Se concluye en que existe una elevada prevalencia de ERGE en este grupo de sujetos. Al aplicar la pauta se logra el siguiente análisis:

Tabla 1
PAUTA DE ANÁLISIS CRÍTICO DE LA LITERATURA, PARA ARTÍCULOS DE PREVALENCIA⁹

I. Aspectos generales del estudio	¿Tiene el estudio un objetivo claro y preciso? ¿Se desarrolló sobre la base de un diseño de investigación de corte transversal? ¿Describe variables y su patrón de distribución? ¿Se realizó la medición deseada en una sola oportunidad? ¿Intenta el estudio buscar alguna asociación?
II. Aspectos referentes a la población en estudio	¿Especifica la población blanco? ¿Cómo se calculó el tamaño de la muestra? ¿La muestra obtenida es representativa de la población en estudio?
III. Aspectos referentes al instrumento de medición	¿Se precisa en el estudio un instrumento de medición? ¿El instrumento utilizado fue previamente validado?, ¿cómo se evaluó? ¿El instrumento utilizado es confiable? ¿qué tipo de confiabilidad se evaluó?
IV. Aspectos referentes a los resultados	¿Se obtiene una medida de prevalencia? ¿Pueden los resultados inferirse a la población en estudio?
V. Aspectos referentes a potenciales sesgos	¿Cómo se evitó el sesgo de selección de la muestra? ¿Cómo se evitó el sesgo de memoria?

Tabla 2
MATRIZ RESUMEN DE LA LECTURA CRÍTICA DE LOS ARTÍCULOS SELECCIONADOS

Item a evaluar	Ref. 10	Ref. 11	Ref. 12
¿Tiene el estudio un objetivo claro y preciso?	*	*	*
¿Se desarrolló un diseño de investigación de corte transversal?	*	*	*
¿Describe variables y su patrón de distribución?	*	*	*
¿Se realizó la medición deseada en una sola oportunidad?	*	*	*
¿Intenta el estudio buscar alguna asociación?	*	*	*
¿Especifica la población blanco?	*	*	X
¿Se efectuó cálculo del tamaño de la muestra?	X	X	X
¿La muestra obtenida es representativa de la población en estudio?	X	X	X
¿Se precisa en el estudio un instrumento de medición?	*	*	*
¿El instrumento utilizado fue previamente validado?	X	X	X
¿El instrumento utilizado es confiable?	X	X	X
¿Se obtiene una medida de prevalencia?	*	*	*
¿Pueden los resultados inferirse a la población en estudio?	X	X	X
¿Se evitó el sesgo de selección de la muestra?	X	X	X
¿Se evitó el sesgo de memoria?	X	X	X
Total de criterios cumplidos	8	8	7

*Cumple con el criterio.

XNo cumple con el criterio.

I. Aspectos generales. El objetivo de este trabajo es estudiar la prevalencia de ERGE en individuos discapacitados mentales hospitalizados y buscar factores predisponentes. A pesar de que no se menciona en forma expresa, el diseño corresponde al de un estudio de corte transversal. Se realiza una descripción de las variables clínicas y biodemográficas a estudiar. Se aplicó la escala y las pruebas diagnósticas en un momento determi-

nado. Se intenta buscar asociación entre ERGE en discapacitados mentales, consumo de drogas de acción sobre el sistema nervioso central y diversas alteraciones mentales.

II. Aspectos referentes a la población en estudio. Se trata de individuos con discapacidad mental hospitalizados (coeficiente intelectual inferior a 50). No se menciona si se realizó determinación del tamaño de la muestra necesario, sólo se

hace referencia a que los sujetos que participaron en el estudio fueron escogidos en forma aleatoria. Con la información que se entregó no es posible establecer el grado de representatividad de la muestra.

III. Aspectos referentes al instrumento de medición. Se utilizó como instrumento de medición una escala para objetivar síntomas de ERGE, tipo y características de las alteraciones mentales del individuo y uso de drogas para su tratamiento. No existen evidencias de validación del instrumento. No existen evidencias de estudios de confiabilidad del instrumento utilizado.

IV. Aspectos referentes a los resultados. Se obtiene la prevalencia de esofagitis y reflujo gastroesofágico patológico. Es difícil inferir resultados a la población blanco, puesto que no se puede establecer representatividad de la muestra.

V. Aspectos referentes a potenciales sesgos. No se menciona cómo se evitaron posibles sesgos de selección. No se menciona como se evitaron posibles sesgos de memoria.

VI. Conclusiones de la revisión crítica. Se trata de un estudio de corte transversal destinado a estudiar la prevalencia de ERGE en discapacitados mentales hospitalizados. No se puede establecer representatividad de la muestra respecto de la población blanco. Se utilizó como instrumento de medición una escala, de la que se desconoce información respecto a estudios de validez y confiabilidad. Se obtiene la prevalencia de esofagitis y reflujo patológico, información aparentemente no inferible debido al problema de representatividad de la muestra. No se mencionan instrumentos metodológicos para minimizar sesgos. En resumen, este trabajo cumple sólo con 8 de los 15 criterios evaluados a través de la pauta confeccionada.

2. *Ruth M. y cols. The prevalence of symptoms suggestive of esophageal disorders.*¹¹ Con el objeto de determinar la prevalencia y gravedad de desórdenes esofágicos en la población general se distribuyó por correo un cuestionario a 407 sujetos con edades entre 20 y 79 años seleccionados en forma aleatoria. De los 407 sujetos seleccionados, sólo 337 contestaron el cuestionario (85%). Se encontraron síntomas sugerentes de reflujo gastroesofágico en el 25% de los participantes, tos en el 27%, dolor torácico en el 13%, disfagia en el 10% y vómitos en el 8%. En la mayoría de los sujetos los síntomas fueron considerados como leves a moderados, y sólo el 1 a 4% de ellos consideró que los síntomas eran graves. No se encontró correlación entre síntomas, edad y sexo, como tampoco con el reporte de consumo de tabaco, alcohol y antiinflamatorios no esteroidales. Mediante un análisis de

regresión logística, se demostró que la tos crónica y/o la disnea se relacionaron en forma significativa con la existencia de síntomas sugerentes de reflujo gastroesofágico. Al aplicar la pauta se logra el siguiente análisis.

I. Aspectos generales. El objetivo de este trabajo es determinar la prevalencia y gravedad de síntomas sugerentes de desórdenes esofágicos en la población general, con énfasis particular en reflujo gastroesofágico. A pesar que no se menciona en forma directa, el diseño corresponde aparentemente al de un estudio de corte transversal. Se realiza una descripción de variables demográficas, clínicas y referentes al consumo de tabaco, alcohol y antiinflamatorios no esteroidales. A pesar que no se menciona en forma expresa, aparentemente se aplicó un cuestionario a través de correo en un momento dado a todos los sujetos en estudio. Se intenta buscar asociación entre síntomas y desórdenes esofágicos, y por otro lado, entre síntomas esofágicos y respiratorios.

II. Aspectos referentes a la población en estudio. Se menciona que la población blanco es la población general. No existe cálculo del tamaño de la muestra necesario. En primer término, se aplicó el cuestionario a individuos entre 20 y 79 años, por lo que no incluye otras edades que también conforman la población general. En segundo lugar, no se aprecia si se efectuó una estratificación de la población que permita representar todos los sectores de la población general. Finalmente, como no se calcula tamaño de la muestra, ni siquiera se puede saber si la muestra estudiada es representativa de los individuos entre 20 y 79 años de la ciudad de Gotenburgo, Suecia, donde se aplicó el cuestionario.

III. Aspectos referentes al instrumento de medición. Se utilizó como instrumento de medición un cuestionario, que se envió a domicilio por correo, junto con una nota explicativa acerca del propósito del estudio y de su carácter voluntario. En ninguna parte del trabajo se menciona la validación del instrumento, sólo que éste ha sido utilizado en su laboratorio en conjunto con otras pruebas diagnósticas por más de cinco años. No se precisa estudios de confiabilidad aplicados, sólo podría considerarse cierto grado de confiabilidad interobservador, si es que este instrumento fue aplicado en este laboratorio por diferentes observadores por más de cinco años con resultados similares, hecho que no se especifica.

IV. Aspectos referentes a los resultados. Se obtiene la prevalencia de los siguientes síntomas: 27% de tos, 26% de pirosis y/o regurgitación, 21% sólo de pirosis, 20% sólo de regurgitación, 15% de

pirosis y regurgitación, 12% de dolor torácico, 10% de disfagia y 8% de vómitos. En primer lugar, la muestra estudiada y a pesar de haber sido obtenida en forma aleatoria entre los ciudadanos de Gotenburgo, no fue sometida a cálculo de tamaño de muestra, y con sólo observar el número de ítems en estudio parece insuficiente. En segundo lugar, la muestra de individuos estudiados sólo podría representar a la población entre 20 y 79 años de edad de esta ciudad y no a la población general.

V. Aspectos referentes a potenciales sesgos. No se menciona cómo se evitaron posibles sesgos de selección, ni cómo se evitaron posibles sesgos de memoria. De hecho, al ser enviado el cuestionario a domicilio, el sesgo de respuesta puede ser considerable.

VI. Conclusiones de la revisión crítica. Se trata de un estudio de corte transversal tendiente a conocer la prevalencia y gravedad de síntomas sugerentes de desórdenes esofágicos en la población general. Existe una población blanco, pero la muestra estudiada no es representativa. Del instrumento aplicado no se precisa validez ni confiabilidad, y los resultados obtenidos no son inferibles. No se utilizaron herramientas metodológicas tendientes a minimizar los sesgos. En resumen, este trabajo sólo cumple con 8 de los 15 criterios evaluados a través de la pauta confeccionada.

3. *Isolauri J, Laippala P. Prevalence of symptoms suggestive of gastro-esophageal reflux disease in an adult population.*¹² Se trata de un estudio realizado con el objeto de determinar la prevalencia de síntomas de ERGE en una población adulta no seleccionada. Para tal efecto se envía un cuestionario por correo a 2.500 personas de edad igual o superior a 20 años. Los síntomas estudiados son pirosis, regurgitación, disfagia, dolor abdominal alto y torácico, y además se pregunta por uso de drogas para el control de los síntomas y consulta en servicios médicos a raíz de los síntomas. Sólo un 68% de los sujetos respondió el cuestionario; de éstos, se obtuvo una prevalencia de pirosis de 9% el día de la encuesta, 15%, 21% y 27% en la semana, el mes y el año precedente a la encuesta respectivamente. Algo similar se constató con el síntoma regurgitación (5%, 15%, 29%, 45%). Se encontró, además, asociación de síntoma de ERGE con edad, embarazo, sobrepeso y hábito de fumar. Por otro lado, el 16% de los entrevistados reconoció utilizar antiácidos para el control de sus síntomas, y un 5,5% haber consultado médico por su sintomatología durante el año. Al aplicar la pauta se logra el siguiente análisis:

I. Aspectos generales. El objetivo de este trabajo es estudiar la prevalencia de síntomas de

ERGE en adultos. El diseño corresponde al de un estudio de corte transversal. Se realiza una descripción de las variables clínicas y biodemográficas a estudiar. Se aplicó el cuestionario en un momento determinado vía postal. Se intenta buscar asociación entre síntomas de ERGE, edad, embarazo, sobrepeso y hábito de fumar.

II. Aspectos referentes a la población en estudio. El trabajo menciona que el estudio se realiza sobre una población no seleccionada de adultos. No se menciona determinación del tamaño de la muestra necesario. Con la información que se entrega no es posible establecer el grado de representatividad de la muestra.

III. Aspectos referentes al instrumento de medición. Se utilizó como instrumento de medición un cuestionario autoadministrado, enviado por correo. No existen evidencias de validación del instrumento. No existen evidencias de estudios de confiabilidad del instrumento utilizado.

IV. Aspectos referentes a los resultados. Se obtiene la prevalencia de los diversos síntomas de ERGE estudiados en forma diaria, semanal, mensual y anual. Es difícil inferir resultados, puesto que no hay una población blanco debidamente identificada.

V. Aspectos referentes a potenciales sesgos. No se menciona cómo se evitaron posibles sesgos de selección. No se menciona cómo se evitaron posibles sesgos de memoria.

VI. Conclusiones de la revisión crítica. Se trata de un estudio de corte transversal, destinado a estudiar la prevalencia de síntomas de ERGE en adultos. No se puede establecer representatividad de la muestra respecto de la población blanco. Se utiliza como instrumento de medición un cuestionario despachado por correo del que se desconoce información respecto a estudios de validez y confiabilidad. Se obtiene la prevalencia diaria, semanal, mensual y anual de los síntomas estudiados, información aparentemente no inferible debido al problema de población blanco. No se mencionan instrumentos metodológicos para minimizar sesgos. En resumen, este trabajo cumple con 7 de los 15 criterios evaluados a través de la pauta confeccionada.

BIBLIOGRAFÍA

1. Streiner DL, Norman GR: Health measurement scales. A practical guide to their development and use. Oxford University Press, New York: United States, 1995; 1-3.
2. Frisbie DA, Brandenburg DC: Equivalence of questionnaire items with varying responses formats. J Ed Meas 1979; 16: 43-8.

3. Cummings SR, Strull W, Nevit MC, Hulley SB. Planning the measurements: Questionnaires, en designing clinical research. In: Hulley SB, Cummings SR. *Designing Clinical Research*. First edition. Baltimore, United States: Williams and Wilkins 1988; 42-52.
4. Last JM: *A dictionary of epidemiology*. Third edition. Oxford University Press. New York, 1995.
5. Steiner DL: Research methods in psychiatry. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry* 1993; 38: 140-8.
6. De Vellis RF: *Reliability, in scale development. Theory and applications*. SAGE Publications, First edition. Newbury Park 1991; 24-42.
7. Fletcher RH, Fletcher SW, Wagner AH: *Clinical epidemiology, the essentials*. 2nd ed. Baltimore, United States: Williams and Wilkins 1982; 42-75.
8. Browner WS, Newman TB, Cummings SR: Designing a new study: III. Diagnostic tests. In: Hulley SB, Cummings SR (eds). *Designing Clinical Research*. First edition. Baltimore, United States 1988; 87-97.
9. Manterola C: *Prevalencia de síntomas de enfermedad por reflujo gastroesofágico en población general urbana de la ciudad de Temuco*. Tesis para optar al Grado de Magíster en Epidemiología Clínica Universidad de La Frontera, Temuco, 1997.
10. Bohmer CJ, Klinkenberg-Knol EC, Niezen-de Boer RC, Meuwissen SG: The prevalence of gastro-oesophageal reflux disease based on non-specific symptoms in institutionalised, intellectually disabled individuals. *Eur J Gastroenterol Hepatol* 1997; 9: 187-90.
11. Ruth M, Mansson Y, Sandberg N: The prevalence of symptoms suggestive of esophageal disorders. *Scand J Gastroenterol* 1991; 26: 73-81.
12. Isolauri J, Laippala P: Prevalence of symptoms suggestive of gastro-oesophageal reflux disease in an adult population. *Ann Intern Med* 1995; 27: 67-70.