

Carta metodológica: Validez de un estudio: validez interna y externa

Methodological letter: Validity of a study: internal and external validity

Si realizamos un estudio para saber si una nueva técnica quirúrgica es mejor que otra, tenemos que saber si las diferencias observadas entre ambas técnicas son debidas al azar o debidas a cuestiones de diseño y desarrollo del estudio para poder descartar explicaciones alternativas de los efectos observados. O al revés, si no observamos diferencias, estar seguros de que en realidad no existen porque realmente no hay una relación causal. Porque ante los resultados de un estudio nos tenemos que preguntar si son correctos, sin errores ni sesgos, si pueden ser atribuidos al azar y si son aplicables a otros contextos¹.

Se definen dos tipos de validez de un estudio: validez interna y validez externa.

La **validez interna** valora el grado en que los resultados obtenidos en un estudio concreto son correctos para las personas estudiadas en este estudio. Nos determina si el resultado observado se puede atribuir a la nueva técnica quirúrgica evaluada. Hay 2 errores principales que podría amenazar la validez interna de un estudio. Son los sesgos (también llamados errores sistemáticos) y los errores aleatorios.

La **validez externa (o generalización)** valora el grado en que las conclusiones extraídas de un estudio se pueden extrapolar o generalizar más allá de la muestra de población estudiada. Depende del tamaño y características de la muestra y del contexto en el que se quiere aplicar.

La falta de validez interna influye negativamente en la calidad de la evidencia que puede derivarse de un estudio. Un estudio con validez interna puede o no tener validez externa, pero un estudio sin validez interna no puede tener validez externa.

Errores que limitan la validez interna

Errores sistemáticos

El error sistemático es un error que consistentemente tiene lugar en cada medición. No son debidos al azar ni dependen del tamaño muestral. Aparece cuando se introduce un error en el diseño del trabajo, ya sea en la selección de los individuos, en la información recogida o en su

análisis e interpretación. Los sesgos sistemáticos pueden dar lugar a una sobreestimación o subestimación del verdadero efecto de una intervención o tratamiento en estudio².

Una de las clasificaciones más extendidas propone agrupar los sesgos en tres categorías generales: selección, información y confusión³.

Sesgo de selección: La muestra no representa la población diana que se quiere estudiar. Se presentan, principalmente, en los estudios de casos y controles y en los estudios retrospectivos de cohortes. Sin embargo, las pérdidas durante el seguimiento de los participantes en estos estudios también pueden dar lugar a un sesgo de selección en las cohortes prospectivas. Por otro lado, los ensayos clínicos controlados y aleatorizados (ECCA) también son susceptibles de sesgo de selección dependiendo del momento en que se informa e invita al paciente, si antes o después de la aleatorización y de las posibilidades de dar el consentimiento informado^{4,5}. No tener en cuenta los pacientes que no dan el consentimiento y que forman parte de la tasa de no respuesta es otra causa de sesgo de selección. También lo es no tener en cuenta características que puedan estar asociadas al pronóstico. La aleatorización puede minimizar este problema.

Sesgo de información: errores no debidos al azar de mediación de la información. Ocurre cuando la información necesaria se recoge de forma sistemáticamente diferente entre los grupos de estudio. En esta categoría encontramos los sesgos del entrevistador, del entrevistado, de memoria, etc. Todos los diseños de estudio son susceptibles del sesgo de información.

Podemos introducir sesgos, por ejemplo, al prestar más atención a los pacientes operados con una nueva técnica y que esto derive en tener menos complicaciones o menos graves porque se detecten antes. O, contrariamente, que al estar más observados se detecten y registren más complicaciones. Si la manera de obtener o interpretar la información difiere en función del grupo de pacientes, especialmente si quien obtiene o quien da la información conoce las hipótesis y objetivos del estudio, hay posibilidad de sesgo de información. Un paciente que conoce el tratamiento que recibe puede estar más atento a efectos secundarios y reportarlos en mayor medida.

En parte estos sesgos se pueden evitar con buenos protocolos que definan de forma exhaustiva qué información hay que registrar y como, incluyendo técnicas de enmascaramiento, y aplicándolos exactamente igual para todos los participantes.

Sin embargo, los estudios en cirugía tienen dificultades añadidas propias de la disciplina⁶, como son la selección de un grupo control (a menudo se recurre a grupos históricos difícilmente comparables), la dificultad de dar placebo, la dificultad de realizar ensayos doble ciego, el efecto del cuidador o cuidados perioperatorios, la experiencia del cirujano, etc. entre otros⁷. Aunque existen estrategias para minimizar estos sesgos, como realizar operaciones simuladas, evaluación ciega de resultados o el uso de un segundo equipo quirúrgico, no siempre son posibles, por lo que debe buscarse un equilibrio entre el riesgo de sesgo y la factibilidad de realizar el estudio en las mejores condiciones posibles².

Confusión

La confusión se produce cuando la asociación observada se debe, al menos en parte, a otras diferencias entre los grupos estudiados distintas a la exposición o intervención bajo estudio, que afectan de forma independiente al riesgo de presentar el desenlace de interés. Se puede limitar durante el diseño del estudio utilizando técnicas que eviten desequilibrios entre los grupos estudiados en aquellas variables que pueden ser confusoras. El emparejamiento, la restricción y especialmente la aleatorización son herramientas para prevenir la confusión en el diseño de un estudio. Y el análisis estratificado y el análisis multivariante, permiten controlarlo en el análisis.

Errores aleatorios y fiabilidad

El error aleatorio se debe a que se trabaja con muestras de individuos y no con toda la población. Procede de la variabilidad inherente al muestreo y depende del tamaño de la muestra: cuando éste aumenta, el error disminuye. La estadística permite cuantificar el error aleatorio y hablamos de potencia estadística.

Validez externa

La validez externa se refiere a la medida en que los resultados de los ensayos o estudios observacionales proporcionan una base correcta para la generalización a otras poblaciones o a otras circunstancias clínicas. La validez interna es un requisito previo para la validez externa. Los resultados de un estudio válido internamente pueden no ser aplicables en otras poblaciones bien porque la muestra no es representativa debido a sesgos o restricciones en criterios de inclusión, bien porque las condiciones metodológicas de estudio no pueden replicarse fácilmente en la práctica habitual o por otros factores como la relación médico-paciente, la información que este tiene sobre el tratamiento que está recibiendo o por sus valores y preferencias. Además, las personas (pacientes e investigadores) reaccionan de modo

diferente cuando observan o son observados en el contexto de un estudio que cuando se encuentran en condiciones normales^{8,9}. Otros problemas que limitan la generalización de resultados son los debidos a factores que afectan a la interpretación e incorporación de la evidencia en la toma de decisiones, tal como han descrito Garcia-Alamino y Lopez Cano en su carta metodológica (**pendiente publicación Cir Española**).

Conclusiones

La evaluación de la validez interna implica descartar fuentes de sesgo y errores aleatorios y tiene que ver con la calidad metodológica. La validez externa se ocupa de valorar la aplicabilidad de los resultados a la población del mundo real e incluye aspectos metodológicos, pero también del entorno o contexto. Para mejorar la calidad de los estudios se han elaborado guía de comprobación (*reporting guidelines*) entre las que destacan: declaración CONSORT (para ensayos clínicos), PRISMA (para revisiones sistemáticas), QUADAS y START (para pruebas diagnósticas), y STROBE (para investigación observacional). Existen nuevas propuestas, algunas en fase de validación, adaptadas a ámbitos específicos de investigación, incluyendo la cirugía.

Bibliografia

1. Kwaan MR, Melton GB. Evidence-Based Medicine in Surgical Education. *Clin Colon Rectal Surg.* 2012;**25**(3):151-5. doi: 10.1055/s-0032-1322552.
2. Gurusamy KS, Gluud C, Nikolova D, Davidson BR. Assessment of risk of bias in randomized clinical trials in surgery. *Br J Surg.* 2009;**96**(4):342-9. doi: 10.1002/bjs.6558.
3. Rothman KJ, Greenland S LT. *Modern Epidemiology.* 3th ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
4. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet.* 2002;**359**(9306):614-8. doi: 10.1016/S0140-6736(02)07750-4.
5. Hotter B, Ulm L, Hoffmann S, Katan M, Montaner J, Bustamante A, Meisel A. Selection bias in clinical stroke trials depending on ability to consent. *BMC Neurol.* 2017;**17**(1):206. doi: 10.1186/s12883-017-0989-9.
6. Demange MK, Fregni F. Limits to clinical trials in surgical areas. *Clinics (Sao Paulo).* 2011;**66**(1):159-61. doi: 10.1590/s1807-59322011000100027.
7. Mcculloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions Obstacles to randomised trials in surgery. *BMJ.* 2002; **324**(7351): 1448–1451.
8. Nguyen VN, Miller C, Sunderland J, McGuinness W. Understanding the Hawthorne effect in wound research-A scoping review. *Int Wound J.* 2018;**15**(6):1010-1024. doi: 10.1111/iwj.12968.
9. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ* 2015;**351** doi: <https://doi.org/10.1136/bmj.h4672>