

Research protocol VII. Validity and reliability of the measurements

El protocolo de investigación VII. Validez y confiabilidad de las mediciones

Miguel Ángel Villasís-Keever,¹ Horacio Márquez-González,² Jessie Nallely Zurita-Cruz,³ Guadalupe Miranda-Navales,¹ Alberto Escamilla-Núñez¹

Abstract

The concept of validity in research refers to what is true or what is close to the truth. It is considered that the results of an investigation will be valid when the study is free of errors. The errors or biases appear in the development of research, are due to methodological problems and, in general, can be grouped into three: selection bias, measurement bias and confusion bias. In this article, measurement biases will be addressed; this type of error has three axes: 1) the research subject, 2) the instrument for the measurement of the variable (s), and 3) those who make the evaluation of the measurement (s). To improve the obtaining of data and to prevent errors, some strategies can be followed: in every study protocol, it is necessary to include the operational definition of the variables; the subjects that will carry out the measurements or surveys must be trained. If measuring instruments are used, their proper functioning must be verified; when questionnaires are included, they must be validated in the language in which they will be applied, they have had a process of adaptation to the language of the participants in the study, and self-applicable ones are preferred. It is necessary to quantify the variability of the measurements from the statistical point of view to increase the validity of a study.

Keywords: Validity; Reliability; Bias

Este artículo debe citarse como: Villasís-Keever MÁ, Márquez-González H, Zurita-Cruz JN, Miranda-Navales G, Escamilla-Núñez A. El protocolo de investigación VII. Validez y confiabilidad de las mediciones. Rev Alerg Mex. 2018;65(4):414-421

ORCID

Miguel Ángel Villasís-Keever, 0000-0002-8566-0811; Horacio Márquez-González, 0000-0001-9041-5813; Jessie Nallely Zurita-Cruz, 0000-0003-1389-7964; María Guadalupe Miranda-Navales, 0000-0003-3262-2608; Alberto Escamilla-Núñez, 0000-0002-5357-9891

¹Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Unidad de Investigación en Síntesis y Análisis de la Evidencia, Ciudad de México, México

²Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Cardiología, Ciudad de México, México

³Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Pediatría, Ciudad de México, México

Correspondencia: Miguel Ángel Villasís-Keever.
miguel.villasis@gmail.com

Recibido: 2018-10-11

Aceptado: 2018-10-15

DOI: 10.29262/ram.v65i4.560



Resumen

El concepto de validez en investigación se refiere a lo que es verdadero o se acerca a la verdad. Se considera que los resultados de una investigación son válidos cuando el estudio está libre de errores. Los errores o sesgos que se presentan en el desarrollo de una investigación se deben a problemas metodológicos y pueden agruparse en tres categorías: sesgos de selección, sesgo en la medición y sesgo de confusión. En este artículo se abordan los sesgos de medición, tipo de error que tienen tres ejes: el sujeto de investigación, el instrumento para la medición de las variables y el evaluador de las mediciones. Para mejorar la obtención de los datos y prevenir los errores se pueden seguir algunas estrategias: incluir la definición operacional de las variables, la capacitación de los sujetos que van a realizar las mediciones o encuestas, la verificación del adecuado funcionamiento de los instrumentos de medición, la validación de los cuestionarios en el idioma en el que se van a aplicar, su adecuación al idioma de los participantes en el estudio y que de preferencia sean autoaplicables. Es necesario cuantificar la variabilidad de las mediciones desde el punto de vista estadístico, para aumentar la validez de un estudio.

Palabras clave: Validez; Confiabilidad; Sesgo

Antecedentes

Desde hace años se ha reconocido que para obtener los mejores resultados en la práctica clínica es fundamental tomar decisiones incorporando los resultados de los estudios de investigación, tanto al realizar un diagnóstico, prescribir medicamentos o establecer el pronóstico de alguna enfermedad. Para lograrlo, la metodología utilizada debe ser la correcta, ya que las fallas en el diseño o en la ejecución del estudio puede llevar a conclusiones erróneas o sin validez.

El concepto de validez en investigación se refiere a lo que es verdadero o lo que se acerca a la verdad. En general se considera que los resultados de una investigación serán válidos cuando el estudio está libre de errores. Para establecer si un determinado estudio es válido, se debe analizar la presencia de sesgos (errores sistemáticos) como mínimo en los siguientes puntos: el diseño de investigación, los criterios de selección y la forma de llevar a cabo las mediciones, es decir, la manera de registrar y evaluar las variables de estudio. De esta forma, se considera que un estudio tiene *validez interna* cuando está libre de sesgos.

En lo que se refiere al diseño de investigación y, como ha sido mencionado en otros artículos de la serie “El protocolo de investigación” de *Revista Alergia de México*, el paso primordial es la selección del diseño más apropiado cuando se planea la realización de cualquier proyecto, tomando en cuenta el objetivo principal de la investigación. Por ejemplo, cuando se va a evaluar la eficacia de un nuevo tratamiento se debe seleccionar un ensayo clínico controlado y aleatorizado, en tanto que el seguimiento de una cohorte es el diseño básico en un estudio donde se desea establecer el pronóstico de una enfermedad. Por otra parte, tanto el estudio de casos y controles como las cohortes son los mejores si se desea analizar las causas o factores de riesgo para el desarrollo de las enfermedades. Cuando por alguna razón los investigadores no han elegido el mejor diseño para dar respuesta al objetivo de su estudio, los resultados de dicha investigación no tendrán la solidez o el grado de validez necesaria para ser incorporados en la práctica, ya que existe cierto grado de incertidumbre.

Respecto a la población de estudio, los investigadores tienen que identificar las características idóneas para conseguir sus objetivos. Así, desde un

principio deberán señalar las características de los participantes, y cuando aplique, la forma de determinar cómo se establece el diagnóstico de la enfermedad, así como su estadio o gravedad. Todo deberá estar consignado en los criterios de selección, tanto de los inclusiones, exclusión y eliminación.

Finalmente, es indispensable describir cómo se van a ejecutar las mediciones que se analizarán, para así determinar la validez de un estudio de investigación. En un artículo previo en esta serie se abordó el tema de variables y se señaló la necesidad de expresar las definiciones por utilizar (conceptual y operacional), así como identificar su tipo (desde el punto de vista metodológico), la escala y las unidades de medición. De esta forma, el objetivo del presente artículo es explicar los aspectos que deben tomarse en cuenta para que las mediciones en un estudio tengan los atributos necesarios para considerarlas válidas.

Definiciones

SI bien, ya señalamos el concepto de validez, antes de continuar es necesario establecer los términos y definiciones que se utilizarán a lo largo del texto, ya que existen conceptos que pueden confundir al lector dado que en la literatura abundan diferentes significados para un mismo concepto, o bien en la traducción del idioma inglés al español se utilizan diferentes términos.

Errores sistemáticos o sesgos

Los sesgos son errores que se presentan siempre en el desarrollo de una investigación, de ahí que se denominen errores sistemáticos. Este tipo de errores se deben particularmente a problemas metodológicos y, generalmente, pueden agruparse en tres: sesgos de selección, sesgo en la medición y sesgo de confusión. El primer tipo ocurre cuando los criterios de selección de la población por estudiar no son congruentes con los objetivos del estudio. Los sesgos de confusión se generan específicamente cuando se lleva a cabo el o los análisis estadísticos, tanto por la selección del tipo de como en su interpretación. Los sesgos de medición (motivo principal de este artículo) suceden al momento de la recolección de las variables de resultado; este tipo de errores tiene tres ejes:

- El sujeto de investigación.
- El instrumento para la medición de la o las variables.
- Quienes hacen la evaluación de las mediciones.

Errores aleatorios

El error aleatorio puede definirse como la fluctuación o variación natural y periódica de cualquier variable que se mide (temperatura, presión arterial, frecuencia cardíaca, saturación arterial de oxígeno, entre otras). En un estudio de investigación clínica se debe tener en cuenta que el error aleatorio ocurre tanto por la variabilidad humana innata y por la casualidad, de ahí su nombre. En general se considera que estos errores no se pueden eliminar, pero pueden ser disminuidos mediante diferentes estrategias, como aumentando el tamaño de muestra y estandarizando las mediciones en los participantes de la investigación.

Confiabilidad

Los resultados de un estudio pueden considerarse confiables cuando tienen un alto grado de validez, es decir, cuando no hay sesgos. Sin embargo, este término se utiliza más cuando se están desarrollando instrumentos o escalas clínicas (por ejemplo, para evaluar depresión, calidad de vida, gravedad de las enfermedades). Así, una vez que se establece que una escala es reproducible y consistente, entonces puede concluirse que es confiable.

Consistencia

Esta característica o atributo se refiere a la capacidad de un instrumento para obtener los mismos resultados al medir una variable; las mediciones deben realizarse bajo las mismas condiciones de los participantes y de su entorno. El término consistencia también se usa cuando distintos estudios de investigación derivan en las mismas conclusiones.

Reproducibilidad

Este término se utiliza más frecuentemente para describir una metodología o técnica para la obtención de resultados, más que la obtención de resultados similares. Es común que también este concepto se describa como repetitividad.

Precisión y exactitud

Ambos términos están relacionados y se refieren a la capacidad que tienen los instrumentos o escalas al momento de la medición de las variables; una medición tendrá mayor validez cuanto más precisa y exacta sea. Para entender mejor ambos conceptos, se presenta la figura 1. Si con un instrumento se obtienen resultados con poca variabilidad entonces se

puede asumir que es más preciso (por ejemplo, una báscula tiene mayor precisión al medir gramos que aquella que solo mide kilos completos). Mientras que la exactitud (en inglés, *accuracy*) de un instrumento se establece por su capacidad de obtener resultados “verdaderos” o “libres de sesgos”

Fuentes de sesgo al hacer mediciones

Las principales fuentes de sesgos de medición pueden ser las siguientes:

- Por el instrumento con que se realiza la medición.
- Por el observador.
- Por las personas que realizan la evaluación
- Por el participante o sujeto en estudio.

Sesgos por el instrumento de medición

Habitualmente para medir las variables es necesario utilizar instrumentos. Estos instrumentos de evaluación pueden ser dispositivos médicos (como esfigmomanómetros, básculas, termómetros), sistemas automatizados (como equipos para medir niveles de hemoglobina, electrolitos séricos, química sanguínea), así como encuestas, cuestionarios o escalas que pueden responder tanto pacientes como médicos. Es necesario tener en cuenta que cualquiera de los anteriores puede tener limitaciones técnicas o metodológicas para medir exactamente lo que se desea. Uno de los objetivos de la utilización de cualquiera de estos instrumentos en la investigación, al igual que la aplicación de pruebas diagnósticas utilizadas

en la clínica, es tratar de recuperar los datos de cada participante para clasificarlo como enfermo, sano, con alguna comorbilidad, o bien, por su gravedad o estadio de la enfermedad. Específicamente cuando no se usa el método ideal para hacer el diagnóstico de la condición de los participantes en estudio, existirá un sesgo denominado de mala clasificación, el cual está incluido dentro del grupo de los sesgos de selección.

Otra finalidad de los instrumentos es la medición de las variables de interés, en particular las variables de desenlace. Por ejemplo, si se desea determinar la eficacia de un medicamento para rinitis alérgica, entonces lo más apropiado sería utilizar un cuestionario donde se determine cómo se presenta la sintomatología, antes y después de brindar el tratamiento. En otro ejemplo, los niveles de hemoglobina obtenidos mediante equipos de laboratorio automatizados servirán para establecer la eficacia de los medicamentos para pacientes con anemia.

Por lo anterior, cuando no se utilizan los instrumentos más apropiados se presentarán sesgos para medir variable en estudio; por ejemplo, si se desea evaluar cómo ha mejorado la anemia de un paciente, la medición de niveles de hemoglobina en sangre será más exacto que cuando solo se usan datos clínicos. También habrá errores en la medición al no calibrar o realizar mantenimiento regular de los instrumentos o equipos utilizados en la investigación; en este último punto, será importante reconocer que,

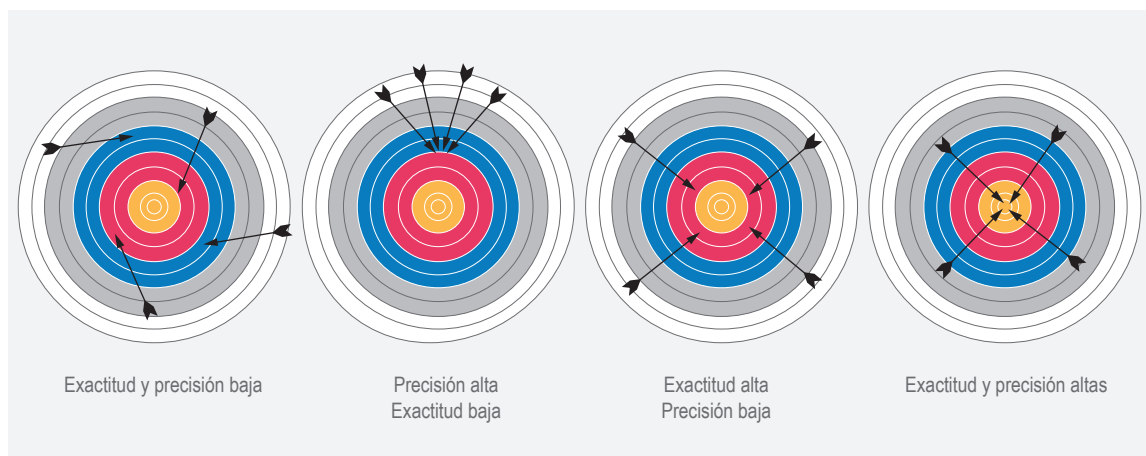


Figura 1. Ejemplos de precisión y exactitud de las mediciones.

al descuidar ambas situaciones, los datos obtenidos tendrán más variabilidad (es decir, menor precisión) por lo que el estudio tendrá menor calidad y, por ende, validez.

Sesgo por el evaluador

Es posible que existan errores en la medición de las variables de estudio por las personas encargadas de recolectar los datos; estas personas también son conocidos como investigadores, observadores o evaluadores. En este rubro, los errores consisten, en general, en tres aspectos principales que suceden durante la ejecución de una investigación:

- En la forma de llevar a cabo el levantamiento o recolección de los datos.
- En la interpretación de dichos datos.
- En el número de los observadores.

Respecto al primer rubro, si en una investigación se requiere que la medición de los datos sea mediante la participación de observadores, los sesgos existirán si desde un inicio la manera en que se realizarán las mediciones no están bien definidas, es decir, cuando no se ha considerado un proceso de estandarización de dichas mediciones. Una persona con *conocimiento del tema* puede suponer que la medición será de acuerdo con lo que realiza regularmente, cuando para los fines del proyecto es necesario se realice de forma diferente. Por lo anterior, es importante tener en cuenta que no siempre es suficiente que la persona encargada de hacer las evaluaciones tenga un perfil determinado (por ejemplo, ser médico, enfermera o nutricionista), sino que es necesario definir la forma como se harán las mediciones, de preferencia usando criterios universalmente aceptados. A fin de garantizar que el estudio sea reproducible debe haber claridad en la manera en la que se van a efectuar cada una de las mediciones, sin importar que las mediciones sean “rutinarias” o “sencillas”, tales como la toma de la presión arterial, el peso o la estatura, o bien, la identificación de signos y síntomas para evaluar la modificación de la gravedad de una enfermedad, como pueden ser las exacerbaciones (o crisis) asmáticas.

En cuanto al aspecto de la interpretación de las variables, es posible que existan sesgos a pesar de los procesos de estandarización, ya que las mediciones (principalmente las que se obtienen basadas

en datos clínicos) están sujetas a ser obtenidas en forma diferente a la realidad. Existen investigaciones donde los observadores pueden, inconscientemente, hacer juicios equivocados sobre las condiciones de los participantes de la investigación. Lo anterior puede ocurrir, por ejemplo, en un estudio donde se evalúa la efectividad de un fármaco en comparación con un placebo; si el observador sabe que un paciente está recibiendo el placebo es posible que interprete que hay poca mejoría, o viceversa. Esta misma situación sucede en investigaciones donde se buscan los factores asociados con las enfermedades; si el investigador conoce que un sujeto determinado tiene la enfermedad en estudio, es posible que interrogue con mayor intensidad sobre los posibles factores asociados con el desarrollo de dicha enfermedad, en comparación con quienes no tienen la enfermedad. Para evitar este sesgo, siempre será necesario que los observadores realicen sus mediciones objetivamente; la mejor estrategia es el cegamiento de los observadores. Siguiendo con los escenarios señalados, al momento de hacer las mediciones los observadores no deberán conocer si están recibiendo el fármaco o placebo, o bien, las condiciones de los participantes, es decir, con o sin enfermedad.

Sesgo por el sujeto de investigación

Durante la ejecución de un proyecto de investigación clínica se pueden presentar sesgos en el momento de hacer las mediciones, al no tomar en cuenta que en los participantes de los estudios puede haber variaciones en sus características clínicas, más cuando estas constituirán una parte fundamental de los datos por recabar. Las condiciones clínicas incluyen tanto datos clínicos (tensión arterial, frecuencia respiratoria, temperatura), como datos obtenidos de muestras de laboratorio (glucosa, electrolitos, saturación de oxígeno). Las modificaciones en estos datos pueden ser normales o fisiológicas, como puede consecuencia de la hora del día, por ejercicio, o bien, antes o después de consumir alimentos.

Por lo anterior, en la planeación de cualquier investigación es primordial que, para obtener confiablemente los datos, debe haber algún proceso de estandarización. Por ejemplo, en el caso que se desee evaluar la presión arterial, será importante que su medición sea en posición sedente, después de un periodo de reposo y en el brazo no dominante; mien-

tras que, si se trata de medir glucemia o lípidos, lo mejor será después de un periodo de ayuno.

Sesgo por el número de evaluadores

En los protocolos de investigación puede ser necesaria la participación de más de un evaluador. Además de considerar que los evaluadores pueden tener variaciones individuales en el momento de la medición de las variables (lo cual se conoce como *variabilidad intraobservador*), será necesario tener en cuenta la posibilidad que se presenten variaciones de las mediciones entre los diferentes evaluadores; de ahí que, para garantizar que un estudio es confiable la *variabilidad interobservador* debe ser mínima. Para lograrlo, es necesaria la estandarización de las mediciones, donde es indispensable llevar a cabo un proceso de capacitación, a fin de que las evaluaciones sean semejantes entre los diferentes evaluadores.

Cómo mejorar la confiabilidad de las mediciones

Uno de los puntos básicos para determinar la validez de un estudio son la forma de realizar las mediciones; así, todo investigador tiene que asegurarse que las mediciones sean apropiadas para los objetivos del estudio, que estén en consonancia con lo conocimiento universal, que tengan una adecuada distribución de respuestas en la población y, que sean objetivas.

Para mejorar la obtención de los datos y para prevenir los errores, se pueden seguir las estrategias que a continuación se señalan, con las cuales, además se logrará reducir el error aleatorio y aumentar la precisión de las mediciones:

1. Todo protocolo de estudio requiere la definición operacional de las variables. Esta definición incluirá las instrucciones específicas para realizar las mediciones. La descripción puede ser extensa y detallada por lo que se recomienda agregarla en un anexo, o bien elaborar un manual de procedimientos.
2. El siguiente paso es capacitar a los sujetos que harán las mediciones, las encuestas, o registrarán las variables. No es suficiente con entregarles el manual, debe verificarse su desempeño y evaluar sus registros. En general, cualquier medición mejora su precisión si se realiza varias veces.

3. De emplearse instrumentos de medición (mecánicos o digitales), deberá verificarse su funcionamiento. Siempre que sea posible deberán emplearse instrumentos automatizados que no requieran la participación de otro sujeto, para así registrar únicamente la respuesta sin intervención del observador.
4. Al incluir cuestionarios, de preferencia deberán seleccionarse los que estén validados, en su caso, aquellos en idioma diferente (en este caso, al español) utilizar los que han tenido un proceso de adecuación y adaptación al idioma de los participantes en el estudio. Son preferibles los instrumentos autoaplicables a aquellos que requieren una entrevista

En el cuadro 1 se resumen las principales fuentes de sesgo y estrategias para aumentar a validez de las mediciones.

Cómo medir la variabilidad de las mediciones

Se ha señalado que cuando hay mínima variabilidad, tanto intraobservador como interobservador, los estudios tendrán mayor validez. Por lo tanto, antes de iniciar cualquier proyecto de investigación es necesario determinar, desde el punto de vista estadístico, la magnitud de dicha variabilidad. Se recomienda hacer un estudio piloto con pocos participantes para poder llevar a cabo la evaluación de los observadores.

Los análisis estadísticos apropiados para analizar la variabilidad dependerán de la escala de medición de la variable de interés. Cuando la variable a medir es cualitativa, ya sea categórica (presencia o ausencia de algún síntoma) u ordinal (gravedad de las enfermedades) se puede utilizar el índice de Kappa. Este índice ayuda a evaluar la concordancia (es decir, lo contrario de la variabilidad) intra e interobservador, para las variables categóricas se deberá seleccionar Kappa simple, y para las ordinales Kappa ponderado. Los valores que se pueden obtener van de 0 a 1; así, entre más cerca del 1.0 se puede asumir muy baja variabilidad.

En el caso de que se la escala sea cuantitativa (peso, estatura, número de leucocitos, etcétera) se puede emplear el coeficiente de variación intraclass, que se expresa e interpreta aproximadamente igual que el índice Kappa. O también, se puede ele-

Cuadro 1. Síntesis de las fuentes de sesgo y forma de control para mejorar la validez de mediciones		
Fuente de sesgo	Posible falla	Forma de control
Observador/investigador	Variabilidad intraobservador Variabilidad interobservador	Estandarización Capacitación
Observador/investigador	Conocimiento previo de la condición clínica del sujeto de investigación Conocimiento de los objetivos del estudio	Cegamiento
Instrumento de medición: equipos	Problema de calibración Falta de mantenimiento	Estandarización
Instrumento de medición: escalas, cuestionarios	No es el apropiado/válido Obtención incompleta de datos	Selección del mejor instrumento* Estandarización
Participante del estudio	Variabilidad "normal" o fisiológica (ciclo circadiano) Variabilidad por incidentes (ej. ejercicio, ingestión de alimentos o medicamentos)	Estandarización Aumentar tamaño de muestra

*Instrumento válido, versión adaptada al idioma de los participantes, de preferencia autoaplicable.

gir análisis de varianza (ANOVA) de medidas repetidas; para este último, después de llevar a cabo los cálculos correspondientes, si indica que el valor de $p > 0.05$, entonces se determina que hay poca variabilidad.

De acuerdo con cada proyecto de investigación, de detectarse por los análisis estadísticos que existe (amplia) variabilidad intra o interobservador, los investigadores deberán realizar un proceso de capacitación para verificar que las mediciones se realizan conforme a lo planeado. Después de la capacitación, nuevamente deberán hacerse los análisis estadísticos correspondientes. El estudio no podrá iniciarse hasta que los resultados de los análisis indiquen que no hay variabilidad o esta es mínima.

Aumentar el tamaño de muestra es una estrategia más que ayuda a aumentar la precisión y, por consiguiente, disminuye la variabilidad

Validez en escalas o cuestionarios

Además de todo lo señalado, para los propósitos de este artículo parece conveniente señalar algo sobre otros términos que se mencionan en diversos textos y artículos, como validez aparente, validez de contenido, validez de constructo, validez de criterio, validez concurrente y validez divergente, entre otros.

Todos esos términos habitualmente son utilizados, particularmente, en el proceso de construcción y desarrollo de escalas o cuestionarios para medir variables o fenómenos que son subjetivos, o bien, cuando se requiere incluir diferentes datos clínicos y de laboratorio para llegar a un diagnóstico o establecer la gravedad de alguna enfermedad, tales como las escalas para evaluar calidad de vida, estilos de vida satisfacción, depresión, ansiedad. Este tema en particular se abordará en un artículo posterior de esta misma serie.

Referencias

1. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58(8):635-641. DOI: 10.1136/jech.2003.008466
2. Hernández-Ávila M, Garrido F, Salazar-Martínez E. Sesgos en estudios epidemiológicos. *Salud Publica Mex*. 2000;42:438-446.
3. Hulley SB, Cummings SR, Broewner WS, Grady DG, Newman TB. *Diseño de investigaciones clínicas*. EE. UU.: Lippincot, Williams & Wilkins; 2014.

4. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg.* 2010;126(2):619-625. DOI: 10.1097/PRS.0b013e3181de24bc
5. Sessler DI, Imrey PB. Clinical research methodology 1: study designs and methodologic source of error. *Anesth Analg.* 2015;121(4):1034-1042. DOI: 10.1213/ANE.0000000000000815
6. Sitthi-amorn C, Poshyachinda V. Bias. *Lancet.* 1993;342(8866):286-288. DOI: 10.1016/0140-6736(93)91823-5
7. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. *Nephron Clin Pract.* 2010;115(2):c94-c99. DOI: 10.1159/000312871
8. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med.* 2004;140(3):189-202.
9. Vetter TR, Mascha EJ. Bias, confounding, and interaction: lions and tigers, and bears, oh my! *Anesth Analg.* 2017;125(3):1042-1048. DOI: 10.1213/ANE.0000000000002332
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;66(10):1093-1104. DOI: 10.1016/j.jclinepi.2013.05.014