

metodológicos a manera de ritual^{2,3}. Con no poca frecuencia se emplean con la ingenua convicción de que proveerán mecánicamente la respuesta adecuada a las preguntas formuladas^{4,5}, más en virtud de una tradición o de una expectativa externa que como un recurso necesario para solucionar el problema de investigación encarado. Tales modas son a menudo estériles, y hasta contraproducentes, ya que pueden terminar por engrosar el caudal de conocimientos erróneos sobre una materia⁶.

Estos problemas se han estudiado desde muchos años atrás, pero más recientemente se han adicionado sólidas evidencias⁷ acerca de su pernicioso impacto sobre la reputación de la estadística en particular y los intereses de la ciencia en general.

En consonancia con el reconocimiento de las deformaciones apuntadas y con el fin de revertirlas, se han desarrollado y puesto en práctica diversas guías para los autores con indicaciones orientadas a optimizar sus manuscritos. Se recomienda asimismo que autores, revisores y editores las empleen en el proceso de confección y evaluación de los trabajos⁸. Estas guías se han ido perfeccionando con el tiempo, y su número ha aumentado a medida que se van refinando para los diferentes tipos de investigación; sin embargo, las dos más destacadas y aceptadas son CONSORT (*Consolidated Standards of Reporting Trials*)⁹ para los ensayos clínicos y STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*)¹⁰ para los estudios observacionales. Una importante parte de estas guías aborda cómo exponer los datos y procedimientos estadísticos. Como es natural, en este contexto comparece el problema de decidir el tamaño muestral, un asunto que desde el punto de vista cualitativo y económico es relevante, que el investigador necesariamente tiene que encarar (de manera implícita o explícita) siempre que se proponga trabajar con datos empíricos. Se trata, sin embargo, de un tema muy controvertido¹¹.

Las guías de publicación antes mencionadas, a las que se adhieren las revistas más prominentes en investigación sanitaria, exigen que los artículos contengan una justificación detallada acerca de cómo fue determinado el tamaño de muestra empleado. Pero dan por sentada la conveniencia de ofrecer tal justificación sin que se discuta o explique por qué pudiera ser útil. Es llamativo que en el propio CONSORT, un documento muy explícito en diversas áreas, no figure un solo argumento que respalde tal demanda.

Que los investigadores deben decidir se pesadamente alcanza está fuera de discusión. Sin embargo, la racionalidad de la exigencia de exponer la forma en que se calculó el tamaño muestral ofrece dudas. No conocemos de un examen abarcador e integral de este problema, que lo enjuicie desde una perspectiva epistemológica, ética y práctica. En torno a él se abren varios interrogantes a los que se procura responder: ¿puede el procedimiento metodológico ser racional, o constituye otro procedimiento metodológico ritual? ¿Cuál es su supuesta utilidad? ¿En qué grado tales demandas son efectivamente satisfechas por autores y exigidas por árbitros y editores de las revistas más destacadas en la actualidad? ¿Hay diferencias en el patrón que muestra la situación según se trate de trabajos observacionales o experimentales? En consecuencia, el presente artículo tiene dos componentes diferentes, aunque íntimamente vinculados: la discusión conceptual en torno a la mencionada exigencia y el grado en que se cumple.

Métodos

Con el fin de identificar los trabajos en que se aborda la necesidad de exponer cómo se determinó el tamaño de la muestra se realizó una búsqueda bibliográfica en las bases especializadas EBSCO, PUBMED y MEDLINE usando *sample size* y sin imponer restricciones temporales. A partir de esa información se identificaron los razonamientos que pudieran haberse esgrimido para respaldar

las exigencias mencionadas, los cuales se sometieron a un análisis crítico.

Para estimar el grado de cumplimiento de la recomendación de ofrecer los datos en que se basó la decisión sobre el tamaño muestral, se realizó un estudio bibliométrico de naturaleza descriptiva. Para ello se eligieron las seis revistas de más alto factor de impacto en el campo de la salud al cierre de 2009, de acuerdo con el *Journal Citation Reports*¹²: *The New England Journal of Medicine*, *The Lancet*, *The Journal of the American Medical Association*, *Annals of Internal Medicine*, *British Medical Journal* y *PloS Medicine*. Los dos autores del estudio examinaron todos los artículos originales publicados a lo largo de ese año que figuraban en las respectivas secciones destinadas a «artículos originales» y que reportasen investigaciones para las cuales hay estándares inequívocos, contenidos en las guías de publicación, sobre la determinación del tamaño muestral. Se incluyeron tanto los trabajos de naturaleza observacional (de cohortes, de casos y controles, y transversales o de prevalencia) como los experimentales (ensayos clínicos controlados aleatorizados de fase III o IV, estudios preclínicos, investigaciones básicas, estudios de eficacia e intervenciones comunitarias), siempre que implicasen el estudio de unidades de análisis individuales y no fuesen metaanálisis ni revisiones sistemáticas.

Para cada artículo seleccionados se establecieron dos aspectos fundamentales: si se justifica o no el tamaño muestral (en particular si se informa del método empleado para la determinación) y si se comunican o no todos los datos necesarios para el cómputo.

Se realizaron estimaciones de porcentajes y distribuciones por tipo de estudio y revista. Se computaron intervalos de confianza del 95% (IC95%) para las estimaciones fundamentales usando el programa EMUCO 2.1¹³.

Resultados

Fundamentación de las exigencias vigentes

Goodman y Berlin¹⁴, si bien enfatizan que la aparición de los cálculos realizados para determinar el tamaño muestral carece de utilidad alguna a los efectos de interpretar los resultados, esbozan algunas razones de dicheos detalles justificarla. Sostienen que la comunicación de dichos detalles pudieran justificarla. Sostienen que la planificación y de la ejecución de la investigación, la cual podría verificarse cotejando el tamaño muestral planificado a priori con el de la muestra finalmente empleada. Añaden que, si estas consideraciones se presentan de manera sofisticada, ello podría aumentar nuestra confianza en que los aspectos metodológicamente complejos fueran bien manejados.

Schultz y Grimes¹⁵ afirman que cuando dicha información aparece en los artículos (específicamente en los que reportan ensayos clínicos controlados) «el lector encuentra en él la información referente al desenlace primario, a si el ensayo presentó dificultades para el reclutamiento o si terminó temprano debido al hallazgo de un resultado estadísticamente significativo», y le atribuyen un valor añadido como posible alerta, ya que «en caso de faltar esta información, los lectores deben interpretar cautelosamente los resultados reportados». Añaden que negarse a reportar el cálculo del tamaño muestral indicaría una «ingenuidad metodológica» que pudiera señalar otros problemas.

El otro polo del debate se expresa embrionariamente en algunos artículos recientes¹⁶⁻¹⁸ en los cuales se ha comenzado a reaccionar contra los rituales arraigados en esta materia. En los dos primeros se reflexiona sobre el daño que las convenciones de investigación al tamaño muestral producen al proceso de investigación; se fundamenta que los informes de estudios completos no deberían incluir la información relativa a cómo se determinó el tamaño muestral, dado que su exigencia promueve la interpretación incorrecta de los

Tabla 1
Artículos originales examinados según revista. Año 2009

Revista	Artículos examinados	(%)	Total
An Intern Med	73	(80)	91
BMJ	169	(69)	244
JAMA	158	(90)	175
Lancet	129	(76)	169
New Engl J Med	176	(78)	226
PloS Med	49	(64)	77
Total	754	(77)	982

estudios, inhibe la innovación, erosiona la integridad de los investigadores que recurren con frecuencia a una variedad de estrategias manipuladoras de los cálculos para justificar un tamaño de muestra que fue realmente elegido por otros motivos, y provee a los revisores de un poder arbitrario. El tercero advierte de que este poder es especialmente bienvenido por las grandes empresas farmacéuticas, que cincelan la idea de que los esfuerzos que no salen de «grandes» estudios (que sólo ellos pueden realizar) son inútiles.

Bacchetti¹⁹ alerta sobre un importante problema derivado de esta exigencia formal: el efecto negativo que su vigencia suele tener en el proceso de inherencia por pares de los artículos. Apoyados en la ambigüedad inherente a la planificación del tamaño muestral, los revisores pudieran siempre objetar dicho tamaño, y no pocas veces lo hacen, incluso sin invocar argumento alguno o esgrimiendo los que dimanan de su propia subjetividad.

Información sobre la determinación del tamaño de muestra en las revistas de alto impacto

Durante el año 2009 se publicaron 982 artículos originales en las seis revistas médicas seleccionadas. Tras aplicar los criterios de exclusión establecidos, se mantuvieron para el análisis 754 (77%) (tabla 1). El análisis se realizó sobre estos 754 trabajos, entre los que predominaban los estudios observacionales, que ascendieron a 415 (55%); de los otros 339, de naturaleza experimental, 327 (97%) eran ensayos clínicos controlados aleatorizados.

Se delimitó el grado en que los artículos que se publican en la actualidad cumplen las dos exigencias establecidas por las revistas: que se explique cómo se arribó al tamaño muestral y que se den los datos necesarios para que el lector pueda corroborar los cálculos.

En el 56% (IC95%: 52-59) de los trabajos no se ofrecían razones que justificasen el tamaño de muestra anunciado. Con excepción de una de las seis revistas estudiadas (*Lancet*), en menos de la mitad de los artículos se justifica el tamaño muestral. La ausencia de justificación fue notablemente mayor entre los trabajos observacionales (92%; IC95%: 89-94) que entre los experimentales (12%; IC95%: 8-15) (tabla 2).

De los 333 artículos en que se justificaba el tamaño muestral, el 98% empleaba fórmulas con ese fin, y de éstos (327) eran observacionales 34 y ensayos clínicos 293. Sólo en 16 de los primeros (47%; IC95%: 45-49) y en 143 de los segundos (49%; IC95%: 43-55) se comunicaron todos los datos necesarios para reproducir el cálculo.

Tabla 2
Artículos examinados en que no se justifica el tamaño de muestra, según tipo de estudio y revista. Año 2009

Revista	Observacionales			Experimentales			Totales			
	No se justifica	(%)	Total	No se justifica	(%)	Total	No se justifica	(%)	Total	IC95%
An Intern Med	40	(98)	41	3	(9)	32	43	(59)	73	48-70
BMJ	110	(96)	115	1	(6)	54	111	(66)	169	59-73
JAMA	90	(84)	107	3	(6)	51	93	(59)	158	51-67
Lancet	40	(85)	47	6	(7)	82	46	(36)	129	27-44
New Engl J Med	69	(97)	71	22	(21)	105	91	(52)	176	44-59
PloS Med	32	(94)	34	5	(33)	15	37	(76)	49	63-88
Total	381	(92)	415	40	(12)	339	421	(56)	754	52-59

IC95%: intervalo de confianza del 95%.

Este es el panorama general, pero no todas las revistas examinadas hacen las mismas exigencias: si bien las seis emplean CONSORT para los ensayos clínicos, ni *JAMA* ni *New Engl J Med* se adhieren explícitamente a STROBE para los observacionales. Tal circunstancia determina que sólo para 564 de los 754 artículos estudiados puede examinarse el grado en que las revistas son consecuentes con sus propias exigencias en materia de comunicación del tamaño de muestra: para el dicho tamaño y también comunicar los datos empleados para el cálculo.

Sólo 151 (27%; IC95%: 23-30) de los trabajos cumplen con ambas cosas. Tal cumplimiento es virtualmente inexistente en los estudios observacionales (3%; IC95%: 1-6), pero también reducido en los experimentales (44%; IC95%: 38-49).

Discusión

La endeblez de los argumentos que se han esgrimido para aconsejar que los autores expliciten las fórmulas y los valores usados para decidir el tamaño de muestra es notoria: justificar cómo se determinó el tamaño muestral no es prueba de calidad metodológica, y tampoco es necesario hacerlo para informar debidamente sobre la variable principal de respuesta ni sobre la magnitud de la muestra planificada. Obviamente, se pueden comunicar el tamaño planificado y el que realmente se empleó sin necesidad de explicar cómo se llegó al primero de ellos. La propia orfandad de argumentos racionales en defensa de tal comunicación sugiere su carácter formal y su naturaleza ceremonial.

Por otra parte, los autores pueden haber realizado un ajuste retrospectivo (*retrofitting*) para lograr a posteriori que ambos valores se correspondan. Schultz y Grimes¹⁵ llaman la atención sobre esta realidad cuando comentan la extendida propensión a hacer ajustes retrospectivos, suscritos por diversos autores como *the sample samba*. Diversos autores suscriben este punto de vista^{5,14,16,17,20,21}. En tal caso, el tamaño muestral figurará como algo planificado de manera coherente con el que realmente se empleó, de modo que la aparición formal de ciertas fórmulas dista de ser garantía de una conducción metodológicamente rigurosa. Expresa más bien que los investigadores no incurren en la «ingenuidad» que Schultz y Grimes aconsejan evitar. Es bien sabido, aunque no suela reconocerse con claridad, que la determinación del tamaño de muestra habitualmente está gobernada por la factibilidad²².

A nuestro juicio, el *retrofitting* es algo bastante natural en realidad. Puesto que la determinación del tamaño muestral es intrínsecamente especulativa, sería absurdo pedir a un investigador que no especule. Lo que suele pedírsele en la práctica termina por ser, en no pocos casos, que simule que no lo hace. Frente a la «dictadura» de los requisitos y sus demandas no lo hace. Los autores muchas veces se «defienden» acudiendo a la inclusión ceremonial de fórmulas que justifiquen el tamaño muestral. Es verosímil que ocurra lo contrario: como se exige determinada conducta, el autor puede verse tentado a presentar la información con el fin de mostrar que «cumple» con ella. Ante una exigencia burocrática, el investigador ofrece una solución burocrática.

Ocasionalmente se hacen juicios según los cuales «el tamaño de muestra es insuficiente». Éste sigue siendo un criterio valorativo muy consolidado en comités de ética, agencias de evaluación de proyectos de investigación y árbitros o editores de revistas. La legitimidad de este tipo de pronunciamientos es, sin embargo, muy discutible. Tales juicios son conflictivos y casi siempre improcedentes por al menos dos razones. En primer lugar, porque la información disponible para determinar con precisión el tamaño muestral «correcto» (un calificativo polémico de por sí) es a menudo menor que la necesaria o directamente inexistente, y en cualquier caso está inexorablemente atravesada por la subjetividad inherente a la elección de los datos que han de emplearse para el cálculo^{11,15,23}. A ello se suma que el pobre dominio de la estadística por parte de muchos revisores²⁴ suele generar críticas injustas y reclamos infundados a los autores. En segundo lugar, porque la noción de «tamaño suficientemente grande» sólo tendría sentido en el supuesto de que pudieran sacarse reglas operativas o conclusiones definitivas de cada estudio aislado. Sin embargo, ésta es una ilusión tan arraigada como incorrecta, por la simple razón de que la ciencia no funciona así.

Nuestras convicciones científicas pueden ser más o ser menos firmes, pero siempre son provisionales, y nuestras representaciones de la realidad tienen en cada momento un cierto grado de credibilidad, pero están abiertas a cambios y perfeccionamientos en la medida en que nuevos datos lo aconsejen. La consolidación del nuevo conocimiento es gradual y cualquier aporte es bienvenido. Unos serán más trascendentes y otros menos, pero todos pueden hacer alguna contribución en este proceso, independientemente del tamaño de la muestra, ya sea por conducto del metaanálisis²¹ como por medio del enfoque bayesiano²⁵, por mencionar las dos opciones más conocidas.

Este elemento es relevante, ya que el tamaño de muestra mínimo que exigen los cánones es aquel lo suficientemente grande como para «detectar» diferencias, expresión ambigua que en realidad alude a un tamaño lo bastante grande como para declarar que la diferencia que mide el efecto es estadísticamente significativa. Es decir, la pertinencia de cierta magnitud del tamaño muestral se subordina casi siempre al dogma de la significación estadística¹⁷. Desde muy temprano se enfatizó²⁶ que cualquier diferencia observada diferirá de manera estadísticamente significativa de la nulidad siempre que el tamaño muestral sea lo suficientemente grande. Siendo así, sorprenden testimonios tales como «es muy frecuente hallar artículos con tamaños de muestra insuficiente para *detectar* los efectos que valoran»²⁷, ya que en rigor esa es la situación no de muchos trabajos donde no se encuentra significación, sino de absolutamente todos ellos.

Lo más atinado es emplear procedimientos que, en lugar de convocarnos a desechar o no una hipótesis, se concentren en la estimación del efecto, tal y como sugieren desde 1988 las recomendaciones del International Committee of Medical Journal Editors²⁸, y con esa información «poner al día» la opinión que determinada hipótesis o punto de vista nos merecen a la luz de los nuevos datos.

En cualquier caso, lo más importante es que los resultados de un estudio pueden ser interpretados de manera adecuada sin contar con la justificación del tamaño de la muestra; lo verdaderamente necesario para ello es conocer las estimaciones realizadas (ya sea de los parámetros o de los efectos) con sus correspondientes intervalos de confianza. La «cautela» que reclaman Schultz y Grimes¹⁵ en la valoración de los resultados no ha de modificarse en función de que se expliciten o no los cálculos que se supone dieron lugar al tamaño de, de forma más precisa, los márgenes de error que de él se derivan.

Al demandar que los investigadores expliquen en los artículos cómo se aplicaron las fórmulas para determinar el tamaño de la muestra, las propias revistas médicas han estado

contribuyendo a la perpetuación del uso ritual de procedimientos estadístico-metodológicos. Como consecuencia, la demanda que estamos valorando podría generar más problemas de los que resuelve en lo que se refiere al reporte transparente de la investigación¹⁶.

El cumplimiento en la práctica de lo pautado acerca de lo que supuestamente ha de comunicarse sobre el tamaño muestral ha sido examinado en profundidad en años recientes, en especial en el área de los ensayos clínicos. Es notable el grupo de trabajos que encuentran prevalencias de justificación del tamaño muestral muy bajas. Un repaso a los más recientes arroja unas altas tasas de incumplimiento en el marco de los ensayos clínicos (nunca inferiores al 50% y con frecuencia por encima del 70%). En el más reciente de ellos, Hopewell et al²⁹ hallaron que el cálculo del tamaño muestral se explicaba sólo en el 45% de los informes de ensayos clínicos indexados en PubMed en 2006, mientras que, en este mismo tipo de estudios, Mills et al³⁰ encuentran que el tamaño muestral se justifica en apenas una quinta parte de los 116 que habían sido indexados en 2002, y Charles et al³¹ observan un patrón similar. Tal panorama concuerda por completo con lo hallado en la presente investigación.

Si nos ceñimos a los artículos que incluyen las fórmulas empleadas para llegar al tamaño muestral, una fracción muy notable reporta de manera incompleta los datos necesarios para corroborar dicho cálculo. Esencialmente, el mismo patrón se expresa con regularidad a lo largo de los últimos años^{32,33}. De nuevo, nuestros datos se ajustan en su totalidad a ese patrón en lo que concierne a los estudios experimentales. La literatura consultada no aborda el tema para los estudios observacionales, para los cuales nuestro estudio encuentra incumplimientos muchísimo mayores. Tal realidad puede deberse a que la insistencia en la necesidad de justificar tamaños, como figura en la literatura de STROBE, ha sido históricamente mucho menor, y a que la controvertida idea de que de cada trabajo aislado hay que sacar una conclusión tiene muchísimo menos arraigo.

En resumen, la justificación del tamaño de la muestra en la investigación sanitaria contemporánea aparece escasamente y con mucha frecuencia de manera incompleta en los artículos. Es un hecho que ni revisores ni editores toman muy en serio la «obligación» de explicar en los trabajos científicos cómo determinar el tamaño de la muestra, aun cuando ello figura en las guías de publicación a las que se adhieren las propias revistas biomédicas que acogen los trabajos.

Que el tamaño muestral no se justifique en más de la mitad de las publicaciones más encumbradas de la investigación sanitaria contemporánea constituye un indicio de que tal información es esencialmente estéril y confirma de manera indirecta su irrelevancia superflua. Si fuera realmente necesario o importante, sólo excepcionalmente se permitiría tal violación. Pero el hecho es que se trata de una información que no agrega nada sustantivo a lo que informa el trabajo sobre la pregunta que éste procura responder, ni a la valoración de su calidad.

El modo en que se arribó a un tamaño muestral es crucial para quien ha de decidir si el trabajo se hace o no, si se financia o no. Pero, una vez realizado, el tamaño es algo ya consumado y lo que interesa al lector es saber qué puede sacar en limpio de lo que objetivamente se ha hecho, no las motivaciones que pudo haber para tomar una u otra decisión (sea del investigador o del financiador) en esta materia. Bien pudiera ocurrir que el tamaño real haya sido la mitad de lo inicialmente deseado en virtud de limitaciones financieras; o que fuera el triple, debido a que el autor decidió emplear el resultado de una herencia en otro caso. Conocer esas razones es tan inútil e irrelevante como saber por qué el número de autores del trabajo fue el que fue, o por qué se seleccionó determinado algoritmo para la generación de números aleatorios.

El tamaño empleado influirá en la calidad de las estimaciones (reflejado en la amplitud de los correspondientes intervalos de confianza) y en la potencia de las pruebas de hipótesis. Tales datos permiten de por sí una interpretación directa y confiable de dichos resultados^{16,34}, pero conocer o no por qué se eligió el tamaño en cuestión no cambia esta interpretación.

En nuestra opinión, defender dicha demanda implica coonestar y hasta promover algunos de los cultos al valor p, en torno a la estadística, tales como el culto al valor p, en torno a cuyo manejo también se incumplen en grado muy considerable las recomendaciones de los propios editores de revistas médicas³⁵.

Conclusiones

La demanda de explicar en un artículo publicado cómo se llegó a cierto tamaño muestral (incluida en las guías de publicación a las que se adhieren la mayoría de las revistas biomédicas) carece de fundamento y utilidad. Por añadidura, contribuye al empleo ceremonial de la estadística y a la sumisión al dogma del valor p, con lo cual nos divorciamos del pensamiento racional. Su vigencia no promueve, sino que más bien menoscaba la racionalidad que reclama el reporte de las investigaciones. Tal exigencia, en suma, debería ser abandonada y, en consecuencia, suprimida de las guías.

¿Qué se sabe sobre el tema?

Desde el punto de vista teórico, hay muy pocas reflexiones en la literatura de muestra la exigencia de hacer trabajos explícitos en el tamaño de muestra con que han trabajado. Siendo así, sobre la fundamentación de hacerlo se sabe muy poco. En cuanto al grado en que tal exigencia se cumple, sí hay varios trabajos, pero se ciñen por lo general a áreas específicas (cirugía, pediatría, etc.) y a los entornos clínicos. Por otra parte, estos esfuerzos datan por lo general de un entorno del año 2000 (5 años).

¿Qué añade el trabajo realizado a la literatura?

Lo que se añade en esencia es un juicio conceptualmente fundamentado acerca de por qué determinadas prácticas rituales, aunque están extendidas, deberían ser erradicadas. Se introducen elementos no tratados anteriormente en torno a este asunto. En cuanto a la zona empírica, se hacen un examen general (todo tipo de estudios) y un escrutinio exhaustivo de lo publicado en época muy reciente (2009) en las seis revistas de más alto impacto.

Contribuciones de autoría

Los dos autores contribuyeron por igual en todos los aspectos del trabajo.

Financiación

Ninguna.

Conflictos de intereses

Ninguno.

Bibliografía

- Silva LC. Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1997. p. 390.
- Gigerenzer G, Krauss M, Vittoz O. The null ritual: what you always wanted to know about significance testing but were afraid to ask. En: Kaplan D, editor. The handbook of methodology for the social sciences. Thousand Oaks, CA: Sage Publications; 2004. p. 391–408.
- Silva LC. Una ceremonia estadística para identificar factores de riesgo. Salud Colectiva. 2005;1:309–22.
- Gore S, Jones IG, Rytter EC. Misuses of statistical methods: critical assessments of articles in BMJ from January to March, 1976. BMJ. 1977;1:85.
- Silva LC. Los laberintos de la investigación biomédica. En defensa de la racionalidad para la ciencia del siglo XXI. Madrid: Díaz de Santos; 2009. p. 492.
- Matthews RA. Facts versus factions: the use and abuse of subjectivity in scientific research. European Science and Environment Forum Working Paper; reprinted in Rethinking Risk and the Precautionary Principle. Morris J, editor. Oxford: Butterworth; 2000.
- Ioannidis JPA. Why most published research findings are false? PLoS Med. 2005;2:e124.
- Fernandes S, Hyun HK, Reeder RN, et al. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. BMC Res Notes. 2011;4:304.
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
- Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:e297.
- Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos; 2000. p. 332.
- Thomson Scientific Institute of Scientific Information. Journal Citation Reports. 2009. (Consultado el 10/12/2011.) Disponible en: http://thomsonreuters.com/products.services/science/science_products/a-z/journal_citation_reports
- Silva LC, Fariñas H. (2006) EMUCO, programa para la estimación de muestras complejas. (Consultado el 10/12/2011.) Disponible en: <http://lcsilva.sbhac.net>
- Goodman SN, Berlin JA. The use of predictive confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med. 1994;121:200–6.
- Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet. 2005;365:1348–53.
- Bacchetti P. Current simple size conventions: flaws, harms and alternatives. BMC Med. 2010;8:17.
- Bacchetti P, Deeks SG, McCune JM. Breaking free of sample size dogma to perform innovative translational research. Sci Transl Med. 2011;3:24.
- Silva LC. Una pincelada estadística con repercusiones extrametedológicas. Salud Colectiva. 2011;7:399–400.
- Bacchetti P. Peer review of statistics in medical research: the other problem. BMJ. 2002;324:1271–3.
- Norman GR, Streiner DL. PDQ statistics. Hamilton, Ontario: B.C. Decker; 2003. p. 208.
- Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter? PLoS Med. 2008;5:e4.
- Senn S. Statistical issues in drug development. Chichester: John Wiley and Sons; 1997. p. 524.
- Martin J. The tyranny of power: is there a better way to calculate sample size? BMJ. 2009;339:b3985.
- Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector. J Gen Intern Med. 1998;13:753–6.
- Malakoff D. Bayer offers a 'new' way to make sense of numbers. Science. 1999;286:1460–4.
- Savage IR. Nonparametric statistics. J Am Statist Assoc. 1957;52:332–3.
- Vickers AJ. Underpowering in randomised trials reporting a sample size calculation. J Clin Epidemiol. 2003;56:717–20.
- International Committee of Medical Journal Editors Uniform requirements for manuscripts submitted to biomedical journals. Updated April 2010. (Consultado el 6/2/2012.) Disponible en: http://www.icmje.org/urm_main.html
- Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010;340:c723.
- Mills EJ, Chan A, Wu P, et al. Design, analysis, and presentation of crossover trials. Trials. 2009;10:27.
- Charles P, Giraudeau B, Dechartres A, et al. Reporting of simple size calculations in randomised controlled trials: review. BMJ. 2009;338:b1732.
- Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet. 2005;365:1159–62.
- Chan AW, Hrobjartsson A, Jorgensen KJ, et al. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008;337:a2299.
- Barnett ML, Mathisen A. Tyranny of the p-value: the deficit between statistical significance and common sense. Guest editorial. J Dent Res. 1997;76:534–6.
- Silva LC, Suárez P, Fernández A. The null hypothesis significance test in health sciences research (1995–2006): statistical analysis and interpretation. BMC Med Res Methodol. 2010;10:44.