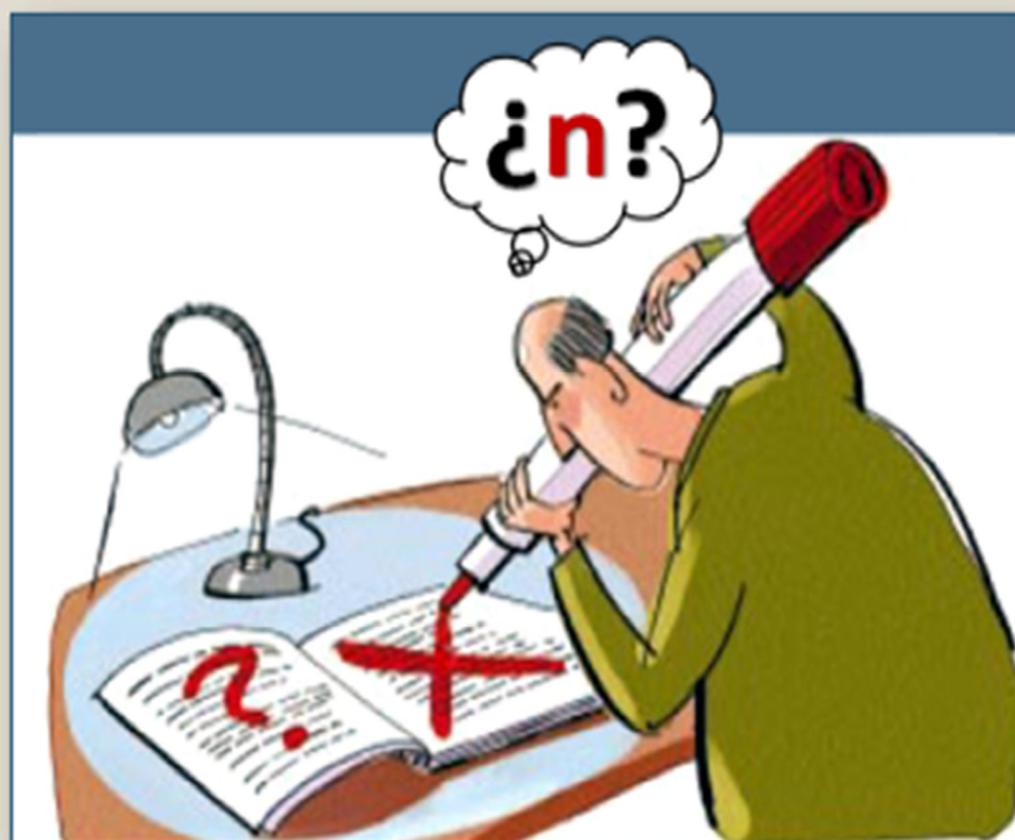


SOBRE LA INFORMACIÓN ACERCA DEL TAMAÑO MUESTRAL EN LAS PUBLICACIONES BIOMÉDICAS



INSTITUTO NACIONAL DE HIGIENE, EPIDEMIOLOGÍA Y MICROBIOLOGÍA

**SOBRE LA INFORMACIÓN ACERCA DEL TAMAÑO
MUESTRAL EN LAS PUBLICACIONES BIOMÉDICAS.**

Estudio para aspirar al título de Especialista de Primer Grado en Bioestadística.

Autora: MCs. Patricia Alonso Galbán

Tutor: DrCs. Luis Carlos Silva Ayçaguer

Ciudad de La Habana, 2010.

A mi hijo Félix Daniel, por su inigualable alegría...

«Cuando se llega al extremo en que los procedimientos estadísticos pasan a suplir nuestros pensamientos en lugar de estar en función de ellos, y somos conducidos así al reino del absurdo, entonces es el momento de regresar al sentido común».

Bakan (1966)

AGRADECIMIENTOS

A mi familia, por el gran apoyo, la paciencia y el acicate permanente para realizar este estudio, paso difícil pero necesario en el camino a la obtención del Título Especialista de Primer Grado en Bioestadística.

A mi tutor, por sus extraordinarias enseñanzas y la confianza depositada en mí para desarrollar esta investigación.

A mis profesores de la Residencia de Bioestadística, que han contribuido con mucho en mi formación académica y como investigador.

A Elías Rovira, Ileana Armenteros y Dicti, por su especial apoyo en la búsqueda y recuperación de la información necesaria para llevar a cabo el presente estudio.

PALABRAS CLAVE

- Tamaño de muestra.
- Guía de publicación.
- Ensayo clínico controlado.
- Estadística.
- Artículo científico.
- Factor de impacto.

RESUMEN

Paralelo al desarrollo de la estadística aplicada a la investigación médica, han ido apareciendo perniciosas desviaciones ajenas a la disciplina propiamente dicha. Una de ellas es la que concierne al uso ritual de fórmulas para el cálculo del tamaño muestral. Y esta conducta se extiende en virtud de las exigencias sostenidas por la mayoría de las revistas biomédicas que plantean que la comunicación detallada de la forma en que fue determinado dicho tamaño debe aparecer en los artículos. Se manejó la conjetura de que tales exigencias carecen de racionalidad y de fundamento real, así como de que son notoriamente desdeñadas e incumplidas por los autores, revisores y editores en los artículos científicos publicados actualmente. Para evaluarla se concibió una investigación predominantemente teórico-metodológica. Con el objetivo de examinar desde una perspectiva teórico-lógica las demandas prevalecientes acerca de lo que debe comunicarse sobre la determinación del tamaño muestral en estudios publicados se realizó un ejercicio de deconstrucción teórica sobre las razones esgrimidas para respaldarlas, así como el análisis de su empleo ceremonial. Para caracterizar la forma en que se comunica la determinación del tamaño muestral en la publicación sanitaria contemporánea de mayor prestigio se realizó un estudio bibliométrico de los artículos originales publicados en el año 2009 en las seis revistas de más alto impacto. La conclusión que surgió del ejercicio teórico realizado fue que las razones planteadas para fundamentar las demandas que criticamos son escasas y endebles, a la vez que existen muchas para no respaldarlas. El profundo examen de las publicaciones actuales convalidó nítidamente que, en la práctica, los propios sujetos sociales llamados a hacerlas cumplir prescinden ostensiblemente de su aplicación. El estudio permite concluir que la demanda de explicar en un artículo publicado cómo se llegó a cierto tamaño muestral carece de fundamento y utilidad. Su vigencia no promueve, sino que más bien menoscaba la transparencia del reporte de las investigaciones. Como consecuencia natural de su esterilidad intrínseca, dicha regla es muy frecuentemente incumplida.

ÍNDICE

Introducción.	1
Objetivos.	12
Metodología.	13
Desarrollo.	24
1. Procedimiento estadístico-metodológico de determinación del tamaño muestral.	24
1.1 Bases teóricas generales.	24
1.2 Consideraciones sobre el empleo ceremonial del procedimiento.	26
2. Análisis de las demandas vigentes acerca de lo que debe comunicarse sobre la determinación del tamaño de muestra en estudios publicados.	37
2.1 Resultados del estudio bibliométrico.	43
Discusión.	54
Conclusiones y recomendaciones.	65
Referencias bibliográficas.	66
Anexos.	77

INTRODUCCIÓN

La investigación científica, como vía utilizada por la ciencia para enriquecer los conocimientos disponibles, constituye hoy día un proceso de vital importancia para la sociedad. En el desarrollo del quehacer investigativo la Estadísticaⁱ-como disciplina- se ha agenciado un papel particularmente importante. Ella contribuye a establecer criterios que valoren la verdad de una teoría científica, dado que aporta medios objetivos que ayudan a identificar los resultados útiles al conocimiento y separarlos de los que no lo son. ^{1,2}

Esta disciplina, aplicada al campo de la salud, ha dado en llamarse Bioestadística y en él ha desempeñado un papel destacado como medio para la inferencia científica y las generalizaciones en el marco médico. ³ Muchas acciones clínicas y de salud pública se basan, aunque sea tras un complejo y largo proceso teórico-práctico, en los métodos empleados para adquirir el conocimiento que permite diseñarlas.

La Estadística se ha consolidado, principalmente a lo largo de los últimos 50 años, en estrecho vínculo con la investigación clínica y epidemiológica, campos en los que ha demostrado ser una herramienta de inestimable valor.¹ Décadas atrás, prácticamente no era utilizada en las investigaciones biomédicas y la mayoría de los investigadores de este campo carecían de entrenamiento profesional y/o de la

ⁱ Estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas. (<http://buscon.rae.es/drael/>)

asesoría adecuada para la aplicación correcta de procedimientos metodológicos, principalmente los de índole estadística.

En la actualidad muchos investigadores sienten que sus trabajos no tienen suficiente rigor científico e incluso temen que su publicación se rechace si no están avalados por al menos un proceder estadístico.^{2,4,5} Ello *“puede dar lugar a que un medio se convierta sin más en una finalidad,”*² realidad esta que ha sido criticada en el sentido de ser un vicio en extremo peligroso, incluso para el propio prestigio de una disciplina como esta que, razonablemente empleada, puede ocupar un espacio relevante como recurso instrumental para operar en el mundo de la investigación médico-social.

En el proceso de transición *“de la marginalidad a la cima”*⁵ que ha experimentado, la Estadística ha encontrado más de un escollo, principalmente en relación con su correcta aplicación. Como tecnología al fin, no es exclusivamente “culpable” de nada aunque en la práctica investigativa algunos la empleen irreflexivamente, de forma inercial o al margen de un razonamiento profundo. Quizás el ejemplo más elocuente de esta realidad sea la inserción indiscriminada de técnicas y procedimientos estadísticos dentro de algoritmos metodológicos a manera de ritual o ceremonia más en virtud de una tradición o de una expectativa externa, que como un recurso adecuado o necesario para solucionar el problema de investigación encarado y con la convicción inocente de que proveerán mecánicamente la respuesta adecuada a las preguntas formuladas.^{1,3,5,6}

Existen no pocos procedimientos estadístico-metodológicos cuyo empleo ritual es, infortunadamente, frecuente en la investigación biomédica.¹ No procede aquí inventarlos exhaustivamente, pero a modo de ilustración corresponde recordar algunos de ellos:

- La implementación de estrictos y complejos diseños muestrales de tipo probabilístico sin la posterior determinación de los errores de muestreo; es decir, sin el cálculo de los intervalos de confianza.^{4,6-10}
- El uso de la regresión logística y la regresión paso a paso para identificar factores de riesgo de enfermedades.^{1,6,11}
- La corroboración de homogeneidad a través de pruebas de significación entre los grupos de tratamiento que fueron determinados mediante asignación aleatoria de un ensayo clínico controlado.^{4, 6- 9, 12-17}
- El empleo de pruebas de hipótesis para corroborar bondad de ajuste.⁶
- La evaluación de la fiabilidad de un instrumento de medición mediante coeficientes de correlación entre resultados obtenidos en dos momentos de aplicación.¹
- El empleo adocenado de fichas técnicas en la comunicación de los resultados de encuestas.^{4, 6}
- El uso ritual de fórmulas para el cálculo del tamaño de muestra mínimo requerido.^{4, 6, 13, 18-21}

La aplicación repetida de recursos estadísticos a toda costa, constituye un serio peligro en la práctica de investigación, no solo por el marcado estancamiento

intelectual que provoca en aquellos que se acomodan a usarlos, sino porque muchas veces son estériles y en gran medida contraproducentes.^{12, 13, 18, 22-28}

Suponen de esta manera un gasto innecesario de recursos y esfuerzos, y en el peor y no menos corriente de los casos, a publicaciones destinadas a engrosar el caudal de conocimientos erróneos sobre una materia.¹

Procede aclarar que se ha estado haciendo referencia a un esquematismo de tipo conceptual que no debe ser confundido con el de índole formal, muchas veces útil y hasta medular. Este último es aquel que define determinado conjunto de pasos o requisitos en aras de estandarizar y viabilizar ciertas acciones sin comprometer su calidad.⁴ Tal es el caso, por poner un ejemplo, de la asignación aleatoria de los pacientes en los ensayos clínicos controlados, la cual ocupa un lugar metodológicamente prominente y esencial en su diseño. Definir los grupos de tratamiento, escoger las unidades de análisis y asignarlas al azar a cada grupo es crucial para eliminar los sesgos determinados por la subjetividad del investigador, lo que permite considerar que las muestras sean independientes y susceptibles de ser comparadas fructuosamente. Aquí se sigue formalmente un orden secuencial y lógico que garantiza el cumplimiento de estándares de calidad ya consagrados.

Otro importante problema que se ha extendido en la aplicación de la Estadística radica en creer que su uso constituye un escudo que nos protege de la subjetividad en el proceso de investigación científica. Aunque la evaluación racional y objetiva de la realidad está sin duda en la base del progreso científico, la

práctica dista de ser tan idílica: la objetividad es un deseo natural y legítimo, pero resulta simplemente inalcanzable en estado puro. Irremediablemente, la subjetividad (no equivalente a la arbitrariedad) está presente en toda actividad humana y la producción científica no es una excepción.⁶

En este sentido, la principal contribución de la Estadística radica en deslindar lo anecdótico de las regularidades, lo cual no significa que los propios instrumentos estadísticos no estén sujetos a la subjetividad; consecuentemente, mal podrán garantizar automáticamente la expectativa que muchos investigadores depositan en su capacidad objetivadora.^{5,6}

Sin embargo, la gravedad del asunto no reside sólo en este hecho, sino en que, al considerar que los procedimientos estadísticos convencionales confieren total objetividad a nuestros análisis-y ocultar la participación de la subjetividad- puede consolidarse la quimera de que constituyen un árbitro final, y enraizarse con ella una peligrosa pereza intelectual. Es un hecho que los investigadores pueden autoengañarse, y desorientar a la comunidad científica disfrazando la subjetividad de objetividad estadística.^{5,6} Como señala Robert Matthewsⁱⁱ, *“los esfuerzos previos para «barrer la subjetividad debajo de la alfombra», han conducido a adoptar procedimientos valorativos que no son ni objetivos ni confiables”*.²⁶ La alternativa más honrada es entonces aceptar que la subjetividad es ineludible y, sin dejar de minimizarla, encaminar nuestros esfuerzos a conseguir que su

ⁱⁱ Físico de la Universidad de Oxford, miembro de la Royal Statistical Society y destacado analista de The Sunday Telegraph de Londres.

participación se refleje explícitamente en el proceso investigativo, tal y como propone esperanzadoramente el enfoque Bayesiano.^{2,6}

No debemos olvidar -como ocurre con cierta frecuencia en el ámbito científico contemporáneo- que el análisis estadístico es solo uno de los componentes de los estudios clínicos y de salud pública y como tal debe considerarse en el proceso investigativo.⁶ Los resultados que se obtienen por su conducto son un elemento más que debe integrarse al marco teórico de conocimientos que existe sobre un tema en cuestión y a la información proveniente de estudios anteriores.

En los últimos tiempos se ha acumulado en la bibliografía más reflexiva importante evidencia sobre estas y otras realidades que lamentablemente dañan la reputación de la estadística en particular y de la ciencia en general.^{1-4, 6, 7, 12, 13, 14, 18, 19, 11, 15, 20,}

²⁹⁻⁴⁶ Se han planteado las condicionantes históricas que han favorecido la aparición de tales tendencias, entre las que se distinguen:¹²

- La falta de cultura investigativa de muchos profesionales.
- La enseñanza masiva, generalizada y despersonalizada de la estadística y la metodología de la investigación (cursos a distancia, teleconferencias, etc.):
- Las conductas inerciales y sumisas a las tradiciones.
- La frecuente exigencia de su uso por parte de los editores.
- La gran popularidad de que gozan.

- El enmascarado trasfondo comercial y publicitario que no pocas veces alienta el empleo de los programas computacionales para el análisis estadístico que facilitan la aplicación acrítica de las técnicas.

También se han desarrollado algunos interesantes ejercicios de deconstrucciónⁱⁱⁱ metodológica ocasionalmente complementados con propuestas concretas que trazan las pautas para la reconstrucción sobre nuevas bases,^{1, 6, 47} y se han propuesto alternativas para encarar estas situaciones.⁴⁸⁻⁵²

Sin embargo, tanto considerar a la Estadística como garante de objetividad como emplearla de forma inercial y ceremonial, son problemas que distan de estar resueltos y que se reflejan con alta frecuencia en los reportes de las investigaciones biomédicas, clínicas o sanitarias.

En consonancia con la denuncia de estas deformaciones, no pocas revistas de avanzada en la investigación sanitaria han realizado desde hace ya algunos años modificaciones en sus políticas editoriales, que pudieran contribuir a revertir estas realidades, lo cual probablemente no ocurra con la velocidad necesaria y deseada.⁵³⁻⁵⁵ Quizás el ejemplo más elocuente de esto sea que desde la versión del año 2000 de las normas que periódicamente publica el *Comité Internacional de Directores de Revistas Médicas* (Grupo de Vancouver), se consigna textualmente en el apartado de requisitos técnicos dedicados al empleo de la estadística “Se evitará la dependencia exclusiva de las pruebas

ⁱⁱⁱ Desmontaje de un concepto o de una construcción intelectual por medio de su análisis, mostrando así contradicciones y ambigüedades. (<http://buscon.rae.es/drael/>)

estadísticas de verificación de hipótesis, tal como el uso de los valores p, que no aportan ninguna información cuantitativa importante”...⁵⁶

La preocupación por la calidad estadística de los artículos publicados es una aspiración lógica y sensata. El artículo científico es, y aparentemente seguirá siendo el mecanismo por excelencia para la transmisión de nuevos conocimientos, aún cuando las formas de comunicación científica son hoy día diversas y se han abierto nuevas alternativas con la expansión de las telecomunicaciones y la informática. Entre los componentes del proceso de su elaboración, la Estadística juega con frecuencia un papel relevante, no solo por la participación que pudo haber tenido en el análisis de los resultados que van a comunicarse, sino porque contribuye a que los mensajes ganen en transparencia y poder persuasivo.⁴

Las deficiencias en el empleo de los recursos estadísticos en los artículos científicos publicados es un hecho estudiado desde muchos años atrás.⁵⁷⁻⁶⁰ El proceso de “peer-review”^{iv} es uno de los elementos que puede contribuir a minimizarlas, siempre que se aplique de forma correcta, desde el punto de vista técnico y ético. Incluso algunas revistas han adoptado la medida de fortalecer el equipo de revisores con personal experimentado en la materia.⁴

Adicionalmente, los comités editoriales de las Revistas Biomédicas Internacionales han desarrollado y puesto en práctica guías de publicación para los autores que contienen listas de recomendaciones^v acerca de cómo exponer los datos y procedimientos estadísticos, así como los restantes aspectos que deben ser

^{iv} En Español “revisión por pares”

^v Guías de publicación, en inglés: Reporting Guidelines.

contemplados en los artículos. Tanto los revisores como los editores hacen uso de ellas en el proceso de evaluación de la calidad de los artículos.

Las guías se han ido modificando con su revisión sistemática y su número se ha incrementado al hacerse más específicas en lo relativo al tipo de investigación.⁶¹

En la actualidad las de uso más frecuente son:

- Consolidated Standards of Reporting Trials (CONSORT).⁶²⁻⁶⁵
- STrengthening the Reporting of Observational Studies in Epidemiology (STROBE).⁶⁶
- Standards for Reporting of Diagnostic Accuracy (STARD).⁶⁷
- Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).^{68,69}
- STrengthening the REporting of Genetic Association Studies (STREGA).⁷⁰
- Meta-analysis of observational studies in Epidemiology (MOOSE).⁷¹

En este contexto, *un requisito incluido en la mayor parte de las guías para la publicación a las que se adhiere el grueso de las revistas- inclusive las de más alto factor de impacto- es el de explicar y justificar cómo fue establecido el tamaño de muestra, así como exponer de forma detallada las consideraciones para su determinación.*^{13, 19, 29, 62-66, 70}

Es natural que un tema ubicuo como el del tamaño muestral sea manejado en este marco, dado que continúa siendo un problema algo enigmático y confuso, carente de una solución inequívoca y aceptable sin objeciones y que tiene que ser encarado siempre que se trate con datos empíricos. Sin embargo, tanto la

pertinencia como la racionalidad del requisito antes mencionado ofrecen dudas. En torno a la demanda de explicar en un artículo publicado cómo se llegó a cierto tamaño muestral, se alzan algunas interrogantes tales como:

¿Está fundamentada o constituye otro procedimiento metodológico ritual?

¿Cuál es -si la tiene- su utilidad?

¿Genera más problemas de los que resuelve en lo referente al reporte transparente de la investigación?

¿En qué medida se cumple en la investigación sanitaria contemporánea?

Al no contarse actualmente con un examen acucioso y serio de esta problemática resultaría sumamente beneficioso desarrollar un estudio en que se analicen los elementos epistemológicos, prácticos y éticos que la enmarcan. Dicho ejercicio es consecuente, además, con la implementación de las recomendaciones emanadas de un trabajo anterior en el que se debaten sistematizadamente e ilustran aspectos y expresiones de las ceremonias estadístico-metodológicas en la investigación sanitaria contemporánea.¹

Adicionalmente, tiene evidente interés el examen de los artículos que se publican en la actualidad, con vistas a caracterizar la manera en que se comunica la determinación del tamaño muestral y aquilatar la medida en que se cumplen las exigencias establecidas por los comités editoriales de las revistas. Algunos esfuerzos se han realizado en este sentido, pero han estado concentrados en el área de los ensayos clínicos controlados.⁷²⁻¹⁰⁸

Dicho estudio permitiría contar con más elementos para identificar todo lo que falta en el proceso de erradicar el extendido uso ceremonial que se hace de la Estadística en la investigación sanitaria y contribuir a advertir acerca del negativo efecto que cotidianamente se produce en detrimento de la buena práctica investigativa. También tendría un valor orientativo para autores, revisores y editores y permitiría proponer recomendaciones racionales, útiles, factibles y lógicas, sobre lo que debe comunicarse en la publicación científica en el campo de la salud.

A tenor de lo anterior, se propone realizar un ejercicio de desconstrucción metodológica sobre el tema del reporte de la determinación del tamaño muestral en los artículos publicados en las revistas médicas, así como examinar los patrones prevalecientes en la investigación sanitaria contemporánea. En esa dirección se orientan los siguientes objetivos:

OBJETIVOS

1. Examinar desde una perspectiva teórico-lógica las demandas prevaletentes acerca de lo que debe comunicarse sobre la determinación del tamaño de muestra en estudios publicados.
2. Caracterizar la forma en que se comunica la determinación del tamaño muestral en la publicación sanitaria contemporánea de mayor prestigio.

METODOLOGÍA

A partir de las conjeturas de que las demandas relativas a lo que debe comunicarse sobre la determinación del tamaño de muestra en la publicación sanitaria carecen de fundamento y racionalidad y que son incumplidas con cierta frecuencia en la actualidad, se concibió una investigación de índole predominantemente teórico-metodológica. El estudio fue vertebrado en torno a un profundo examen bibliográfico del tema y a un estudio bibliométrico de naturaleza descriptiva.

El trabajo se dividió en dos etapas:

En la primera de ellas se realizó una revisión de la bibliografía sobre el tema de la determinación del tamaño muestral necesario en la investigación sanitaria y su reporte en la publicación científica. Se compilaron aspectos teóricos que permitieran valorar la racionalidad, pertinencia y fundamentación de las demandas actuales sobre lo que debe comunicarse en relación a dicho cómputo, así como examinar su carácter potencialmente ceremonial.

La recopilación del material de referencia estuvo centrada en libros, artículos originales y otras publicaciones de carácter científico que trataran elementos teóricos de metodología para la investigación, estadística aplicada a las ciencias de la salud y sobre la elaboración y revisión de artículos científicos en este ámbito. También se revisaron textos que incursionaran en análisis críticos sobre las

consecuencias nocivas de utilizar de forma inadecuada determinados procedimientos estadísticos en el marco de la investigación sanitaria.

Sobre estas bases, se realizó una reflexión sistematizada que abarcó, introductoriamente, los elementos teóricos generales del procedimiento estadístico-metodológico para determinar el tamaño de muestra y algunas consideraciones sobre su empleo ritual. A continuación se desarrolló un ejercicio de deconstrucción teórico-metodológica centrado en las demandas vigentes acerca de lo que debe comunicarse sobre la determinación del tamaño de muestra en estudios publicados, que incluyó una valoración de las razones esgrimidas hasta el momento para su respaldo, así como el análisis del empleo ceremonial de este reporte y su improcedencia.

Complementariamente -en la segunda fase- se examinaron los patrones de la forma en que se comunica la determinación del tamaño muestral en la investigación sanitaria contemporánea de alto impacto. Con este fin se realizó un estudio bibliométrico de los artículos originales publicados en el periodo enero-diciembre del 2009 en las seis revistas de más alto factor de impacto en el campo de la salud al cierre del propio año, de acuerdo con lo informado por el Journal Citation Reports.¹⁰⁹

Estas fueron:

- The New England Journal of Medicine (New Engl J Med),
- The Lancet (Lancet),
- The Journal of de American Medical Association (JAMA),

- Annals of Internal Medicine (Ann Intern Med),
- British Medical Journal (BMJ) y
- PloSMedicine (PloS Med).

La elección obedeció a que estas revistas representan lo que hoy se admite como el mejor reflejo de la “ciencia exitosa”^{vi} en las grandes áreas de la investigación médica: la Clínica, la Epidemiología y la Salud Pública. Se seleccionó el año 2009 por constituir el más reciente con respecto al inicio de la presente investigación.

En cada una de las revistas los artículos originales fueron localizados en las secciones destinadas a su publicación, como se describe a continuación:

Revista	Secciones
New Engl J Med	Original Research
Lancet	Articles
JAMA	Original contributions Caring for the critically ill patient Preliminary communications
Ann Intern Med	Articles Academia and clinic Improving patient care
BMJ	Research
PloS Med	Research articles

En correspondencia con nuestros objetivos, se decidió incluir solo aquellos artículos que reportasen investigaciones para las cuales existen estándares

^{vi} El concepto de “ciencia exitosa alude a que no se puede considerar como tal aquella producción que no se lee o se lee de manera muy marginal.

inequívocos sobre la determinación del tamaño muestral, a los cuales se refieren esencialmente las guías de publicación y que adicionalmente constituyen la mayoría de los estudios publicados. Los trabajos incluidos fueron entonces:

- Estudios Epidemiológicos Observacionales (de cohorte, de casos y controles y transversales o de prevalencia)
- Estudios Experimentales (ensayos clínicos controlados aleatorizados de fases avanzadas (III ó IV)-exceptuando los diseños adaptativos-, estudios preclínicos, investigaciones básicas, estudios piloto, estudios de eficacia e intervenciones comunitarias)

Este eje de clasificación está conectado con la manera de encarar el problema del tamaño muestral, por lo que consideramos oportuno realizar el examen previsto de manera diferenciada entre ambos tipos de investigaciones. Los estudios experimentales -y más específicamente los ensayos clínicos - son necesariamente de aliento pragmático, lo cual los distingue claramente, en principio, de los observacionales, y su metodología es por lo general más sólida. Adicionalmente, van dirigidos casi siempre a dirimir un problema inferencial específico, relativo a un único parámetro (sobre la base del cual se suele realizar el cálculo del tamaño muestral). En cambio, en los observacionales, se desea por lo general estimar varios parámetros al mismo tiempo (porcentajes, odds ratios, medias, etc), lo cual hace de la elección del tamaño muestral un acto más confuso, ya que no habría un objetivo único en que basar el proceso de cálculo.

Aunque al definir los criterios de inclusión ya queda de hecho claro cuáles trabajos fueron excluidos, no es ocioso esclarecer que, de acuerdo a las restricciones arriba mencionadas, no fueron incluidos en este estudio los trabajos susceptibles de ser clasificados como: reporte de un caso, revisión narrativa, revisión sistemática y/o meta-análisis, estudio de casos, análisis de costo-efectividad, investigación cualitativa, cuasi-experimental ó metodológica, estudio ecológico ó de modelación matemática, análisis de la eficacia de medios diagnósticos, ensayo clínicos fase I, II, I-II ó con diseño adaptativo y análisis conjunto de varios ensayos clínicos. Se excluyeron también aquellas investigaciones básicas que no implicasen el estudio de unidades de análisis individuales (como por ejemplo, cuando el objeto de estudio son códigos genéticos, tejidos, características biológicas de un agente infeccioso, procesos celulares, entre otros.)

Para todos los artículos seleccionados se revisó el reporte completo de la investigación así como cualquier material extra disponible ó trabajo citado que contuviera aspectos del diseño del estudio. Con la información recopilada se confeccionó una base de datos que incluyó los siguientes aspectos de cada uno de los artículos:

- Revista en la que fue publicado.
- Referencia bibliográfica.
- Título.
- Clasificación de acuerdo al eje seleccionado.
- Información sobre la determinación del tamaño de muestra.
- Valor del tamaño muestral anunciado.

- Guía de publicación conforme a la cual debió elaborarse y revisarse según la revista en que fue publicado y el tipo de estudio de que se trate.
- Valor del tamaño final del estudio.

Las variables empleadas para el análisis de los artículos incluidos se operacionalizaron como se explica a continuación:

Variable	Clasificación	Operacionalización	
		Escala	Descripción
Revista	Politómica	New Engl J Med Lancet JAMA Ann Intern Med BMJ PloSMed	Nombre de la Revista en que fue publicado el artículo.
Clasificación del estudio	Dicotómica	Observacional Experimental	Se clasificó cada trabajo de acuerdo con el siguiente criterio: Observacional: no se manipula el factor en estudio. Experimental: manipulación del factor en estudio con aleatorización.
Tamaño muestral justificado	Dicotómica	Sí justificado No justificado	“Sí justificado” en caso de que en el artículo se comunique y justifique cómo fue determinado el tamaño de muestra.

			(es decir, que en el artículo aparezcan las razones que condujeron a que el tamaño muestral fuese el resultante -y no otro-). "No justificado" en caso contrario.
Método empleado para la determinación	Dicotómica	Fórmula Otro	En caso de haberse comunicado y justificado cómo se determinó el tamaño muestral, se examinó la naturaleza de la forma en que fue determinado. Se consideró "Fórmula" si se comunicaba que el tamaño muestral había sido calculado mediante algún algoritmo, explícita o implícitamente mencionado. Para los restantes casos se consignó la categoría "Otro"
Comunica datos necesarios para el cómputo	Dicotómica	Sí No	En caso de que el método empleado para la determinación del tamaño muestral anunciado fuese "Fórmula" se examinó si se comunicaban cada uno de

			<p>los datos necesarios para reproducir dicho cómputo. (Véase Anexo 1)</p> <p>Para cada uno de los datos se consignó “Sí” en caso afirmativo y “No” en caso contrario.</p>
<p>Reporte de <u>todos</u> los datos necesarios para el cómputo</p>	Dicotómica	<p>Sí</p> <p>No</p>	<p>En caso de que el método empleado para la determinación del tamaño muestral anunciado fuese “Fórmula” se examinó si se comunicaban <u>todos</u> los datos necesarios para el cómputo según el tipo de estudio. (Véase Anexo 1)</p> <p>Se consignó “Sí” en caso afirmativo y “No” en caso contrario.</p>
<p>Consideración de la tasa de no respuesta</p>	Dicotómica	<p>Sí</p> <p>No</p>	<p>En caso de que se comunicase el método empleado para el cómputo del tamaño muestral anunciado y este fuese fórmula se revisó si la tasa de no respuesta fue considerada en el cálculo.</p>
<p>Comunicación del valor del tamaño muestral anunciado.</p>	Dicotómica	<p>Sí</p> <p>No</p>	<p>“Sí” en caso de que se comunicase el <u>valor</u> determinado a priori del tamaño de muestra.</p>

			"No" en caso contrario
Diferencia relativa entre tamaño muestral anunciado y el tamaño final del estudio	Cuantitativa continua	Valor	En caso de que se comunicase el valor determinado a priori del tamaño de muestra se calculó la diferencia relativa entre este y el correspondiente al tamaño de la final del estudio.
Publicado en revista que demanda justificar el tamaño muestral.	Dicotómica	Sí No	Para cada trabajo y según su clasificación se verificó si la revista en que fue publicado se adhiere a una guía de publicación que incluye la demanda de comunicar y justificar cómo fue determinado el tamaño muestral. Para esto se revisó la sección de "Instrucciones a los autores" de cada revista. (Véase anexo 2) Se consignó "Sí" en caso afirmativo y "No" en caso contrario.
Cumplimiento de las demandas.	Dicotómica	Cumple No cumple	Si el artículo fue publicado en una revista que, de acuerdo con el tipo de estudio que este reporta, demanda justificar el tamaño muestral se

			analizó si cumplía o no con este requisito. Se consideró como “cumple” en caso de que se correspondiera con lo planteado en la guía y “no cumple” en caso contrario.
--	--	--	---

Procesamiento de los datos

- Codificación de los datos. Las hojas de recogida de datos se diseñaron teniendo en cuenta la necesidad de posterior vaciado de la información. Cada uno de los posibles resultados de las variables se codificó para tal fin.
- Revisión de los datos transcritos. Tras ser registrados en la base de datos, se efectuó una revisión de toda la información contenida con respecto a los artículos analizados.
- Depuración de datos. Se efectuó una depuración de los datos introducidos, estudiando una a una todas las variables. Se analizaron los posibles valores ilógicos o imposibles.
- Soporte informático. El registro y procesamiento de la información se realizó mediante el programa Microsoft Office Excel 2007 para Windows XP.

Análisis de los datos.

Para caracterizar la forma en que se comunica la determinación del tamaño muestral se realizaron estimaciones de la proporción de artículos clasificados en

cada una de las categorías de las variables cualitativas, así como de la media de la diferencia relativa entre el tamaño muestral y el tamaño final del estudio.

Las estimaciones se obtuvieron según la distribución de los artículos por tipo de estudio y revista, así como para el total de los mismos. Se computó un intervalo de confianza al 95% para las estimaciones fundamentales.^{vii} Para el análisis estadístico se utilizaron los programas EMUCO (Estimación en muestras complejas) versión 2.1 y SPSS.inc versión 17.0.

^{vii} Los datos disponibles para el cálculo del intervalo de confianza para las restantes estimaciones están disponibles en la sección de resultados. Debe tenerse en cuenta que en algunos casos debido a la magnitud reducida de la muestra, estos pueden ser muy amplios.

DESARROLLO

1. Procedimiento estadístico-metodológico de determinación del tamaño muestral.

1.1) Bases teóricas generales.

En la mayoría de las investigaciones que se desarrollan en el campo de las ciencias biomédicas y de la salud resulta conveniente, generalmente por motivos económicos, recurrir al estudio de la población a través de una muestra.^{4, 18, 110}

Con frecuencia el muestreo se asume incluso como única alternativa en situaciones donde es imposible acceder a cada uno de los integrantes del universo de estudio por ser éste en teoría infinito, o también cuando se realiza una intervención que implica riesgo para los participantes, situación en la que se impone involucrar, por razones éticas, la menor cantidad de personas posible.¹

Cualquiera sea la naturaleza de la muestra, es natural que se desee operar con la menor cantidad posible de unidades de análisis con el fin de economizar recursos. Lógicamente, mientras más pequeño sea el número de efectivos estudiados, la inversión de tiempo, recursos financieros, materiales y humanos será menor. El problema consiste en hallar ese número mínimo.⁶ Se trata de una tarea muy singular y contradictoria porque en ella confluyen dos circunstancias especiales:

- es un asunto al que hay que atender inevitablemente: por una u otra vía, no hay más remedio que decidir un tamaño muestral;

- no existe una solución formal que sea enteramente satisfactoria.

El hecho de que todo proceso de muestreo lleva implícito futuros errores inferenciales (precisamente porque la información disponible se obtiene a partir de datos que son solo una fracción de todos aquellos que reflejan exactamente la realidad que se observa) determina que optar por un número pequeño de unidades de análisis produzca imprecisiones en los resultados de la investigación, conocidas como errores de muestreo. En ocasiones, tales errores son de tal magnitud que pueden cancelar su valor a los efectos de aportar conocimientos provechosos.¹¹⁰⁻¹¹²

De lo anterior se deriva un razonamiento tan importante como intuitivo: el tamaño de muestra también gravita (inversamente) de manera determinante sobre la magnitud del error muestral. En síntesis, es necesario seleccionar la menor cantidad posible de unidades de análisis, que suponga errores inferenciales aceptables dentro de una exigencia económica asumible.

En auxilio de tal demanda se han desarrollado numerosas fórmulas que teóricamente permiten calcular el tamaño mínimo necesario. Cada una de estas expresiones matemáticas concierne específicamente a algún parámetro poblacional y a la acción básica que se vaya a realizar en relación con éste: bien estimarlo, bien valorar mediante pruebas de significación si su magnitud difiere de otro.^{1, 4, 18, 110,112}

Usualmente, las estimaciones se realizan en el marco de los llamados estudios descriptivos donde el objetivo consiste en caracterizar determinada población a

través de diversos parámetros. Generalmente se trata de medidas de resumen (proporciones, porcentajes, razones, índices, medias y medianas), de posición (deciles, cuartiles y, en general, percentiles) y totales absolutos, aunque también con propósito descriptivo pueden realizarse estimaciones sobre varianzas y coeficientes de correlación y regresión. La fórmula específica para calcular el tamaño muestral para la mayoría de estos estadígrafos suele ser bastante compleja. Por lo general son poco tratadas en la literatura no especializada y, menos aun, aplicadas en la práctica. Dada su sencillez, aquella a la que más se acude es la relacionada con la estimación de un porcentaje y, en menor grado, de una media aritmética.^{4,18}

Lo mismo sucede en el ámbito analítico, donde usualmente se aborda la problemática para calcular el tamaño muestral cuando se desea contrastar, mediante respectivas pruebas de significación, porcentajes o medias aritméticas correspondientes a sendas poblaciones.

Las fórmulas que se emplean con mayor regularidad pueden ser revisadas en el Anexo 1.

1.2) Consideraciones sobre el empleo ceremonial del procedimiento.

Curiosamente, a pesar de ser el problema del tamaño de muestra uno de los temas más borrosamente solucionados por la estadística, la uniformidad del tratamiento que le dan los textos es casi total. Ello hace pensar a muchos que la interfase entre la teoría y la práctica correspondiente está completamente consolidada y carece de fisuras. Pero esto es falso.

Las inconsistencias en esta materia son tan notables que resulta asombrosa la hipnosis colectiva a la que asistimos. Salvo alguna pincelada aquí o allá, el conflicto es predominantemente pasado por alto por textos y profesores y el grado en que se ha tratado con rigor y entereza, de manera frontal, es ridículamente reducido. Ante tal contingencia, aparecen de inmediato las “soluciones” rituales.⁶ El sensible y crucial asunto de determinar el tamaño muestral adecuado para un estudio constituye un ejemplo difícilmente superable de procedimientos estadísticos en los que el mecanicismo suple al pensamiento crítico que lo originó.

No es que las fórmulas antes planteadas no aporten supuestamente la solución, sino que ellas plantean muchos más problemas de los que resuelven por la simple razón de que, salvo situaciones excepcionales, para aplicarlas hay que fijar de antemano valores desconocidos y sólo susceptibles de ser determinados mediante una generosa dosis de imaginación, e incluso fabulación^{1, 6} (ej. la preestimación del parámetro de “mayor importancia” que se desea estimar, el nivel de confiabilidad, el error mínimo admisible, la potencia de la prueba de hipótesis, el beneficio mínimo que resulte clínicamente interesante con el nuevo tratamiento, entre otros).

Esto hace que, independientemente de los objetivos de un estudio y su diseño, la determinación del tamaño de muestra mínimo requerido contenga, entre otros aspectos generalmente escamoteados en los textos, una inevitable carga de subjetividad.^{1, 4, 18, 110} Algunas de sus fuentes son ocasionalmente reconocidas en tal calidad por la literatura, pero lo típico es que no se haga mención alguna

a ellas, se den «explicaciones» doctrinarias,⁶ o se propongan recetas mágicas ante las dificultades que ellas generan.¹¹³⁻¹²⁰

Para ilustrar la subjetividad implícita en este procedimiento tomaremos como ejemplo la determinación de cada uno de los parámetros que exige la fórmula del tamaño de muestra para estimar un porcentaje poblacional (Véase fórmula [1.1] en el Anexo 1). Cabe enfatizar, sin embargo, que las consideraciones para esta situación particular no se diferencian conceptualmente de aquellas a tener en cuenta para fórmulas tales como, [1.2], [1.3], [1.4] (Véase Anexo 1) u otras.

- ***El tamaño poblacional (N):***

El cálculo del tamaño de muestra requiere de un valor que represente el total de unidades que componen la población de la cual va a ser extraída la muestra de estudio (N).

Es un hecho que en la mayoría de los casos, es extremadamente difícil, y con frecuencia imposible, conocer exactamente cuántas unidades componen una población, a menos que ésta sea realmente tan pequeña como para que sus integrantes puedan ser fácilmente identificados. Esto determina que cuando se trabaja con poblaciones relativamente grandes su tamaño se obtiene mediante estimados que brindan con mejor o peor calidad una aproximación.

Sin embargo, el valor de N tiene poca influencia y, además, se suele conocer con bastante precisión, por lo cual no es este un elemento que aporte sensiblemente subjetividad a la determinación del tamaño muestral.

- ***El porcentaje para la variable “de mayor importancia” y su preestimación (P):***

La fórmula para el cálculo del tamaño muestral en el caso que nos ocupa exige, paradójicamente, que se anticipe el propio valor P que se desea estimar. Puesto que en cualquier estudio se maneja un nutrido grupo de porcentajes poblacionales y como el tamaño de muestra debe ser único para todo el estudio, la *solución* usualmente sugerida consiste en seleccionar aquel que se considere “de mayor relevancia” y hacer el análisis correspondiente al tamaño de muestra para este.

Esta decisión impone una clara carga de subjetividad. Basta reparar en que usualmente una tabla con propósito descriptivo exige una estimación por cada una de sus celdas y en ocasiones suelen construirse muchas tablas con numerosas casillas, donde todas contienen un parámetro “de interés”. Resulta obvio que la decisión sobre cuál ha de elegirse inevitablemente entraña una decisión de fuerte corte subjetivo.

Por otra parte, suponiendo que un porcentaje se distinga de otros, por su importancia para la investigación, su selección implica que el error máximo que se admite (E_0) sea válido solo para éste mientras que para los otros tal consideración queda en un limbo.

Pero mucho más serio es el conflicto que crea la demanda de comunicar de antemano el valor del porcentaje que se desea estimar. En principio se trata desde luego, de un problema sin solución real por obvias razones lógicas. Para

escapar de ese laberinto (hay que comunicar anticipadamente el valor cuyo desconocimiento es la razón del estudio) se sugiere usar un valor de P reportado por estudios similares, o el que dimane de un estudio piloto que aporte una cifra aproximada. La demanda resulta, además de paradójica, en cualquier caso dependiente de la subjetividad del investigador, quien habrá de determinar de dónde escoger su preestimación, pues pueden existir diferentes estudios con disímiles resultados de P . La paradoja se resume en que el valor preestimado de P , sin ser conocido a ciencia cierta, debe ser cercano al verdadero parámetro poblacional, pues si su estimación previa fuera muy errónea influirá de manera negativa en el resultado del cálculo del tamaño muestral; al mismo tiempo si se conociera verdaderamente o con bastante precisión el parámetro, entonces carecería de toda utilidad estimarlo, ya que la finalidad del proceso es justamente obtener una buena aproximación del parámetro en cuestión.

En la fórmula para el cálculo de muestra en relación a una media aritmética el proceso se agrava pues la preestimación no corresponde a la media poblacional sino a su varianza, parámetro aun más esquivo.

Una de las soluciones que ha sido sugerida con éxito (su aplicación en la práctica resulta bastante frecuente) consiste en asumir, caso de que la magnitud de P en la población sea totalmente desconocida, que ésta es igual a 50%. Con este artificio matemático se consigue que la cantidad $P(100-P)$ adopte su mayor valor produciendo el mismo efecto en el tamaño muestral, dada la relación directamente proporcional que guarda con él. ¹¹³⁻¹¹⁶

Pero el argumento es falaz, pues esto solo sería válido siempre que E_0 esté fijado de antemano. Sin embargo, el máximo error que se puede admitir al estimar P no puede establecerse razonablemente hasta que no se suponga su magnitud, y cuando ésta se desconoce totalmente no existen criterios para considerar aceptable un error en términos absolutos. Lo que sí pudiera fijarse de antemano es un error relativo máximo admisible cuyo cálculo se realiza mediante la siguiente expresión:

$$E_r = \frac{E_0}{P} \quad [2.1]$$

La fórmula correspondiente para el cálculo del tamaño muestral puesta en términos del error relativo en vez del error absoluto (sustituyendo [2.1] en [1.1]) sería:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad ; \quad n_0 = \frac{Z_{1-\alpha/2}^2 \frac{(100-P)}{P}}{E_r^2} \quad [2.2]$$

Pero al fijar el error en términos relativos, el tamaño de n_0 (y por ende, de n) crece en la medida que P disminuye, un resultado coherente con la intuición: si el rasgo cuya prevalencia se quiere estimar es muy poco frecuente, entonces el tamaño de muestra necesario ha de ser muy alto.

Siendo así, la regla del 50% pierde todo sentido, ya que si se trabaja con errores absolutos, éstos no se pueden establecer racionalmente bajo el supuesto de que no se tiene idea alguna de cuál es el valor de P ; y si se trabaja con errores

relativos, deja de ser válida la afirmación de que n está acotada superiormente en virtud de que lo está el producto $P(100-P)$.

- ***El nivel de confiabilidad:***

No existe criterio que determine que, dada determinadas circunstancias, la confiabilidad planificada para las inferencias de una investigación requiera de una magnitud específica. Se suele elegir 95%, pero este número no tiene más méritos que 90% o 99%, o 93,7%. La subjetividad, por tanto, también aparece en este caso.

- ***La determinación del error mínimo admisible, E_0 :***

Cuando se calcula el tamaño de muestra mínimo requerido se debe realizar una preestimación de P , no solo porque aparece explícitamente en la fórmula, como ya vimos, sino porque sin este conocimiento es imposible decidir el valor del error absoluto que también aparece en ella.

En la identificación a priori del error máximo admisible se deberá considerar qué valor de este resulta lo suficientemente pequeño como para considerarlo admisible, decisión que no deja de ser esencialmente subjetiva pues, tal y como sucede con la determinación del nivel de confianza, no se puede establecer un juicio carente de subjetividad que precise, según las particularidades de la investigación, si determinado error es más adecuado que otro.

Existen además dos elementos que en determinadas ocasiones culminan el bloque de fórmulas para el cálculo del tamaño muestral: el efecto de diseño y la

tasa de no respuesta. Estos no se deben pasar por alto, sobre todo debido a que la determinación de sus valores también entraña decisiones subjetivas. Su participación generalmente tiene lugar en estudios observacionales donde la aleatorización solo se garantiza mediante el muestreo, comúnmente aplicado antes de obtener el consentimiento de los efectivos y que en ocasiones debe adoptar formas complejas mediante la estratificación de la población, a veces en varias etapas.

- ***El efecto de diseño (deff):***

Todas las expresiones matemáticas que permiten calcular el tamaño de muestra se basan en el muestreo simple aleatorio, aunque con frecuencia los diseños muestrales que se implementan en las investigaciones son de naturaleza compleja, con etapas de selección que incluyen conglomerados y/o estratos.^{1,4,18, 113}

En estos casos, los errores no solo dependen del tamaño total de la muestra sino también del número de conglomerados en cada etapa y de sus tamaños cuando se usa un muestreo polietápico.

Para alcanzar bajo estas circunstancias un grado de precisión previamente establecido (E_0) es necesario seleccionar una muestra de mayor tamaño que el que demanda el muestreo simple aleatorio. De ahí que resulte necesario, tras obtener el valor aportado por la fórmula inicial, multiplicarlo por un número (el llamado efecto de diseño) que contrarreste el efecto desfavorable del diseño muestral sobre la exactitud de las inferencias.

La cuantificación de este valor es totalmente subjetiva. Algunos autores plantean como regla asumir un d_{eff} igual a 2,^{1,18} mientras que otros sugieren emplear, tras hacer ciertas especulaciones que no detallaremos, valores entre 1,5 y 2, o entre 1,5 y 4. Cálculos en algunos estudios orientados a precisarlo han evidenciado que su valor cambia radicalmente de un diseño muestral a otro, de una muestra a otra, y que dentro de un mismo diseño inclusive suele modificarse notablemente en dependencia del parámetro elegido.^{1, 4} El valor empleado como d_{eff} es quizás lo que con más subjetividad se anticipa.

- ***La tasa de no respuesta:***

La muestra sufre generalmente una reducción de su tamaño cuando en la práctica algunos de los efectivos que al inicio fueron seleccionados en el proceso de muestreo no consienten participar en el estudio o no pueden ser interrogados o medidos por alguna razón. La magnitud en que se presenta este fenómeno, conocida como “tasa de no respuesta”, debe ser tomada en cuenta por el investigador quien, para atenuar su efecto, deberá incrementar el valor resultante de la fórmula en una cantidad que compense por las unidades que no podrán ser estudiadas.

Establecer previamente tal cantidad es literalmente imposible. Solo puede especularse atendiendo a características muy particulares de cada estudio y de la población sobre la que se trabaja. El valor que se determine finalmente como tasa de no respuesta, incluso cuando se tomen como ejemplo

estudios de iguales características, será en cualquier caso un acto con importantes “cuotas de adivinación”.^{4,18}

De todo lo anterior se deduce que, como ha sido ilustrado por algunos autores,^{7,18} la realidad es tal que dos investigadores independientes, siguiendo la misma estrategia general (usando las mismas fórmulas), pueden obtener tamaños abismalmente diferentes, sólo por el hecho de que han fijado —legítimamente ambos— de manera diferente, incluso ligeramente distinta, los valores que tienen que incluir dentro de las fórmulas para llegar al tamaño muestral.

Es claro que justificar el tamaño muestral basándose en expresiones matemáticas no confiere objetividad a tal decisión ni aporta elementos de rigor científico. Todo dependerá de la elección de los valores que serán sustituidos en cada fórmula lo cual está determinado por la percepción individual del fenómeno, la intuición y la experiencia del investigador, adscritas al desarrollo de los conocimientos alcanzados en la materia que se estudia y fundamentalmente a los recursos disponibles.

Tales realidades conducen a asegurar que no existe para cada investigación un óptimo tamaño de muestra único.¹²¹ Calcularlo algebraicamente se diferencia poco de establecerlo usando el sentido común. En este sentido, Rothman,¹²² reconociendo con crudeza la inviabilidad de una solución teórica, plantea:⁶

"En resumidas cuentas, el problema de determinar el tamaño de muestra más adecuado no es de naturaleza técnica; no es susceptible de ser resuelto por vía de los cómputos, sino que ha de encararse mediante el juicio, la experiencia y la intuición."

En el centro de esta problemática se encuentra el hecho incuestionable de que **es la cantidad de recursos disponibles lo que rige la determinación del tamaño muestral.**

En este sentido, numerosas propuestas teóricas se han elaborado para incluir en las fórmulas parámetros que responden a aspectos de orden económico, principalmente en lo relativo al costo-eficiencia.¹²³ Un número importante de métodos Bayesianos, propuestos para seleccionar el tamaño muestral, toman en consideración los costos.¹²⁴⁻¹³³ Sin embargo, éstos son por lo general desconocidos, o bien desdeñados por constituir expresiones extremadamente complejas y abigarradas cuya aplicación exige especificaciones de costos y variabilidades que escapan del campo del investigador en salud sin dejar de estar también sujetas a un alto –incluso mayor– grado de subjetividad.^{123, 134-136}

Sin embargo en la práctica no es posible decidir la cantidad de unidades que pueden ser estudiadas sin tener en cuenta las cuestiones relativas a la factibilidad.

4,18, 19, 110, 123, 137

En consecuencia, la metodología frecuentemente se encamina a determinar el tamaño muestral mínimo necesario basándose en la simple conjunción entre el sentido común, y los costos, y posteriormente ajustar los datos de las fórmulas en

correspondencia a este valor, “invirtiendo la ecuación” para el cómputo del tamaño muestral.^{4,18,110, 123, 126}

Esta manipulación ha sido descrita en varios artículos especializados con el nombre de *ajuste retrospectivo*.^{viii 13, 20, 123} Su aplicación constituye la más clara expresión del empleo ceremonial de las fórmulas para la determinación del tamaño muestral.

2. Análisis de las demandas vigentes acerca de lo que debe comunicarse sobre la determinación del tamaño de muestra en estudios publicados.

El problema del uso ritual del procedimiento para determinar el tamaño muestral se extiende al marco de las obligaciones que tiene el investigador de hacer explícito este proceso, esto es, la exigencia de que a la hora de su publicación los artículos contengan una justificación detallada de cómo fue determinado el tamaño de muestra empleado. Es este un requisito incluido en las principales guías de publicación a las que se adhiere la mayor parte de las revistas de avanzada en la investigación sanitaria.^{62-66, 70}

Se trata de un fenómeno muy llamativo, ya que casi nadie discute o explica (aunque quizás tampoco sean muchos los que lo acaten realmente) por qué esto pudiera ser útil o conveniente.

Dos de las escasas razones esgrimidas a favor de esta demanda aparecen en un artículo publicado en la revista *Annals of Internal Medicine* en 1994.¹³⁸ En él los autores sostienen que la comunicación del cálculo del tamaño de muestra puede

^{viii} En Inglés: retrofitting

dar una idea de la calidad de la planificación y ejecución de la investigación-lo cual puede verificarse revisando la correspondencia entre el tamaño muestral planificado a priori y el de la muestra finalmente empleada- así como que si las consideraciones para hacerlo son presentadas de manera sofisticada, eso podría incrementar nuestra confianza en que los análisis difíciles fueron manejados apropiadamente. En un trabajo posterior, Faragher apoya el primero de estos argumentos.¹³⁹

Este es un débil argumento a nuestro juicio, ya que el hecho de que el artículo cumpla con dicho requisito no demuestra nada. En rigor, los autores pueden haber incurrido en el ajuste retrospectivo y así “lograr” a posteriori, que ambos valores se correspondan, al tiempo que el tamaño muestral aparece planificado de manera coherente. En este sentido, es imposible demostrar la veracidad de lo que afirma en el artículo, lo cual agrega debilidad a dicha tesis y apoya la idea de que lo que en ella se exige tampoco es garantía de la conducción apropiada de ningún análisis. Más bien, es verosímil que ocurra lo contrario: como se exige determinada conducta, se comunica que se esta ha sido seguida; cuando en realidad no se ha seguido a raíz de que se exige.

Como parte del debate que se ha generado en los últimos años en torno si es o no necesario comunicar el cálculo del tamaño muestral en los artículos, en un trabajo trascendente (especialmente por haber sido publicado en Lancet), Schulz y Grimes¹³ admiten la idea de que dicho requisito debe aparecer en los artículos- específicamente en los que reportan ensayos clínicos controlados- debido a que “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 (de ser así los autores deben informar una regla formal de parada estadística)”; y le atribuyen un valor añadido en tanto posible alerta, ya que “en caso de faltar esta información los lectores deben interpretar cautelosamente los resultados reportados.”¹³

Tales planteamientos son asombrosamente endeble. En primer lugar, es obvio que los datos que ellos consideran que deben encontrarse en el reporte de la determinación del tamaño muestral podrían comunicarse en los artículos independientemente de que se informe o no cómo se realizó el cómputo. Por otra parte, los resultados de un estudio pueden ser adecuadamente interpretados sin que aparezca esta información, siempre y cuando se provea al lector de las estimaciones realizadas y sus correspondientes intervalos de confianza. Dicho de otro modo, la *cautela* en la valoración de los resultados no se ha de incrementar o disminuir en función de que se expliciten o no las fórmulas para calcular el tamaño de muestra sino estrictamente en función de cuál fue el tamaño empleado, y más precisamente, de los márgenes de error que dicho tamaño contribuyó a producir.

Schulz y Grimes plantean además que negarse a reportar el cálculo del tamaño muestral indicaría una ingenuidad metodológica que pudiera señalar otros problemas. Lo que no aclaran estos autores es cuáles son estos problemas.

En otro orden, y en defensa de la pertinencia del reporte del cálculo del tamaño muestral en los artículos publicados, un argumento que pudiera ser esgrimido es el que se refiere a la replicabilidad de los estudios. Es cierto que el artículo

científico debe contener una detallada exposición de los procedimientos utilizados en el terreno experimental o para la observación, así como el diseño utilizado y los métodos estadísticos empleados para el análisis en correspondencia con las preguntas formuladas en el trabajo; ⁴ ya que esta información es necesaria dado el caso que se desee replicar el estudio.

En realidad, la lógica y la práctica habitual indican que cuando un estudio merita ser replicado, generalmente es porque se desea responder a la misma pregunta científica con un mayor grado de certeza, lo que casi siempre implica no sólo trabajar con un tamaño de muestra mayor, sino que con frecuencia se emplean los parámetros estimados en el estudio previo para “introducirllos” en las fórmulas empleadas en el cómputo de tamaño para la nueva investigación. Pero ni siquiera en el caso de que se deseara realizar este procedimiento utilizando exactamente el mismo tamaño muestral es preciso conocer cómo fue determinado.

Es una excelente noticia que en trabajos más recientes, se ha comenzado a reaccionar desde importantes tribunas científicas contra los rituales arraigados en esta materia ^{13, 37, 121, 138, 140}

Tal es el caso que, en el artículo citado más arriba, ¹³⁸ Goodman y Berin destacan que los detalles de un cálculo de tamaño de la prueba no son de ninguna ayuda para la decodificación estadística de los resultados. En este sentido -refiriéndose a los estudios experimentales- algunos plantean que son útiles porque muestran la opinión del investigador sobre el efecto más probable. Sin embargo no existe ninguna certeza de que la diferencia empleada en el cálculo refleje la opinión

previa de nadie y en cualquier caso esta solo es importante en la medida en que se base en evidencias objetivas. Si así fuera, estas deben presentarse como parte de la discusión para que el lector pueda interpretar por él mismo el efecto observado en el estudio, a la luz de la evidencia previa.

Por otra parte, Bacchetti alerta sobre otro problema que entraña la formalidad de reportar cómo se determinó el tamaño muestral: el efecto que la perpetuación de esta norma pudiera tener en el proceso de revisión por pares de los artículos.²⁹

La información disponible para determinar con precisión un tamaño muestral correcto (un calificativo impropio en sí) es a menudo menor que la necesaria y adicionalmente-como ha sido mencionado- la subjetividad es inherente a la elección de los datos a utilizar para el cálculo.^{13, 21} Esto, unido al pobre dominio de la Estadística por parte de muchos revisores¹⁴¹ -especialmente en lo referido a que las reglas deben cumplirse ciegamente- pudiera devenir en que el área del tamaño muestral sea motivo de críticas injustas e inútiles.²⁹

Debido a la ambigüedad inherente a la planificación del tamaño muestral- los revisores pueden siempre objetar las justificaciones reportadas para su determinación, y no pocas veces lo hacen. Un ejemplo ilustrativo de esta realidad ha sido proporcionado por Bacchetti, McCulloch y Segal.¹²³

Frente a la hegemonía de la estructura de los revisores, la inclusión ceremonial de fórmulas en los artículos que justifiquen el tamaño muestral es la defensa de los autores ante una demanda endeblemente fundamentada e ineficaz, carente de

una lógica persuasiva y que constituye una fuente para el mantenimiento de muchas de las ideas inerciales o irracionales relacionadas con el controvertido problema del tamaño muestral, que dimanar a su vez, en buena medida, de las incomprendiones asociadas con las pruebas de hipótesis. Ella representa, “*una solución operacional a un problema real*”.¹³

En este sentido, la práctica del ajuste retrospectivo- bautizada por Shultz y Grimes como “*the sample size samba*”- parece ser bastante frecuente.^{13, 138} Con ello la existencia de la demanda que criticamos genera más problemas de los que resuelve en lo que se refiere al reporte transparente de la investigación.¹²¹

Al demandar que los investigadores expliquen en los artículos cómo fue planificado el tamaño de muestra y las razones para haber seleccionado ese y no otro, ni siquiera las propias Revistas Médicas Internacionales-incluso las que se adhieren al Consenso de Vancouver-han escapado de participar en la perpetuación del uso ritual de procedimientos estadístico-metodológicos.

A las imprecisiones inherentes al procedimiento de la determinación del tamaño muestral, a su naturaleza subjetiva y su empleo formal, se añade una demanda improductiva, que lo sería incluso si el problema tuviese una solución inequívoca.

La orfandad de argumentos racionales de tal acción comunicativa y el hecho de que los revisores deban tenerla en cuenta en el proceso editorial, pone de manifiesto su carácter formal y favorece su empleo ceremonial.

A tono con esta realidad, resulta interesante caracterizar la manera en que se comunica la determinación del tamaño muestral y aquilatar la medida en que se

cumplen las exigencias establecidas por los comités editoriales de las revistas en los artículos que se publican en la actualidad.

Los resultados de este examen se exponen a continuación:

2.1 Resultados del estudio bibliométrico.

Durante el año 2009 el total de artículos originales publicados en las seis revistas médicas seleccionadas ascendió a 982 (Tabla 1).

Tabla 1. Porcentajes de artículos originales en las diferentes revistas. Año 2009.

Revista	n (%)
Ann Intern Med	91 (9)
BMJ	244 (25)
JAMA	175 (18)
Lancet	169 (17)
New Engl J Med	226 (23)
PloS Med	77 (8)
TOTAL	982 (100)

De acuerdo con los criterios definidos a los efectos del presente estudio, se excluyeron del análisis 228 (23%) trabajos. Entre estos, los más frecuentes fueron los que reportaron revisiones sistemáticas y/o meta-análisis. (Tabla 2)

Tabla 2. Distribución de artículos excluidos según tipo de estudio.

Tipo de estudio	n (%)
Revisión sistemática y/o meta-análisis	60 (26,3)
Estudio ecológico	30 (13,2)
Análisis de costo -efectividad	25 (11)
Reporte de un caso	21 (9,2)
Ensayo clínico fase I, II ó I-II	18 (7,9)
Investigación cuasi-experimental	17 (7,5)
Estudio de casos	11 (4,8)
Revisión narrativa	10 (4,4)
Modelación matemática	9 (3,9)
Investigación cualitativa	7 (3,1)
Investigación básica que no implica unidades de análisis	6 (2,6)
Estudio de eficacia de medios diagnósticos	6 (2,6)
Análisis conjunto de varios ensayos clínicos	5 (2,2)
Investigación metodológica	2 (0,9)
Ensayo clínico con diseño adaptativo	1 (0,4)
TOTAL	228 (100)

En consecuencia todo el análisis que subsigue se realiza a partir de los 754 trabajos incluidos. Entre estos predominaron los estudios observacionales, que ascendieron a 415 (55%).

De los restantes 339-de naturaleza experimental- 327 (97%) fueron ensayos clínicos controlados aleatorizados y de ellos 36 (11%) emplearon diseño con conglomerados aleatorizados (cluster randomized trials). Entre los estudios experimentales se encontraron además, 7 investigaciones básicas (2%), 2 estudios preclínicos (0,6%), 1 intervención comunitaria (0,3%), 1 estudio de eficacia y 1 trabajo calificado por los autores como “estudio piloto”.

En cuatro de las seis revistas estudiadas se observó predominio de los estudios observacionales. (Tabla 3)

Tabla 3. Porcentajes de artículos incluidos por tipo de estudio en las diferentes revistas.

Revista	Observacionales	Experimentales	Total
	n (%)	n (%)	n
Ann Intern Med	41 (44)	32 (56)	73
BMJ	115 (32)	54 (68)	169
JAMA	107 (32)	51 (68)	158
Lancet	47 (64)	82 (36)	129
New Engl J Med	71 (60)	105 (40)	176
PloS Med	34 (31)	15 (69)	49

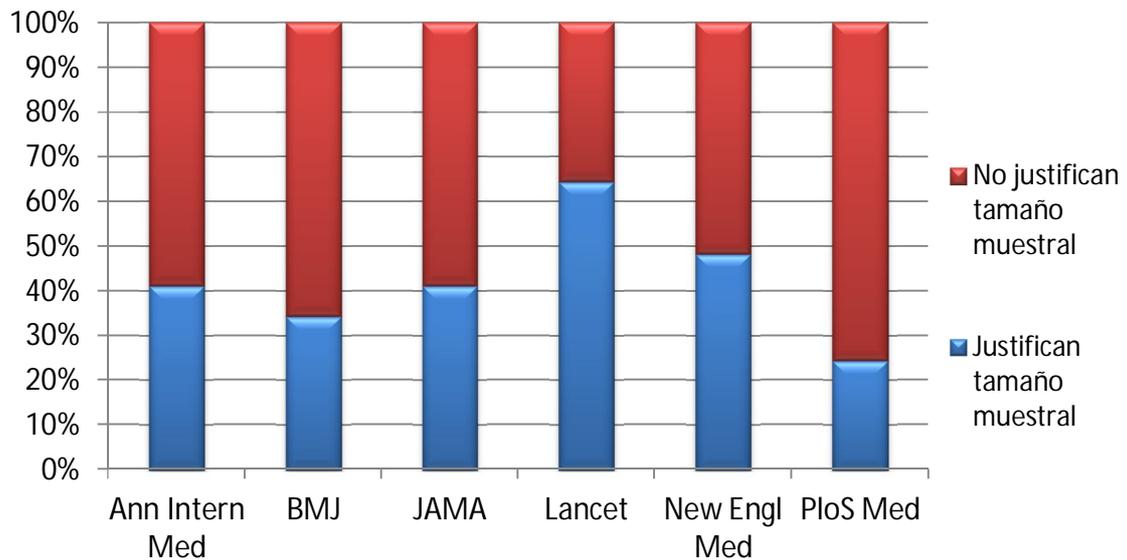
2.1.1) Información sobre la determinación del tamaño de muestra.

En el 56% (IC: 52-59) de los trabajos no se ofrecieron razones que justificasen el tamaño de muestra anunciado. Solo en una de las seis revistas estudiadas (Lancet) se observó un predominio de los artículos donde se justifica el tamaño muestral sobre aquellos en los que esto no ocurre. (Tabla 4, Gráfico 1)

Tabla 4. Porcentajes (IC) de artículos donde se justifica el tamaño muestral en las diferentes revistas.

Revista	Si justificado		No justificado	
	n	% (IC)	n	% (IC)
Ann Intern Med	30	41 (30-53)	43	59 (48-70)
BMJ	58	34 (27-42)	111	66 (59-73)
JAMA	65	41 (33-49)	93	59 (51-67)
Lancet	83	64 (56-73)	46	36 (27-44)
New Engl J Med	85	48 (41-56)	91	52 (44-59)
PloS Med	12	25 (12-37)	37	76 (63-88)
Total	333	44 (41-48)	421	56 (52-59)

Gráfico 1. Distribución porcentual de artículos según el tamaño muestral haya sido o no justificado en las diferentes revistas.



Fuente: Tabla 4

La ausencia de la justificación del tamaño de muestra fue notablemente mayor entre los trabajos observacionales (92%; IC: 89-94) con respecto a los experimentales (12%; IC: 8-15). (Tabla 5)

Tabla 5. Porcentajes de artículos por tipo de estudio y justificación del tamaño muestral en las diferentes revistas.

Revista	Observacionales		Experimentales	
	Sí justificado n (%)	No justificado n (%)	Sí justificado n (%)	No justificado n (%)
Ann Intern Med	1 (2)	40 (98)	29 (91)	3 (9)
BMJ	5 (4)	110 (96)	53 (98)	1 (2)
JAMA	17 (16)	90 (84)	48 (94)	3 (6)
Lancet	7 (15)	40 (85)	76 (93)	6 (7)
New Engl J Med	2 (3)	69 (97)	83 (79)	22 (21)
PloS Med	2 (6)	32 (94)	10 (67)	5 (33)
Total	34 (8)	381 (92)	299 (88)	40 (12)

En todos los artículos en que el tamaño de muestra fue justificado, se examinó el **método utilizado para su determinación** y se encontró que en 327 de ellos (98%; IC:97-100) se computó mediante fórmulas. En ningún estudio observacional y en sólo 6 de los 299 (2%; IC:0.4-4) trabajos experimentales (4 ensayos clínicos, 1 estudio piloto y 1 trabajo de intervención comunitaria) en que se justificó el tamaño muestral anunciado se comunicó haberlo determinado sin emplear algoritmos matemáticos. Entre ellos, 5 ofrecieron razones de factibilidad y 1 basó su decisión en información de estudios anteriores.

En la mayor parte de los estudios observacionales en los que se determinó el tamaño muestral mediante fórmulas el análisis estadístico se basó en la comparación mediante pruebas de hipótesis de medias o porcentajes entre dos poblaciones (30 de 34). Con relación a la **comunicación de los datos necesarios para el cómputo** en este grupo de artículos se encontraron los siguientes resultados:

- En 23 (77%) se reportó el valor predeterminado del nivel de confiabilidad, que fue de 0,05 en 22 trabajos y de 0,1 en el restante.
- En 27 (90%) se ofreció el valor prefijado de la potencia de la prueba.
- En 14 (47%) se comunicaron los valores de las preestimaciones del parámetro de interés (en el que se basó el cálculo del tamaño muestral) en cada población involucrada en el estudio, mientras que en 15 se reportó este dato para solo una de las poblaciones y en el trabajo restante no fue comunicado en lo absoluto.

Los cuatro estudios observacionales restantes emplearon fórmulas para la determinación del tamaño muestral basadas en la estimación de una proporción poblacional. Dos de ellos comunicaron todos los datos necesarios para el cálculo. Un trabajo reportó sólo los valores predeterminados del nivel de confiabilidad y del error relativo, mientras que otro reportó este último dato y el correspondiente a la preestimación del parámetro poblacional.

En resumen, entre los 34 estudios observacionales en los que determinó el tamaño muestral mediante fórmulas sólo en 16 (47%) se **comunicaron todos los**

datos necesarios para el cómputo. Por otra parte, sólo tres (9%) consideraron la tasa de no respuesta para el cálculo.

Se revisó asimismo, en el grupo de los estudios experimentales donde el tamaño muestral se determinó mediante fórmula, la información referente **los datos necesarios para el cómputo.** En este grupo se ubicaron 292 ensayos clínicos y 1 estudio preclínico, en los que el cálculo del tamaño de muestra se basó en la comparación mediante pruebas de hipótesis. Un análisis conjunto de esos trabajos arrojó los siguientes resultados:

- En 214 (82%; IC:78-87) se reportó el valor predeterminado del nivel de confiabilidad.
- En 280 (96%; IC:93-98) se ofreció el valor prefijado de la potencia de la prueba.
- En 172 (59%; IC:53-64) se comunicó el valor esperado de la variable principal de respuesta en el grupo de referencia.
- En 278 (95%; IC:92-98) se comunicó el valor mínimo de interés para la diferencia entre grupos en relación con la variable principal de respuesta.

Ciento-cuarenta y tres estudios experimentales con cómputo del tamaño muestral mediante fórmula **comunicaron todos los datos utilizados** para el mismo, lo que representó sólo el 49% (IC:43-55). Adicionalmente se encontró que en 113 (39%) artículos de este último grupo se tuvo en cuenta la tasa de no respuesta en el cálculo.

2.1.1.A) Análisis de la información sobre la determinación del tamaño muestral en el caso de los ensayos clínicos controlados.

El número de ensayos clínicos publicados en las seis revistas seleccionadas en el periodo estudiado ascendió a 327 y representó el 96% de los estudios experimentales. El **tamaño muestral fue justificado** en 296 de ellos (91%; IC: 87-94), de los cuales la casi totalidad emplearon fórmulas: 292 (99%; IC:97-100).

La información referente a la comunicación de los datos necesarios para el cálculo se presenta en la tabla 6.

Tabla 6. Porcentajes de ensayos clínicos que reportan información empleada para el cálculo del tamaño muestral para cada uno de los datos necesarios.

Dato	Frecuencia del reporte (%)
Nivel de confiabilidad (alfa)	240 (82)
Potencia	279 (96)
Valor pre-estimado de la variable principal de respuesta en el grupo de referencia.	172 (59)
Beneficio mínimo clínicamente relevante.	278 (95)

En adición a la información allí resumida, se encontró que el valor estimado de la desviación estándar de la respuesta entre sujetos fue comunicado en 56 (69%) de los 81 ensayos clínicos cuya variable principal de respuesta fue cuantitativa.

Además, en los 36 ensayos clínicos con diseño con conglomerados la frecuencia de comunicación de los demás datos necesarios para el cálculo se comportó como sigue: método empleado: 29 (81%), cantidad de conglomerados: 31 (86%),

tamaño de los conglomerados: 26 (72%), coeficiente intraconglomerados: 27 (75%).

En resumen, 143 (49%; IC:43-55) de los ensayos clínicos en los que el tamaño de muestra fue determinado mediante fórmulas **comunicaron todos los datos necesarios para el cómputo**. Se encontró además que 113 (39%; IC:33-44) refirieron haber considerado la tasa de no respuesta en el cálculo.

2.1.2) Tamaño muestral anunciado vs tamaño final del estudio.

En 29 de los 333 estudios en los cuales el tamaño muestral fue justificado no se informó el valor resultante. (9%; IC:6-12) (Tabla 7) La frecuencia de la comunicación de este dato fue menor en los estudios observacionales (29 artículos, 85%; IC:73-97) que en los experimentales (275 artículos, 92%;IC:89-95).

Tabla 7. Porcentajes de artículos que anticipan el valor del tamaño muestral a emplear en las diferentes revistas.

Revista	n (%*)
Ann Intern Med	27 (90)
BMJ	56 (97)
JAMA	59 (91)
Lancet	79 (95)
New Engl J Med	72 (85)
PloS Med	11 (92)
Total	304 (91)

**con respecto al total de artículos con tamaño muestral justificado.*

En solo 35 (12%) de los 304 trabajos que anticiparon el valor del tamaño muestral, este coincidió con el tamaño final. Entre los restantes estudios, la diferencia relativa entre esos dos valores fue positiva en la mayor parte de los casos (177; 66%) y su valor medio fue de 14,3% (IC:5,9-22,7). En 94 (35%) de esos trabajos la diferencia relativa fue superior a 15%.

2.1.3) Cumplimiento de las demandas planteadas en las guías de publicación.

De manera general, los requisitos establecidos en las guías de publicación para la información sobre la determinación del tamaño muestral incluyen: explicar y justificar cómo fue decidido, comunicar los datos necesarios para el cálculo en caso de que el método empleado sea algún algoritmo matemático e informar el tamaño muestral determinado a priori. Tomando en cuenta esta información se analizó la coherencia entre lo que realmente se incluye en los artículos y lo que reclaman las guías de publicación.

De los 754 artículos incluidos en el análisis, 564 (75%) fueron publicados en revistas que, según el tipo de estudio, se adhieren a una guía de publicación que incluye el requisito de justificar el tamaño muestral, así como comunicar los datos empleados en caso de que este fuese calculado.

Se encontró que en 311 de esos trabajos se justificó el tamaño de muestra (55%;IC: 51-59) En 307 (99%) de ellos se emplearon fórmulas, pero solo se informaron todos los datos necesarios para el cómputo en 151. En tres de estos trabajos no se informó el valor del tamaño muestral calculado a priori.

Por otra parte, de los 4 estudios que no emplearon fórmulas para la justificación (todos ensayos clínicos) uno no informó el valor del tamaño muestral determinado a priori.

En consecuencia, de los trabajos publicados en revistas que, según el tipo de estudio, se adhieren a una guía de publicación que incluye el requisito de justificar el tamaño muestral, **solo 151 (27%; IC:23-30) cumplieron con todas las demandas que figuran** en la correspondiente guía.

Se encontró mayor cumplimiento en los estudios experimentales (44%; IC:38-49) que en los observacionales (3%; IC:1-6). (Tabla 8)

Tabla 8. Porcentajes de artículos que cumplen con todas las demandas de las propias revistas acerca de la información a dar sobre la determinación del tamaño muestral, en las diferentes revistas.

Revista	Observacionales	Experimentales	Total
	n (%)	n (%)	n(%)
Ann Intern Med	1 (2)	19 (59)	20 (27)
BMJ	3 (3)	26 (48)	29 (17)
JAMA	-	17 (35)	17(35)
Lancet	3 (6)	42 (52)	45(35)
New Engl J Med	-	34 (33)	34(33)
PloS Med	1 (3)	5 (56)	6(14)
Total	8 (3)	143 (44)	151 (27)

DISCUSIÓN

Los resultados de nuestro estudio sugieren que ni revisores, ni editores ni mucho menos los autores, toman muy en serio la norma de explicar en los artículos científicos cómo determinaron el tamaño de muestra, aun cuando ella figura en las guías de publicación a las que se adhieren las principales revistas biomédicas-incluidas las de más alto factor de impacto. Acaso sea este el indicio más claro de que tal información es esencialmente estéril.

De hecho, el tamaño muestral solo aparece justificado en menos de la mitad de las publicaciones representativas de la investigación sanitaria contemporánea; situación que se refleja en mayor medida en la de tipo observacional. Incluso en aquellos trabajos en los que se justifica la elección del tamaño de muestra, algunos de los parámetros requeridos para el cálculo están frecuentemente ausentes en los reportes, aunque esta es también una exigencia explícita.

Estos datos reflejan la escasa adherencia a las demandas que figuran en las guías de publicación en relación con este tema. El no cumplimiento de las mismas por parte de autores, revisores y editores constituye, en nuestra opinión, un indicador irrefutable de su irracionalidad e inoperancia. Esta realidad toma mayor relevancia dado que ocurre en revistas de alto impacto con un proceso de revisión por pares.

Previo al presente estudio, no pocas revisiones de publicaciones habían mostrado que el reporte de la determinación del tamaño muestral está ausente con frecuencia en los artículos.

Sin embargo, los trabajos publicados hasta la fecha que abordan este asunto se constriñen casi exclusivamente al área de los ensayos clínicos, y la mayor parte de ellos incluye solo artículos sobre determinadas disciplinas médicas. En este punto cabe enfatizar que una fortaleza de nuestra investigación radica en que no se restringe a un área específica del campo de la salud y a que abarca tanto estudios observacionales como experimentales.

La frecuencia con que aparece la información sobre la determinación del tamaño de muestra según estudios publicados, es enteramente consistente con la que se halló en la presente investigación. Es abultado el grupo de trabajos previos que han encontrado prevalencias de justificación del tamaño muestral inferiores al 50%,^{72-74, 77-81, 84, 86-88, 90-95, 97-99, 101,142} e incluso menores del 10%.^{83, 85, 86, 96, 100, 102, 143, 144} Entre ellos sólo unos pocos examinan el área de los estudios observacionales,^{96, 142-144} pero ponen de manifiesto que la ausencia de dicha información ha sido incluso notablemente mayor que la que se registra en trabajos experimentales. También este panorama es coincidente con nuestros hallazgos.

No resulta ocioso detenernos en algunos detalles expresivos entresacados de la profusa información revisada. En el más reciente de dichos artículos, publicado en 2010, Hopewell et al.⁸⁰ encontraron que el cálculo del tamaño muestral fue

reportado en solo el 45% de los informes de ensayos clínicos indexados en PubMed en el año 2006, mientras que, refiriéndose a este mismo tipo de estudios- específicamente a aquellos donde el sujeto es su propio control- Mills et al.⁸² encuentran que el tamaño muestral fue justificado en apenas una quinta parte de los 116 que habían sido indexados en el 2002. Asimismo, Chan et al.⁷⁹ observaron que en la cohorte de ensayos clínicos indexados dos años antes, el reporte de la determinación del tamaño de muestra fue mayormente inadecuado, y estuvo ausente en casi tres cuartas partes (73%) de los artículos revisados.

En esta misma línea, en sendas revisiones que incluyeron mayoritariamente trabajos publicados en revistas de alto impacto en fechas tan recientes como 2008 y 2003 y 2008, Tiruvoipati et. al.⁷⁷ y Mathieu et. al.⁸¹ encontraron que el tamaño muestral fue justificado en el 20% y 36% respectivamente, cifras incluso inferiores a las observadas en nuestro trabajo.

Por otra parte, y en lo que se refiere a los estudios observacionales Bachmann et.al.¹⁴⁴ revisaron 43 artículos publicados en revistas de prestigio (4 de ellas generales y 4 especializadas) entre los cuales solo el 5% (IC:1-16) comunicó la información referente al cálculo del tamaño de muestra.

A la realidad descrita se suma otro hecho relevante: los trabajos publicados hasta el momento muestran que la casi totalidad de los artículos en que se justifica el tamaño muestral comunican que el procedimiento empleado se basó en algún algoritmo matemático, y que entre estos, el reporte de los datos necesarios para el

cálculo es frecuentemente incompleto o impreciso.^{19, 77, 78, 88, 89, 98, 100, 108} Este hecho constituye una fuerte evidencia adicional del no cumplimiento de las demandas incluidas en las guías de publicación. En este sentido dichos estudios también se corresponden con los resultados de nuestra investigación.

Como parte de los trabajos que reflejan esta situación puede citarse uno recientemente publicado en la revista *British Medical Journal*.¹⁹ Sus autores revisaron 215 artículos de ensayos clínicos de superioridad, grupos paralelos y dos brazos de tratamiento, con una sola variable principal de respuesta, publicados entre el 1ro de enero del 2005 y el 31 de diciembre del 2006 en las revistas de más alto factor de impacto. Encontraron que en el 45% de los que reportaron el cálculo del tamaño muestral planificado faltaba al menos uno de los datos necesarios para el cómputo, situación similar a la observada por Tiruvoipati et al.⁷⁷.

Asimismo, en un trabajo publicado en *Pediatrics* en el año 2006-que incluye ensayos clínicos sobre parálisis cerebral infantil- Anttila et al. encontraron que la cifra fue de 33%,⁸⁰ al tiempo que Lesaffre et al.¹⁰⁸ estudiaron 34 ensayos clínicos que emplearon diseño con conglomerados aleatorizados entre los cuales solo uno cumplió con todas las demandas incluidas en la correspondiente guía de publicación en lo que al reporte del tamaño muestral se refiere.

Por último, en un trascendental estudio en el que Chan et al.⁷² cotejan la información reflejada en protocolos de ensayos clínicos con los artículos en los

que se informan sus resultados, estos autores encontraron que en el 30% de los trabajos que reportaron el cálculo del tamaño de muestra faltaba al menos uno de los datos necesarios para el cálculo. Como dato interesante encontraron además que en el 47% de los trabajos que reportaron el cálculo del tamaño de muestra en ambos sitios- protocolo y artículo- existieron discrepancias entre ellos en al menos uno de los datos para el cálculo.

En contraposición a lo antes expuesto y a lo encontrado en nuestra revisión, pocos estudios comunican haber encontrado que la mayoría de los artículos justifica el tamaño muestral planificado. Dos de ellos conciernen ^{19, 103} a revistas biomédicas generales de alto impacto, 3 a disciplinas médicas específicas ¹⁰⁵⁻¹⁰⁷ y otro incluye trabajos de ambos tipos de publicaciones. ¹⁰⁴

En resumen, en lo que se refiere al reporte de la justificación del tamaño de muestra en la investigación sanitaria contemporánea, la realidad actual parece susceptible de ser caracterizada del modo siguiente: tal información está muy mayoritariamente ausente en los artículos, y cuando aparece es con mucha frecuencia incompleta. Lo significativo de este resultado –cabe reiterarlo- es que refleja directamente que las demandas que figuran en las guías de publicación sobre la información relativa al tamaño muestral son frecuentemente incumplidas.

Ahora bien, ¿cómo explicar el divorcio entre la existencia de estos requisitos y el hecho de que no se cumpla con ellos? Es un hecho que tanto autores como revisores y editores confieren poca o ninguna importancia a exponer en un

artículo los elementos considerados para elegir el tamaño de muestra, ya sea mediante cálculos matemáticos o análisis de disponibilidad de recursos. Téngase en cuenta que, además de las evidencias empíricas presentadas sobre la escasa aparición de esta información, hay algo más en relación con los trabajos que sí la tienen. Se trata de que, a todas luces, el hecho de que dicha información haya sido comunicada no necesariamente dimana de que haya sido exigida, pues sería absurdo que, una vez que el autor disciplinadamente la incluye, alguien vaya a reclamar que se suprima.

La respuesta emerge con nitidez: tal y como habíamos discutido de manera genérica, se trata de información simplemente irrelevante para un lector o para un editor; carece de toda utilidad porque no agrega absolutamente nada a lo que informa el trabajo propiamente dicho a los efectos de la pregunta que éste procura responder.

El modo en que se arribó a un tamaño muestral es sumamente relevante para quien ha de decidir si el trabajo se hace o no, si se financia o no. Pero al lector no le interesa para nada conocer las motivaciones de los autores para trabajar con un tamaño de muestra u otro, ni las consideraciones que pudieran haber hecho quienes lo financiaron (las cuales, por otra parte, pudieron haber sido las determinantes, a la vez que nadie las informa ni ninguna guía las exige). Se trata, en suma, de una información absoluta y totalmente superflua a los efectos de valorar la calidad del trabajo o de entender sus resultados o sus interpretaciones.¹³⁸

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, pero conocer o no por qué se eligió el tamaño en cuestión no cambia ni unos ni otras. El tamaño es algo ya consumado y lo que interesa al lector es conocer qué se puede sacar en limpio de lo que objetivamente se ha hecho, no las motivaciones que pudieran haber existido para hacer una cosa u otra en materia de tamaño muestral. Conocer esas razones es tan útil como saber por qué el número de autores del trabajo fue el que fue, o la calidad de la computadora que emplearon, o por qué se seleccionó determinado algoritmo para la generación de números aleatorios.⁶

En nuestra opinión, defender dicha demanda implica contemporizar, cohonestar y hasta promover algunos de los problemas actuales en el uso de la Estadística, como ocultar las dificultades inherentes al proceso de selección del tamaño muestral.^{4-6, 18}

Se suman a este otros no menos relevantes. Uno de ellos es la arraigada e incorrecta idea mitológica de que pueden sacarse conclusiones definitivas y reglas operativas a partir de cada estudio aislado, la cual se vincula estrechamente con el tema de la determinación del tamaño muestral.^{6, 7, 13, 21, 138,145.}

La legitimidad de que el “tamaño de muestra suficiente” siga siendo un criterio de evaluación muy consolidado en comités de éticas, grupos de evaluación de

protocolos de investigación y revisores de revistas es muy discutible. Tal precaución solo tendría sentido si de un solo estudio pudieran sacarse conclusiones definitivas; pero en realidad 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 abierto a cambios y perfeccionamientos en la medida que nuevos datos lo aconsejen. Lo ideal sería contar con procedimientos que, en lugar de convocarnos a desechar o no una hipótesis, asuma lo que de hecho siempre termina haciéndose en la investigación verdadera, "poner al día" la opinión que la hipótesis nos merece a la luz de dichos datos.^{6, 9, 20, 36}

Al decir de Marks (2003):¹⁴⁵

"En la era en que el tránsito de los hallazgos clínicos que van desde New England Journal of Medicine al Wall Street Journal se hace más corto que nunca, es también mayor que nunca la necesidad de tener en cuenta la advertencia de Fisher acerca de la naturaleza provisional del conocimiento empírico y sobre la necesidad de medir la incertidumbre."

En este orden, las decisiones binarias (muestra grande/muestra pequeña, significativo/no significativo) responden al problema equivocado y se impone reconocer la necesidad de sustituir decisiones basadas en valores puntuales -

refiriéndose al valor p de las pruebas de significación- por intervalos de valores, que son los que nos dan la respuesta que como científicos necesitamos. ^{4, 6, 8-10,}

^{12, 14, 16, 22, 25, 39}

Un experimento en particular casi nunca producirá una respuesta definitiva a la pregunta planteada sino que se reduce a dar una contribución al proceso de permanente reajuste de nuestras convicciones. En ese sentido todo aporte es útil, unos más y otro menos, para ir construyendo respuestas científicamente fructíferas y no hay una respuesta única a la pregunta sobre cuál es el tamaño de muestra apropiado. ^{13, 20}

Por las razones expuestas, tanto teóricas como empíricas, nosotros sostenemos que el reporte de la determinación del tamaño muestral en los artículos debiera ser abandonado, en consonancia, por otra parte, con lo planteado por algunos autores. ^{18,19,121}

Cabe recordar que las justificaciones esgrimidas para sustentar la demanda de que dicha información aparezca en los artículos publicados son endeble. Justificar cómo fue determinado el tamaño muestral no es prueba de calidad metodológica. Tampoco es necesario hacerlo para informar debidamente sobre la variable principal de respuesta ni sobre la magnitud de la muestra planificada. ¹²¹

Asimismo es irrelevante para interpretar los resultados de las investigaciones, porque las estimaciones y sus correspondientes intervalos de confianza permiten una interpretación directa y confiable de estos. ^{104, 121, 146}

Como ha escrito Senn,¹⁴⁷ *“el cálculo del tamaño de muestra es una excusa para un tamaño muestral, no una razón, y la determinación del tamaño de muestra es habitualmente gobernada por la factibilidad”*. De hecho- tal y como muestran los resultados de la presente investigación- en solo una pequeña parte de los trabajos que anticipan el valor calculado del tamaño muestral, este coincide con el finalmente empleado. Esto hace que la exigencia de comunicar cómo aquel fue computado no promueva, sino que más bien menoscaba el reporte transparente de las investigaciones.¹²¹

Consecuentemente, las guías que demandan que esta información sea comunicada deben ser modificadas para, en cambio, desestimular su inclusión, tal y como ha sido sugerido recientemente por Bacchetti.¹²¹

Es difícil cambiar un modo de pensar tan arraigado. Por lo general las propuestas revolucionarias y novedosas encuentran en sus inicios pocos adeptos. Sin embargo, a la luz del análisis teórico realizado y los resultados encontrados, esta investigación cumplirá sus propósitos si contribuye a abrir un frente de debate en la comunidad científica en lo que a la pertinencia y racionalidad de esta demanda se refiere.

Aunque muchos piensan que la tradición de actuar con rigor y no en función de inercias insuficientemente fundamentadas (o directamente sin fundamento) es de muy vieja data, la realidad es que hace solo 40 años el *peer review* de las grandes revistas médicas era por lo general un acto informal, incompleto, no sistematizado

ni realmente respetado. Algo parecido pasaba con los criterios para la práctica clínica y con el empleo de la estadística.

La consolidación del *peer review* como accionar riguroso, ineludible, y susceptible de estudio crítico, el movimiento de la Medicina Basada en la Evidencia y los nuevos enfoques inferenciales tales como la estadística bayesiana, son solo algunos ejemplos de cómo en años recientes se han venido produciendo modificaciones sustantivas en el accionar que se derivaba de dichas tradiciones. Siendo así, no cabe sorprenderse de que asuntos más sutiles, como el que aquí se ha abordado, no solo merezcan rectificación sino que la consigan en plazos no demasiado largos.

CONCLUSIONES Y RECOMENDACIONES

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 adhiere la mayoría de las revistas biomédicas- carece de fundamento y utilidad. Ella constituye una expresión más del empleo ceremonial de la estadística y en tanto tal, conspira contra la racionalidad. Su vigencia no promueve, sino que más bien menoscaba la transparencia del reporte de las investigaciones. Como consecuencia natural de su esterilidad intrínseca, dicha regla es muy frecuentemente incumplida.

Se recomienda:

- A) Promover el debate sobre este tema en espacios destinados a promover el acrisolamiento y mejoría de las formas de realizar la comunicación científica, tales como revistas de amplio impacto o sitios especializados (e.j <http://www.equator-network.org>)
- B) Obrar en función de que tal demanda sea eliminada de las guías de publicación que las contienen.
- C) Monitorizar el grado de adherencia a las exigencias establecidas mediante nuevas investigaciones bibliométricas aplicadas a otros períodos y revistas. Sería particularmente oportuno sondear la opinión de investigadores y editores sobre las causas de su existencia.

REFERENCIAS BIBLIOGRÁFICAS

1. Hernández PL. Ceremonias estadístico-metodológicas en la Investigación médica contemporánea. [Tesis]. Ciudad de la Habana: Instituto Nacional de Higiene Epidemiología y Microbiología (INHEM); 2007.
2. Silva LC. Pensamiento clínico, métodos estadísticos clásicos y enfoque bayesiano. *Dimensión Humana*. 2003; 5 (4):182.
3. Silva LC. Límites y mitos de la estadística en la valoración de los resultados de los ensayos clínicos y otros estudios. *Atención Primaria*. 2007; 39 (Supl 2):12.
4. Silva LC. Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1997.
5. Silva LC. Apuntes sobre subjetividad y estadística en la investigación en salud. *Rev Cubana Salud Pública* 2003;29(2):170-73.
6. Silva LC. Los laberintos de la investigación biomédica. En defensa de la racionalidad para la ciencia del siglo XXI. Madrid: Díaz de Santos; 2009.
7. Silva LC. Los límites de las pruebas de significación estadística y los valores p. *AMF*. 2008;4(2):62-63.
8. Cohen J. The earth is round ($p < .05$). *Am Psychol* 1994;49:997-1003.
9. Rozeboom WW (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin* 56: 26-47.
10. Bakan D. The test of significance in psychological research. *Psychological Bulletin* 1966;66: 423-437.
11. Silva LC. Una ceremonia estadística para identificar factores de riesgo. *Salud Colectiva*. 2005; 1(3): 309-322.
12. Silva LC, Sarría M. La pruebas de significación estadística en tres revistas biomédicas: una revisión crítica. *Revista Panamericana de Salud Pública*. 2004; 15(5)
13. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365:1348-53.
14. Silva LC, Suárez P, Fernández A, Álvarez V, Iglesias T. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. Trabajo presentado en el Sexto Congreso Internacional sobre revisión por pares y publicación biomédica. Septiembre 10-12, 2009. Vancouver,

BC, Canadá. Resumen disponible en: http://www.ama-assn.org/public/peer/program_2009.html

15. Ioannidis JPA. Why most published research findings are false? *PLoS Medicine* 2005;2: e124.
16. Altman DG, Martin J. Absence of evidence is not evidence of absence. *BMJ* 1995;311-485.
17. Lipset SM, Trow MA, Coleman JS. Statistical problems. Apéndice 1-B de *Union Democracy*, Glencoe, Illinois: Free Press;1956.
18. Silva LC. Diseño razonado de muestras y captación de datos para la investigación sanitaria. Madrid: Díaz de Santos; 2000.
19. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of simple size calculations in randomised controlled trials: review. *BMJ*. 2009;338:b1732.
20. 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.
21. Martin J. The tyranny of power: is there a better way to calculate sample size? *BMJ* 2009;339:b3985.
22. Benavides A, Silva LC. Contra la sumisión estadística: un apunte sobre las pruebas de significación. *Metas de Enfermería*. 2000; 3(27).
23. Gigerenzer, G. We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences* 1998; 21, 199-200.
24. Ordúñez P, Silva LC, Rodríguez MP, Robles S. Prevalence estimates for hypertension in Latin America and the Caribbean: are they useful for surveillance? *Revista Panamericana de Salud Pública*. 2001; 10(4).
25. Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*. 2000; 5(2): 241–301.
26. 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*. Ed: Morris, J. Oxford: Butterworth;2000.
27. Comité Internacional de Directores de Revistas Médicas. Requisitos uniformes para la preparación de manuscritos. *Revista Panamericana de Salud Pública*. 2004; 15(1).
28. Goodman SN. Toward evidence based medical statistics (I): The p value fallacy. *Annals of Internal Medicine*. 1999; 130: 995-1004.

29. Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ* 2002;324:1271-3.
30. Bacchetti P, Wolf LE, Segal MR, McCulloch CE. Ethics and sample size. *Am J Epidemiol* 2005;161:105-10.
31. Senn S. Two cheers for p-values and two cheers for Bayes. *Journal of Epidemiology and Biostatistics* 2001;6:193-210.
32. Loftus GR. On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology* 1991; 36:102-105.
33. Anderson DR, Burnham KP. Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* 2002; 66:912–918.
34. Rozeboom WW. Good science is abductive, not hypothetico-deductive. En LL Harlow, SA Mulaik, & JH Steiger (Eds.). *What if there were no significance tests?* (pp. 335–391). Hillsdale, NJ: Erlbaum;1997.
35. Edwards SJ, Lilford RJ, Braunholtz D, Jackson J. Why “underpowered” trials are not necessarily unethical. *Lancet* 1997;350:804-7.
36. Cohen J. Things I have learned (so far). *American Psychologist* 1990;45: 1304-1312.
37. Thompson B. In praise of brilliance: Where that praise really belongs. *American Psychologist* 1998;53: 799–800.
38. Howard GS, Maxwell SE, Fleming KJ. The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods* 2000; 5:315-332.
39. Barnett ML, Mathisen A. Tyranny of the p-value: The conflict between statistical significance and common sense. Guest Editorial. *Journal of Dental Research*. 1997;76: 534-536.
40. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *Journal of the American Medical Association* 1998;280: 280-282.
41. Los valores P y los intervalos de confianza: ¿en qué confiar? *Revista Panamericana de Salud Pública /Pan American Journal of Public Health* 15: 293-296.
42. Smith GD. Uncertainty and significance. *International Journal of Epidemiology* 2003; 32: 683-685.

43. Fidler F et al. Editors can lead researchers to confidence intervals, but can't make them think: statistical reform lessons from medicine. *Psychology Science* 2004;15:119–126.
44. Guthery FS, Lusk JJ, Peterson MJ. The fall of the null hypothesis: liabilities and opportunities. *Journal of Wildlife Management* 2001;65: 379–384.
45. Almenara J et al. Historia de la bioestadística: la génesis, la normalidad y la crisis. Cádiz: Quórum SA; 2003.
46. Rushton SP, Ormerod SJ, Kerby G. New paradigms for modelling species' distributions? *Journal of Applied Ecology* 2004; 41: 193–200.
47. Lecoutre B. Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science* 2006;4: 207-232.
48. Silva LC. Muñoz A. Debate sobre métodos frecuentistas vs bayesianos. *Gaceta Sanitaria* 2000;14: 482-494.
49. Silva LC. La alternativa bayesiana. *Brotos* 2000;1:1-4.
50. Silva LC, Benavides A. El enfoque bayesiano: otra manera de inferir. *Gaceta Sanit* 2001;15:341-6.
51. Malakoff D. Bayes offers a 'new' way to make sense of numbers. *Science* 1999;286:1460-4.
52. Spiegelhalter DJ et al. (1999). An introduction to bayesian methods in healthy technology assessment. *British Medical Journal* 319: 508-512.
53. Wilkinson L. Task force on statistical inference. *Statistical methods in psychology journals*. *American Psychologist* 1999; 54: 594–604.
54. Marshall S. Testing with confidence: The use (and misuse) of confidence intervals in biomedical research. *Journal of Science and Medicine in Sport* 2004; 7: 135-137.
55. Thompson B. Various editorial policies regarding statistical significance tests and effect sizes. 2003. Disponible en <http://www.coe.tamu.edu/~bthompson> en Septiembre del 2009.
56. Comité Internacional de Editores de Revistas Médicas. Requisitos de uniformidad para manuscritos enviados a revistas biomédicas: Redacción y preparación de la edición de una publicación biomédica. Actualizado en Febrero del 2006.
57. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *Journal of the American Medical Association*. 1966; 195: 1123-1128.

58. Gore S, Jones IG, Rytter EC. Misuses of statistical methods: critical assessments of articles in BMJ from January to March, 1976. *British Medical Journal* 1977;1:85.
59. Sheehan TJ. The medical literature: Let the reader be aware. *Archives of Internal Medicine* 1980; 140: 472-474.
60. Freiman JA et al. The importance of beta, the type II error and sample size in the design and Interpretation of the randomized control trial: survey of 71 «negative» trials *New England Journal of Medicine* 1978;299: 600-694.
61. Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest* 2010; 40 (1): 35–53.
62. 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.
63. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
64. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004; 328;702-708.
65. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of Noninferiority and Equivalence Randomized Trial. An Extension of the CONSORT Statement. *JAMA*. 2006;295:1152-1160.
66. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med* 4(10):e297.
67. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. for the STARD steering group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–4.
68. Moher D, Liberati A, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med*. 2009;151:264-269.
69. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis J P.A, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Ann Intern Med*. 2009;151:W-65–W-94.

70. Little J, Higgins J P.T, Ioannidis J P.A, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA):An Extension of the STROBE Statement. *Ann Intern Med.* 2009;150:206-215.
71. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in Epidemiology. *JAMA* 2000;283:2008-2012.
72. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
73. Ziogas DC, Zintzaras E. Analysis of the Quality of Reporting of Randomized Controlled Trials in Acute and Chronic Myeloid Leukemia, and Myelodysplastic Syndromes as Governed by the CONSORT Statement. *AEP* Vol. 19, No. 7 July 2009: 494–500.
74. Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY, Sample size calculations in surgery: Are they done correctly? *Surgery* 2003;134:275-9.
75. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
76. Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984;5:328-47.
77. Tiruvoipati R, Balasubramanian SP, Atturu G, Peek GJ, Elbourne D. Improving the quality of reporting randomized controlled trials in cardiothoracic surgery: The way forward *J Thorac Cardiovasc Surg* 2006;132:233-40.
78. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22-4.
79. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
80. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723.
81. Mathieu S, Boutron I, Moher D, et al. Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials. *JAMA.* 2009;302(9):977-984.

82. Mills EJ, Chan A, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials*. 2009; 10: 27.
83. Wang G, Mao B, Xiong Z, Fan T, Chen X, Wang L, et al. The Quality of Reporting of Randomized Controlled Trials of Traditional Chinese Medicine: A Survey of 13 Randomly Selected Journals from Mainland China. *Clinical Therapeutics*. 2007; 29(7): 1456-1467.
84. Rios LP, Oduyungbo A, Moitri MO, Rahman MO, Thabane L. Quality of Reporting of Randomized Controlled Trials in General Endocrinology Literature. *The Journal of Clinical Endocrinology & Metabolism* Vol. 93, No. 10 3810-3816.
85. Thoma A, Chew RT, Sprague S, Veltri K. Application of the CONSORT statement to randomized controlled trials comparing endoscopic and open carpal tunnel release. *Can J Plast Surg*. 2006 Winter; 14(4): 205–210.
86. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000; 136:381–5.
87. Alvarez F, Meyer N, Gourraud PA, Paul C. CONSORT adoption and quality of reporting of randomized controlled trials: a systematic analysis in two dermatology journals. *British Journal of Dermatology* 2009 161, pp1159–1165.
88. Anttila H, Malmivaara A, Kunz R, Autti-Rämö I, Mäkelä M. Quality of Reporting of Randomized, Controlled Trials in Cerebral Palsy. *Pediatrics*. 2006;117;2222-2230.
89. Thabane L, Chu R, Cuddy K, Douketis J. What is the quality of reporting in weight loss intervention studies? A systematic review of randomized controlled trials. *International Journal of Obesity* (2007) 31, 1554–1559.
90. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122–4.
91. Stinson JN, McGrath PJ, Yamada JT. Clinical Trials in the Journal of Pediatric Psychology: Applying the CONSORT Statement. *Journal of Pediatric Psychology*, Vol. 28 No. 3, 2003, pp. 159–167.
92. Karri V. Randomised clinical trials in plastic surgery: Survey of output and quality of reporting. *Journal of Plastic, Reconstructive & Aesthetic Surgery* (2006) 59, 787–796.
93. Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol* 2007;14:390-8

94. Smith BA, Lee H, Lee J, Choi M, Jones DE, Baker-Bausell R, et al. Quality of reporting Randomized Controlled Trials (RCTs) in the nursing literature: Application of the Consolidated Standards of Reporting Trials (CONSORT). *Nurs Outlook* 2008;56:31-37.
95. Uetani K, Nakayama T, Ikai H, Yonemoto N, Moher D. Quality of Reports on Randomized Controlled Trials Conducted in Japan: Evaluation of Adherence to the CONSORT Statement. *Inter Med* 48: 307-313, 2009.
96. Fung AE, Palanki R, Bakri SJ, Depperschmidt E, Gibson A. Applying the CONSORT and STROBE Statements to Evaluate the Reporting Quality of Neovascular Age-related Macular Degeneration Studies. *Ophthalmology* 2009;116:286–296.
97. Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: A systematic review. *International Journal of surgery*. 5 (2007) 413-722.
98. Sánchez-Thorin JC, Cortés MC, Montenegro M, Villate N. The Quality of Reporting of Randomized Clinical Trials Published in Ophthalmology. *Ophthalmology* 2001;108:410–415.
99. Ladd BO, McCrady BS, Manuel JK, Campbell W. Improving the quality of reporting alcohol outcome studies: Effects of the CONSORT statement. *Addictive Behaviors* 35 (2010) 660–666.
100. Bhandari M, Guyatt GH, Lochner H, Sprague S, Tornetta P. Application of the Consolidated Standards of Reporting Trials (CONSORT) in the Fracture Care Literature. *J Bone Joint Surg Am*. 2002;84:485-489.
101. Piggott M, McGee H, Feuer D. Has CONSORT improved the reporting of randomized controlled trials in the palliative care literature? A systematic review. *Palliative Medicine* 2004; 18: 32-38.
102. Latronico N, Botteri M, Minelli C, Zanotti C, Bertolini G, Candiani A. Quality of reporting of randomised controlled trials in the intensive care literature: A systematic analysis of papers published in Intensive Care Medicine over 26 years. *Intensive Care Med* (2002) 28:1316–1323.
103. Mills EJ, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Contemp Clin Trials* 2005;26:480-7.
104. Mills E, Wu P, Gagnierc J, Heels-Ansdella D, Montoria VM. An analysis of general medical and specialist journals that endorse CONSORT found that

- reporting was not enforced consistently. *Journal of Clinical Epidemiology* 58 (2005) 662–667.
105. Greenfield ML, Mhyre JM, Mashour GA, Blum JM, Yen EC, Rosenberg AL. Improvement in the Quality of Randomized Controlled Trials Among General Anesthesiology Journals 2000 to 2006: A 6-Year Follow-Up. *Anesth Analg* 2009;108:1916–21.
106. Sinha Si, Sinha Sh, Ashby E, Jayaram R, Grocott M. Quality of Reporting in Randomized Trials Published in High-Quality Surgical Journals. *J Am Coll Surg* 2009;209:565–571.
107. Halpern SH, Darani R, Douglas MJ, Wight W, Yee J. Compliance with the CONSORT checklist in obstetric anaesthesia randomised controlled trials. *International Journal of Obstetric Anesthesia* (2004) 13, 207–214.
108. Lesaffre E, Garcia Zattera M-J, Redmond C, Huber H, Needleman I. Reported methodological quality of split-mouth studies. *J Clin Periodontol* 2007; 34: 756–761.
109. Thomson Scientific. Institute of Scientific Information. Journal Citation Reports. 2009. Disponible en: http://thomsonreuters.com/products_services/science/science_products/az/journal_citation_reports
110. Martínez M, Sánchez A, Faulin J. *Bioestadística amigable*. (2ªEd) Madrid: Díaz de Santos; 2006.
111. Daniel W. *Bioestadística. Base para el análisis de las ciencias de la salud*. Limusa. Distrito Federal; 2004.
112. Camel F. *Estadísticas médicas y de salud pública*. La Habana: Pueblo y Educación; 1978.
113. Lemeshow S, Hosmer Jr DW, Klar J, Lwanga SK. *Adequacy of sample size in health studies*. New York: Wiley; 1990.
114. Azorín F, Sanchez-crespo JL. *Métodos y aplicaciones del muestreo*. Madrid: Alianza; 1986.
115. Argimón JM, Jiménez J. *Métodos de investigación. Aplicados a la atención primaria de salud*. Barcelona: Doyma; 1991.
116. Jenicek M, Cleroux R. *Epidemiología: principios-técnicas-aplicaciones*. Barcelona: Salvat; 1987.

117. Marrugat J, Vila J, Pavesi M, Sanz F. Réplica. *Medicina Clínica*. 1999; 112:798.
118. Marrugat J, Vila J, Pavesi M. Supuesto de máxima indeterminación: ¿error absoluto o error relativo en el cálculo del tamaño de la muestra? *Gaceta Sanitaria*. 1999; 13: 491-493.
119. Suárez P, Alonso JC. Sobre el supuesto de máxima indeterminación, el tamaño muestral y otras consideraciones sobre muestreo. *Gaceta Sanitaria*. 1999; 13: 243-246.
120. Silva LC . Nueva visita al supuesto de máxima indeterminación y al empleo de errores absolutos y relativos. *Gaceta Sanitaria*. 1999; 14: 255-257.
121. Bacchetti P. Current simple size conventions: Flaws, harms and alternatives.
122. Rothman JK . *Modern Epidemiology*. Boston: Little, Brown and Col; 1986.
123. Bacchetti P, McCulloch CE, Segal MR. Simple, defensible simple sizes based on cost efficiency. *Biometrics* 2008;64:577-85; discussion 586-94.
124. Bernardo, J. M. Statistical inference as a decision problem: The choice of sample size. *The Statistician*. 1997; 46, 151–153.
125. Claxton, K. and Posnett, J. An economic approach to clinical trial design and research priority-setting. *Health Economics*. 1996; 5, 513–514.
126. Detsky, A. S. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. *Statistics in Medicine*. 1990; 9, 173–184.
127. Gittins, J. and Pezeshk, H. How large should a clinical trial be? *The Statistician*. 2000; 49, 177–197.
128. Gittins, J. and Pezeshk, H. A behavioral Bayes method for determining the size of a clinical trial. *Drug Information Journal*. 2000; 34, 355–363.
129. Halpern, J., Brown, B. W., and Hornberger, J. The sample size for a clinical trial: A Bayesian-decision theoretic approach. *Statistics in Medicine*. 2001; 20, 841–858.
130. Lindley, D. V. The choice of sample size. *The Statistician*. 1997; 46, 129–138.
131. Tan, S. B. and Smith, A. F. M. Exploratory thoughts on clinical trials with utilities. *Statistics in Medicine*. 1998; 17, 2771–2791.
132. Walker, S. G. How many samples? A Bayesian non-parametric approach. *The Statistician*. 2003; 52, 475–482.

133. Willan, A. R. and Pinto, E. M. The value of information and optimal clinical trial design. *Statistics in Medicine*. 2005; 24, 1791–1806
134. Mejía JM et al. El tamaño de muestra: un enfoque práctico en la investigación clínica pediátrica. *Boletín Médico del Hospital Infantil de México*. 1995; 52: 381-391.
135. Hansen MH, Hurwitz WN, Madow WG. *Sample survey methods and theory*. New York: Wiley; 1953.
136. Yokota, F. and Thompson, K. M. Value of information analysis in environmental health risk management decisions: Past, present, and future. *Risk Analysis*. 2004; 24, 635–650.
137. Torgerson DJ, Campbell MK. Cost effectiveness calculations and sample size. *BMJ* 2000;321:697.
138. Goodman SN, Berin JA. The use of predictive confidence intervals when planning experiments and de misuse of power when interpreting results. *Annals Int Med* 1994;121:200-206.
139. Faragher B. How confident are you of your study. *Dev Med Child Neurol*. 2003 Apr;45(4):219.
140. Bacchetti P. Author's thoughts on power calculations (letter). *British Medical Journal*. 2002; 325: 492-493.
141. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: Caveat lector? *J Gen Intern Med* 1998;13:753-6.
142. Papathanasiou AA, Zintzaras E, Assessing the Quality of Reporting of Observational Studies in Cancer. *Ann Epidemiol* 2010;20:67–73.
143. Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006;332:1127–9.
144. Bochmann F, Johnson Z, Azuara-Blanco A. Sample size in studies on diagnostic accuracy in ophthalmology: a literature survey. *Br J Ophthalmol* 2007;91:898–900.
145. Marks HM. Rigorous uncertainty: why RA Fisher is important. *International Journal of Epidemiology* 2003;32: 932–937.
146. Hoenig JM, Heisey DM: The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician* 2001, 55:19-24.
147. Senn SJ: Power is indeed irrelevant in interpreting completed studies. *British Medical Journal* 2002, 325:1304-1304.

ANEXOS

Anexo 1. Fórmulas empleadas con mayor frecuencia para calcular el tamaño de muestra.¹¹⁰

a) *Para la estimación de un porcentaje o una media:*

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad ; \quad n_0 = \frac{Z_{1-\alpha/2}^2 V}{E_0^2} \quad [1.1]$$

Datos necesarios:

- $Z_{1-\alpha/2}^2$ (percentil $1-\alpha/2$ de la distribución normal estándar) y $1-\alpha$ es la confianza o confiabilidad que se desea conferir a la estimación
- $V=P(100-P)$ si se trata de un porcentaje (donde P es la preestimación de dicho porcentaje que se quiere estimar) y V es la varianza de la variable (en caso de que se trate de estimar una media).
- N (tamaño de la población muestreada)
- E_0 (máximo error absoluto admisible), expresado en porcentaje para el primer caso y en las unidades de la variable empleada en el segundo

b) *Para la comparación mediante pruebas de hipótesis de las medias o porcentajes de dos poblaciones mediante estudios observacionales analíticos (casos y controles, estudios de cohorte) ó experimentales.*

Si partimos de que en cada grupo se tomarán n sujetos (suposición habitual), se tiene:

$$n = \frac{2V(Z_{1-\alpha/2}^2 + Z_{1-\beta}^2)}{\Phi^2} \quad [1.2]$$

Datos necesarios:

- α [probabilidad máxima admisible de rechazar indebidamente la hipótesis nula (probabilidad de cometer el error de tipo I)]
- β [probabilidad máxima admisible de no rechazar la hipótesis nula cuando ella es falsa (probabilidad de cometer el error de tipo II)]
- Φ (beneficio mínimo clínicamente relevante)

Si llamamos P_A y P_B a los porcentajes de “éxito” en los grupos A y B respectivamente, entonces Φ es el menor valor de $|P_A - P_B|$ que se considere relevante; si estamos en el caso de comparación de medias, Φ será la menor diferencia entre estas medias que se considere relevante.

- V es igual a la varianza de la variable que se promedia (que usualmente se considera común para ambos grupos) en caso de que se comparen medias; y que en el caso de comparación de proporciones es igual a $P_M(1 - P_M)$

$$\text{donde } P_M = \frac{P_A + P_B}{2}$$

Anexo 2. Guías de publicación correspondientes a los artículos incluidos, a las que se adhiere cada una de las seis revistas estudiadas.

Revista	Guías
AIM	CONSORT ⁶²⁻⁶⁵ STROBE ⁶⁶
BMJ	CONSORT ⁶²⁻⁶⁵ STROBE ⁶⁶
JAMA	CONSORT ⁶²⁻⁶⁵
LANCET	CONSORT ⁶²⁻⁶⁵ STROBE ⁶⁶ STREGA ⁷⁰
NEJM	CONSORT ⁶²⁻⁶⁵
PLOSM	CONSORT ⁶²⁻⁶⁵ STROBE ⁶⁶