

ERRORES EN LA LITERATURA BIOMÉDICA: UN CATÁLOGO PARA EL LECTOR CRÍTICO

JORGE DAGNINO S.¹

- Los errores en la literatura biomédica son frecuentes y escaparon a la atención de autores, revisores y editores.
- Hay errores frecuentes en el análisis, la presentación o la interpretación que pueden ser corregidos en el proceso editorial o por una lectura atenta.
- Hay otros errores, llamados mortales, que no pueden ser corregidos y vician insalvablemente los resultados y conclusiones.
- El lector crítico hace bien en incorporar entre sus competencias la capacidad de detectar estos errores y evitar que se traduzcan en cambios que podrían dañar a sus pacientes.

El uso de métodos estadísticos en la literatura biomédica se empezó a generalizar a contar de la segunda mitad del siglo XX. Tempranamente se pudo detectar que los errores eran frecuentes y que muchos de ellos hacían dudar de las conclusiones de buena parte de lo publicado y que ello ocurría generalizadamente, incluso en las revistas de mayor prestigio. Aunque pareciera ser que los errores más básicos han disminuido, el análisis estadístico en las revistas biomédicas se ha ido haciendo más complejo por lo que aumenta la oportunidad para que se multipliquen los errores y también la necesidad de un mayor conocimiento para detectarlos y sopesarlos en relación con las conclusiones alcanzadas por los autores.

Los errores son tantos que difícilmente se pueden explicitar todos aquí; incluso hay libros dedicados a ellos como el de Anderson o el de Good y Hardin. En un artículo como este, hay diversas maneras de aproximarse al tema. Por ejemplo, revisando los errores más frecuentes como los que

aparecen en la Tabla 1; algunos de estos errores pueden ser rectificadas en la etapa de revisión del trabajo o bien pueden ser advertidos por un lector crítico, pero sin duda reflejan descuido, ignorancia o mala intención. En general, se pueden dividir en aquellos derivados de un análisis inapropiado de los datos, aquellos relacionados a una presentación incompleta o poco clara, y aquellos causados por una interpretación no justificada. Muchos de ellos pueden ser corregidos en el proceso editorial si son detectados y algunos luego de una lectura crítica por el lector interesado. Por ejemplo, cuando en el trabajo se anota el error estándar como dimensión de la variabilidad de la muestra, cualquiera puede calcular la desviación estándar o la varianza al conocer el tamaño de la muestra.

Otra aproximación es la de destacar aquellos errores que invalidan las conclusiones y que en la literatura se han citado como “pecados o errores mortales” pues su presencia no puede deshacerse con un nuevo cálculo o discusión de los resultados. En general, estos se refieren a problemas en el diseño del estudio pues estos no pueden ser deshechos ni con el más sofisticado de los análisis (Tabla 2).

Esta revisión entonces sólo pretende hacer una

Tabla 1. Errores estadísticos corregibles, frecuentes en la literatura biomédica

1	Uso de error estándar en vez de la desviación estándar
2	Mala interpretación del significado de p
3	Uso inadecuado de pruebas estadísticas
4	Muestras pequeñas y ausencia de cálculo de potencia
5	Interpretación inadecuada de la correlación
6	Inferencia limitada o extralimitada

¹ Profesor Titular
División de Anestesiología. Pontificia Universidad Católica de Chile.

Tabla 2. “Errores mortales” de la literatura biomédica

1	Información insuficiente: un investigador independiente no puede replicar el trabajo
2	Muestras insuficientes: el error tipo II es inaceptable, pues la imprecisión de las estimaciones muy grande
3	Muestras no representativas; elección de controles inadecuados
4	Definiciones operacionales inadecuadas o insuficientes
5	Falla en medir variables que pueden tener incidencia en los resultados
6	Pérdidas en seguimiento o en la medición de variables
7	Mal manejo del sesgo y variables de confusión: inadecuado manejo del enmascaramiento y métodos de selección
8	Asignación al azar inexistente o inadecuada
9	Resultados vagos y descontrol con los números
10	Hipótesis y/o conclusiones errantes, cuando se plantea una en la introducción y se termina concluyendo otra cosa

enumeración somera de errores o defectos que se pueden encontrar en la literatura biomédica. Su amplia ocurrencia hace que la responsabilidad de detectarlos recaiga en los lectores que quieren evitar que dichos errores se trasladen al cuidado de sus pacientes. El énfasis está puesto en ayudar al lector a descubrir esos errores más que en quien quiera hacer y publicar un trabajo de investigación. Si bien el presente texto puede serles de utilidad, existen guías escritas por diversos grupos de expertos como el CONSORT, QUORUM, PRISMA o STARD que servirán mejor a ese propósito (<http://www.consort-statement.org/>, <http://www.prisma-statement.org/>, <http://www.stard-statement.org/>).

Los errores se presentan en este artículo como un catálogo ordenado según la parte del manuscrito donde suelen aparecer: diseño y descripción del estudio, análisis de los datos, descripción de los métodos estadísticos usados, presentación de los resultados e interpretación de estos. Esto deja afuera errores que se pueden cometer en otras etapas del trabajo, como en las mediciones o en la recolección y registro de los datos, pues no quedan reflejados en el trabajo publicado. No hay intento de explicar con detalle porqué constituyen errores pues extendería demasiado este artículo y porque esas justificaciones aparecen en los capítulos respectivos. Por ejemplo, porqué es un error usar el error estándar de la media en lugar de la desviación estándar para describir la variabilidad de la muestra que aparece en el artículo sobre la distribución normal. Esta decisión puede disminuir la legibilidad y quizás la amenidad de este artículo, pues estamos conscientes que quienes pueden considerar ameno un artículo de errores estadísticos son decididamente una minoría, pero facilita su uso como una lista de verificación para una lectura crítica.

ERRORES EN EL DISEÑO

Repetiendo un concepto dicho muchas veces antes, en otros artículos, los errores y olvidos cometidos en la planificación del trabajo muchas veces determinan que su ejecución y resultados no serán capaces de responder a las preguntas planteadas. Esto se traduce en riesgos para los sujetos del estudio y en gastos inútiles, el escándalo de la investigación mal hecha que ha sido denunciado por muchos. Por ello, el tiempo y pensamiento, incluyendo una asesoría estadística oportuna, que se inviertan en esta etapa son esenciales para mejorar las probabilidades que el trabajo pueda representar un aporte real al conocimiento. Son errores frecuentes de diseño:

- La hipótesis principal y/o las secundarias no son explícitas o claramente reconocibles. Esto impide seguir el hilo del trabajo, desde la pregunta a la respuesta o conclusión, y no permite evaluar la justedad del análisis estadístico usado.
- Las definiciones operacionales, especialmente los criterios de inclusión y exclusión, son vagas, lo que impide replicar el trabajo y, quizás peor, la generalización, pues la población de interés queda indefinida.
- No queda claro qué se midió, cuándo y cómo se midió; particularmente frecuente es la ausencia de datos sobre el error y calibración de los métodos de medición. Por ejemplo, son frecuentes frases como “se midió la presión arterial antes y después de tal o cual procedimiento”, sin precisar quién hizo la medición, en qué condiciones, con qué aparato y si este o estos fueron calibrados o no.
- No siempre se describe con detalle el método usado para la aleatorización y el reclutamiento o elección de las muestras. Frases como “se hizo una asignación aleatoria”, sin especificar

- cómo se hizo, son insuficientes.
- Con frecuencia tampoco se describen los métodos usados para asegurar el ciego de los sujetos y el ciego de quienes realizan las mediciones o estimaciones.
- No hay referencia a resultados previos, cuál es el efecto principal estimado o esperado y tampoco referencia al cálculo del tamaño muestral basado en esta presunción.
- No se define bien el grupo control o este es inapropiado para la pregunta que se quiere contestar. Esto deja a las conclusiones sin sustento.
- Muchos trabajos son demasiado pequeños para detectar efectos importantes. Se falla en reportar el tamaño inicial de la muestra y dar cuenta de las pérdidas ocurridas en el desarrollo del estudio.
- No hay demostración de la comparabilidad basal de los grupos o de los ajustes hechos cuando hay diferencias evidentes. Otras muchas veces, se dice que “no había diferencias basales entre los grupos” sin que se haga una consideración sobre un posible error tipo II derivado de un tamaño muestral reducido.

ANÁLISIS DE LOS DATOS

En general, los errores más frecuentes se deben a que se ignoran los principios generales básicos del diseño experimental y de la bioestadística, tales como la relación que existe entre el tipo de estudios y las preguntas que pueden o no contestar, el tipo de análisis apropiado para cada caso, o a no comprobar que se cumplen las presunciones o requisitos de cada test. La amplia disponibilidad de programas computacionales que permiten analizar los datos con facilidad ha aumentado este tipo de errores. A esto se agrega que ese desconocimiento impide que los autores hagan una descripción precisa de los

métodos usados y de su justificación cuando esta es necesaria. Son errores frecuentes de análisis de datos:

- Inadecuada descripción de los métodos usados. Un investigador independiente debería ser capaz de reproducir el análisis de los datos basándose en la descripción del manuscrito. Es de especial preocupación el hecho que los errores sean frecuentes incluso cuando se trata de métodos básicos pues la creciente complejidad de los análisis aumenta las probabilidades de error.
- En general, los errores más frecuentes se deben a no comprobar que se cumplan las presunciones o requisitos de cada test: distribución normal o no, semejanza de las varianzas, escala de medición de los datos, tamaño de las muestras, número de grupos, análisis de una o dos colas, datos pareados o independientes, número de comparaciones, presencia de una tendencia en categorías ordenadas, linealidad de la relación entre dos o más variables.
- Son particularmente frecuentes los errores que se relacionan con el uso e interpretación del test t de Student, correlación (Tabla 3) y chi cuadrado (Tabla 4). Están descritos y resumidos en los respectivos artículos.
- Muchas veces se reduce la precisión de la medición y se pierde información cuando se reducen datos continuos a ordinales sin explicación ninguna. Además, se generan confusiones con los límites entre cada grupo que no siempre son explicados.
- Nombres erróneos de tests o, con mayor frecuencia, incompletos. Por ejemplo, limitarse a decir “test t de Student” sin especificar si es independiente o pareado o de una o dos colas. Esto es especialmente importante en esta era de acceso fácil a programas computacionales que a un set de datos permiten aplicar muchos tests o, peor, arrojan simultáneamente resulta-

Tabla 3. Mal uso de la correlación

1	No analizar nube de puntos
2	Observaciones no independientes: más de un valor por individuo
3	Dragado de datos: múltiples correlaciones
4	Efecto del tiempo
5	Muestreo restringido
6	Muestras mezcladas
7	Evaluación de cambios en relación con un valor inicial
8	Relacionar una parte con el total
9	Uso en la validación de un método de medición (se debe usar el método del agreement de Bland y Altman)

Tabla 4. Mal uso de chi cuadrado

1	Uso de porcentaje o proporciones en el cálculo de χ^2
2	Uso con n pequeños; no usar si: - n total < 20 - Más del 20% de las celdas con n esperado ≤ 5
3	No usar corrección de Yates (n total < 100 o cualquier celda ≤ 10)
4	Análisis post hoc tablas 2 x n
5	No tomar en cuenta categorías ordenadas en tablas 2 x n

dos de muchos tests. Otros errores en el uso de las pruebas es no usar corrección por comparaciones múltiples, uso inapropiado de un análisis *post hoc* de subgrupos y no explicar las razones por las cuales se decidió usar un análisis o una prueba inusual.

- Cuando hay diversas comparaciones posibles, con frecuencia no se especifica los tests usados en cada una de las comparaciones. En este sentido la expresión “donde apropiada” o similares debiera ser evitada pues deja en la incertidumbre si el uso fue realmente apropiado.
- Con frecuencia aparecen en los resultados variables que no aparecen descritas antes. Además de dejar en la duda cómo se obtuvieron, revela falta de prolijidad en el manejo de los datos.
- Falla en describir si hubo pérdida de datos y cómo se lidió con este problema. Otros, usan soluciones inadecuadas, como omitir casos incompletos o usar la imputación simple. Muchos programas estadísticos, especialmente aquellos usados en el análisis de medidas repetidas, simplemente omiten los sujetos que no tienen todas las mediciones completas; el programa entrega igual una p pero su significado es incierto y los autores pueden no darse cuenta de la disminución del tamaño muestral que se refleja en los grados de libertad usados en el cálculo del estadístico en cuestión.
- Otro error relativamente frecuente es el confundir las unidades de medición en el análisis y presentación de los resultados. Y otro, particularmente importante en análisis de variable múltiple, es confundir el tamaño de la muestra con el número de resultados de interés (por ejemplo, cuando se reclutan 1.000 pacientes pero sólo hubo 18 con un infarto del miocardio) o confundir variables de resultado como predictoras (por ejemplo, duración de la cirugía o hipotensión intraoperatoria en índices de riesgo

preoperatorio).

- Uso de “*propensity scores*” sin consideración sobre los sesgos potenciales de factores de confusión no medidos; el uso de la técnica podría aumentar los sesgos en estas circunstancias.
- Es frecuente el error en el análisis de datos seriados donde existen dificultades de análisis e interpretación que no son advertidos excepto por estadísticos profesionales, por lo que estos deben estar entre los autores o figurar como consultores excepto si se ha recurrido a medidas sumatorias en el análisis.
- Falla en el manejo de datos agrupados (“*clustering*”) cuando esto ocurre, por ejemplo, cuando los datos se registran a lo largo del tiempo y los datos sucesivos dependen de los previos, situación muy frecuente en estudios médicos. Por otro lado, muchas veces se realizan análisis de subgrupos sin que hayan sido especificados en el protocolo y se falla en especificar su carácter exploratorio.

DOCUMENTACIÓN Y PRESENTACIÓN

El objetivo de un escrito es entregar toda la información necesaria para reproducir un trabajo en detalle, entregar todos los resultados pertinentes, seguidos de una discusión que soporte las conclusiones. Todo esto, que no es poco, en forma clara, precisa y concisa, para que se entienda perfectamente bien ocupando el menor espacio posible. Son errores frecuentes de documentación y presentación:

- La repetición de datos en texto, tablas o gráficos, que alarga innecesariamente la presentación y muchas veces la hace más confusa. El problema se agrava si no concuerdan entre si.
- Otro problema frecuente es que la descripción de lo que se hizo no deja clara la línea de tiempo en la ejecución del estudio.
- En gráficos y figuras el principio general es que, excepto aquellas muy complejas, deben entenderse sin recurrir al texto del artículo. En su confección se repiten algunos errores: uso inapropiado de las escalas en gráficos, en particular la decisión de no partir desde cero.
- No indicar la variabilidad de los datos. También es un error hacerlo inapropiadamente: usar varianza o desviación estándar cuando no hay o no se comprobó una distribución normal o usar el error estándar de la media (ESM) para describir la variabilidad de la muestra; el ESM se refiere a la precisión de la estimación de la media poblacional hecha a través de la media muestral.

- No analizar los datos para detectar valores extremos (“*outliers*”) o no hacer mención que se hizo, por ejemplo, con algún tipo de gráfico. También es un error, no explicar qué se hizo con éstos si aparecieron.
- No usar *p* exactas y citarlas como menor o mayor al límite α elegido ($< 0,05$ o $> 0,05$ u otro). Este modo de citar la *p* se justificaba cuando se consultaba su valor en tablas poco precisas; los programas computacionales hoy en uso arrojan el valor de *p* preciso con hasta tres decimales o más. Por otro lado, también es un error usar niveles de precisión innecesarios o más allá de la precisión con que se midieron las variables; por ejemplo, citar el peso o la edad con dos decimales cuando se midieron en kilos o al último cumpleaños.
- Sólo citar el valor de *p* y no los intervalos de confianza, especialmente aquellos referidos a una diferencia en el caso de comparaciones.
- Un error frecuente en trabajos que comparan drogas es el no comunicar los efectos adversos o complicaciones observadas. Así, un trabajo puede demostrar la eficacia de una droga en reducir la presión arterial pero no permite hacer una estimación de la frecuencia o gravedad de los efectos colaterales.

INTERPRETACIÓN

Los errores cometidos en etapas previas generalmente desembocan en errores en la interpretación de lo que los resultados significan y, por ende, en la explicación de las conclusiones presentadas. Esta situación, en el mejor de los casos, significa

que todo el esfuerzo, gasto y riesgos del estudio fueron en vano; en el peor, las conclusiones erradas, sin sustento, pueden trasladarse al cuidado de los pacientes con consecuencias insospechadas. Son errores frecuentes de interpretación:

- Una *p* no significativa es interpretada como que no hay diferencia o como demostración que no hay efecto.
- Con una *p* no significativa no se comenta la posibilidad de un error tipo II por un tamaño muestral insuficiente: no se calcula la potencia del experimento. Otras posibilidades derivan de un uso inapropiado de pruebas estadísticas: por ejemplo, usar un test no paramétrico cuando se pudo usar la alternativa paramétrica, o usar un test de dos colas cuando se debió usar uno de una cola.
- Por otro lado, también es un error frecuente interpretar una *p* significativa como significado clínico o de magnitud de efecto. Es frecuente la frase “el efecto fue muy significativo” cuando está asociado a una *p* muy pequeña. Tan frecuente como los anteriores es interpretar como efecto demostrado cuando la *p* fue “casi” significativa o “tendió a ser significativa”.
- Todo trabajo debe discutir limitaciones, posibles sesgos o factores de confusión: su ausencia constituye un error que impide al lector aquilatar la solidez de las conclusiones.
- Generalizar o extrapolar más allá de lo que la muestra y los datos permiten. Un error particularmente frecuente y grave es inferir causalidad desde la correlación o la regresión.
- Son frecuentes los errores en la interpretación de aumento o disminución del riesgo asociado a una intervención, particularmente la diferencia entre riesgo absoluto y riesgo relativo.

REFERENCIAS

- Altman DG, Bland JM. Generalisation and extrapolation. *BMJ* 1998; 317: 409-410.
- Altman DG. Poor-Quality Medical research. What Can Journals Do? *JAMA* 2002; 287: 2765-2767.
- Altman DG. Statistics and ethics in medical research. VII-Interpreting results. *BMJ* 1980; 281: 1612-1614.
- Altman DG. The scandal of poor medical research. *BMJ* 1994; 308: 283-84.
- Anderson B. Methodological errors in medical literature. An incomplete catalogue. Oxford:Blackwell, 1990.
- Bacchetti P. Current sample size conventions: Flaws, harms, and alternatives. *BMC Medicine* 2010; 8: 17.
- Chan AW, et al. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008; 337: a2299.
- Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. *J Prosthodontic Res* 2011; 55: 1-6.
- Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AHS. Common statistical and design problems in manuscripts submitted to high-impact medical journals. *BMC Research Notes* 2011; 4: 304.
- Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation* 1980; 61: 1-7.
- Good PI, Hardin JW. *Common Errors in Statistics (and How to Avoid Them)*. Hoboken: Wiley & Sons, 2003.
- Greenfield MLVH, Rosenberg AL, O'Reilly M, et al. The Quality of Randomized Trials in Major Anesthesiology Journals. *Anesth Analg* 2005; 100: 1759-1764.
- Habibzadeh F. Common statistical mistakes in manuscripts submitted to biomedical journals. *European Science Editing* 2013; 39: 92-94.
- Kozak M. How to Show That Sample Size Matters. *Teaching Statistics* 2009; 31: 52-54.
- Lang T. Twenty statistical errors

- even YOU can find in biomedical research articles. *Croat Med J* 2004; 45: 361-370.
16. Martínez-Sellés M, Prieto L, Herranz I. Frequent mistakes in the statistical inference of biomedical data. *Ital Heart J* 2005; 6: 90-99.
17. Morgan PP. The seven deadly scientific sins of clinical studies. *CMAJ* 1986; 134: 1225.
18. Pua HW, Lerman J, Crawford MW, Wright JG. An Evaluation of the Quality of Clinical Trials in Anesthesia. *Anesthesiology* 2001; 95: 1051-1053.
19. Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001; 165: 1226-1237.
20. Simundic AM, Nikolac N. Statistical errors in manuscripts submitted to *Biochemia Medica*. *Biochem Med* 2009; 19: 294-300.
21. Vaisrub N. Manuscript Review From a Statistician's Perspective. *JAMA* 1985; 253: 3145-3147.
22. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals. A Systematic Sampling Review. *JAMA* 2007; 297: 1233-1240.

Correspondencia a:
Dr. Jorge Dagnino S.
jdagnino@med.puc.cl