

Impacto sobre el tamaño de las muestras en estudios nacionales si cambiara el nivel de significación estadística de $\alpha = 0,05$ a $\alpha = 0,005$

LUIS VILLARROEL DEL PINO^{1,a}

The impact on sample sizes of studies if the significance level is changed from an α of 0.05 to 0.005

Background: The statistical significance $\alpha = 0.05$ is the cut-off point used to decide whether a hypothesis is statistically significant. When p -value is less than 0.05, we reject the null hypothesis. Although this criterion has been used for almost a century to generate new knowledge, there is currently an international discussion about the need to decrease the significance to $\alpha = 0.005$. **Aim:** To determine the effects that changing the p value would have on the sample size of different types of studies. **Material and Methods:** A series of formulas for calculating the sample size of cross-sectional and comparative studies were used to create case scenarios. **Results:** By changing $\alpha = 0.05$ to $\alpha = 0.005$, the sample sizes in cross-sectional studies would double and in comparative studies would increase between 60% and 70%, depending on the statistical power chosen. **Conclusions:** Considering the sample size implications, the change in the level of significance would have important effects on the Chilean science. The cost of a randomized clinical trial could increase by at least 27% to 32%. This increase could be similar for cross-sectional studies. With an investment of less than 0.4% of gross domestic product in science and technology, national scientific research would become more expensive, distributing the few available resources among fewer projects. This effect should be considered in any discussion about national budget for science and technology.

(Rev Med Chile 2021; 149: 45-51)

Key words: Cross-Sectional Studies; Data Interpretation, Statistical; Sample Size; Uncertainty, Randomized Controlled Trial.

Cada vez que se plantea una hipótesis estadística, esta tiene asociada un nivel de significación, que es la máxima probabilidad de rechazar erróneamente la hipótesis de nulidad. Habitualmente esta probabilidad se fija en 0,05 (5%), de modo que basta con que el valor p obtenido al hacer el test estadístico sea inferior a ese nivel de significación para rechazar la hipótesis de nulidad en favor de la alternativa.

El razonamiento anterior, que ha servido para comparar tratamientos, validar test diagnósticos o identificar factores de riesgo para una enfermedad, entre otras pruebas, surge a principios del siglo XX de una manera peculiar que ha sido identificada como una de sus debilidades.

En el año 1925, el británico Ronald Fisher sentó las bases del planteamiento y resolución de pruebas de hipótesis al proponer que, para

¹Departamento de Salud Pública, Escuela de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.
^aEstadístico, PhD.

Este trabajo no recibió financiamiento. El autor declara no tener conflictos de interés.

Recibido el 4 de octubre de 2020, aceptado el 4 de diciembre de 2020.

Correspondencia a:
Luis Villarroel
Departamento de Salud Pública,
Diagonal Paraguay 362 piso 2,
Santiago, CP 8330077, Chile.
lvillard@uc.cl

determinar si dos poblaciones difieren en forma significativa, sus promedios debieran estar al menos a una distancia de 1,96 desviaciones estándar. Como la probabilidad de alejarse 1,96 desviaciones estándar del promedio en distribución Normal es igual a 0,05 (o 1/20), este criterio, siguiendo a Fisher, debiera llevarnos a rechazar erróneamente la hipótesis H_0 de igualdad de los promedios solo una vez en 20 ensayos¹.

Algunos años más tarde, Neyman y Pearson² desarrollaron una teoría para resolver hipótesis estadísticas simples (de la forma $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$), que incluía conceptos como el error tipo I (o error α), error tipo II (o error β) y potencia estadística, además de una regla para rechazar H_0 sujeto a la condición de que la probabilidad de rechazarla por error fuera menor que α .

No pasó mucho tiempo para que la regla de decisión de Neyman-Pearson y la probabilidad 0,05 sugerida por Fisher se fusionaran para convertirse en el valor $p < 0,05$, y éste se convirtiera en sinónimo de “estadísticamente significativo”. De esta forma, la sugerencia de Fisher de 1925 pasó a constituir el criterio estadístico más importante utilizado hasta nuestros días para dirimir si un hallazgo científico debe considerarse estadísticamente significativo.

Más allá de que Fisher probablemente no pensó que esa sugerencia se convertiría en la regla para la toma de decisiones científicas, el punto de corte 0,05 y el uso del valor p nunca han estado exentos de críticas, las cuales han aumentado considerablemente en los últimos años³⁻⁵. Esto podría deberse, en parte, a que estamos cerca de cumplir 100 años usando este criterio para generar nuevo conocimiento.

Entre las críticas al valor p , hay autores que plantean que este se utiliza en exceso y sugieren reemplazar su cálculo por la estimación del tamaño del efecto (diferencia de medias, medidas de riesgo, etc.) y el uso de intervalos de confianza para mostrar el grado de precisión de estas estimaciones³, mientras otros centran sus críticas en la automaticidad de las decisiones basadas en el valor p , que hace que las inferencias fluyan automáticamente a partir de los datos⁴.

Aunque hay autores que buscan fomentar buenas prácticas en el uso del valor p ⁵, es conocida la tendencia a publicar estudios con resultados estadísticamente significativos en desmedro de estudios sin hallazgos, lo que conocemos como

sesgo de publicación⁶, o la búsqueda sistemática de diferencias estadísticamente significativas en un mismo conjunto de datos sin planteamiento de hipótesis por adelantado, lo que se conoce (entre otros nombres) como *p-hacking*⁷, prácticas que hacen proliferar la publicación de falsos positivos, produciendo un efecto dañino sobre la verdadera evidencia científica^{8,9}.

Sin embargo, es el punto de corte 0,05 donde se centran las críticas más importantes en la actualidad, y el argumento que se repite con frecuencia es que el estándar $p < 0,05$ es muy poco exigente y resulta en tasas de falsos positivos muy altas (hallazgo de evidencia significativa que en realidad no existe), aun sin considerar los posibles errores no aleatorios (sesgos) propios del experimento o procedimiento seguido en el estudio¹⁰.

Aunque la sugerencia de cambiar el estándar 0,05 por uno más exigente no es algo nuevo^{10,11}, la discusión alcanzó un nuevo nivel con la publicación de una propuesta en *Nature Human Behaviour*¹² consistente en cambiar el estándar $p < 0,05$ (una probabilidad de error de 5%) por $p < 0,005$ (una probabilidad de error de 5 por mil). Más allá de la aparición de nuevas publicaciones que se oponen a este cambio¹³ o lo apoyan¹⁴, lo claro es que este cambio tendría importantes implicancias para la ciencia, ya que muchos hallazgos que con el estándar 0,05 son considerados importantes, dejarían de serlo con el nuevo estándar.

Considerando esta tendencia internacional a cuestionar el uso de $\alpha = 0,05$, un cambio de estándar en el corto o mediano plazo no solo es posible, sino que es un escenario altamente probable, por lo que se requiere de un análisis de las implicancias que esto podría tener para el desarrollo de la ciencia en el país.

El objetivo de este trabajo es estimar el efecto que tendría el cambio de nivel de significación de 0,05 a 0,005 respecto al tamaño de la muestra de los estudios de diseño más frecuente en Chile y discutir acerca de las consecuencias que podría tener este cambio sobre otros aspectos de un estudio, promoviendo una reflexión anticipatoria del impacto para la generación de conocimiento local.

Material y Método

En este trabajo se eligieron diseños de estudio de amplio uso en investigación científica nacional e internacional: i) estudios transversales, en los que

se calculan tamaños muestrales para estimar una proporción o un promedio poblacional; ii) estudios caso-control o ensayos clínicos aleatorizados, en los que se calculan tamaños muestrales para comparar dos proporciones o dos promedios independientes; iii) otros diseños con fórmulas para el cálculo de tamaños muestrales con estructura matemática similar a las usadas en (ii), que permitió identificar un patrón común en el cambio en el tamaño de la muestra, al cambiar el nivel de significación.

Para el cálculo de tamaños muestrales se usaron los programas estadísticos R¹⁵, PS Power and Sample Size¹⁶ o las fórmulas disponibles para cada situación.

Resultados

Efecto sobre el tamaño de la muestra para estimar una proporción o un promedio poblacional

En un diseño transversal, el cálculo del tamaño muestral para estimar un promedio poblacional requiere conocer la varianza de la variable en estudio (σ^2) y fijar un error de estimación (e), de modo que para un nivel de significación α , el tamaño de la muestra está dado por:

$$n \geq \frac{z_{1-\alpha/2}^2 \times \sigma^2}{e^2}$$

Donde $z_{1-\alpha/2}$ es el percentil $1-\alpha/2$ de la distribución normal estándar. En particular, para $\alpha = 0,05$ se tiene que $1-\alpha/2 = 0,975$ y el percentil a utilizar en la fórmula es $z_{0,975} = 1,96$. Si se cambia a $\alpha = 0,005$, el valor $1-\alpha/2 = 0,9975$ y el percentil es $z_{0,9975} = 2,807$.

Por otra parte, el cálculo del tamaño muestral para estimar una proporción poblacional requiere indicar el valor máximo que podría tomar esta proporción (P) y un error de estimación (e). Luego, para un nivel de significación α , el tamaño de la muestra está dado por:

$$n \geq \frac{z_{1-\alpha/2}^2 \times P \times (1 - P)}{e^2}$$

Donde $z_{1-\alpha/2}$ toma los mismos valores indicados antes: $z = 1,96$ cuando $\alpha = 0,05$ y $z = 2,807$ cuando $\alpha = 0,005$.

En las fórmulas anteriores, al disminuir $\alpha = 0,05$ a $\alpha = 0,005$, asumiendo que el resto de las condiciones se mantienen sin cambios, el tamaño de la muestra solo depende del valor $z_{1-\alpha/2}$ que se esté utilizando. Luego, si se calcula la razón en que aumenta el tamaño muestral al estimar un promedio con $\alpha = 0,005$ respecto a $\alpha = 0,05$ (tamaños que llamaremos $n_{0,005}$ y $n_{0,05}$, respectivamente), se tiene:

$$\frac{n_{0,005}}{n_{0,05}} = \frac{z_{0,9975}^2 \times \sigma^2 / e^2}{z_{0,975}^2 \times \sigma^2 / e^2} = \frac{z_{0,9975}^2}{z_{0,975}^2} = \frac{2,807^2}{1,96^2} = 2,05$$

Es decir, el tamaño muestral necesario para estimar un promedio poblacional se duplica al disminuir $\alpha = 0,05$ a $\alpha = 0,005$. Lo mismo ocurre al estimar una proporción, si el único cambio es el nivel de significación:

$$\frac{n_{0,005}}{n_{0,05}} = \frac{z_{0,9975}^2 \times P \times (1 - P) / e^2}{z_{0,975}^2 \times P \times (1 - P) / e^2} = \frac{z_{0,9975}^2}{z_{0,975}^2} = \frac{2,807^2}{1,96^2} = 2,05$$

La Tabla 1 muestra algunos ejemplos de cálculo de tamaño muestral para estimar una proporción, asumiendo diferentes errores de estimación y que la proporción en estudio no es mayor que 0,1, 0,3 y 0,5.

Por ejemplo, consideremos el estudio de prevalencia en salud más importante de nuestro país: la Encuesta Nacional de Salud, que en su versión

Tabla 1. Ejemplos de cálculo de tamaño muestral para estimar una proporción poblacional

P	Error de estimación	n ($\alpha = 0,05$)	n ($\alpha = 0,005$)
0,1	0,05	138	284
0,1	0,02	864	1.773
0,1	0,01	3.457	7.091
0,3	0,05	323	662
0,3	0,02	2.017	4.137
0,3	0,01	8.067	16.546
0,5	0,05	384	788
0,5	0,02	2.401	4.925
0,5	0,01	9.604	19.698

2016-2017 tuvo un tamaño muestral de 6.233 encuestados¹⁷, y su costo fue de \$1.482.894.125¹⁸. Con ese tamaño de muestra se pueden estimar prevalencias del orden de 1% con un margen de error inferior a 0,25 puntos porcentuales, asumiendo $\alpha = 0,05$. Si se usara $\alpha = 0,005$, se requeriría un tamaño muestral de 12.481 encuestados para tener el mismo nivel de precisión, y su costo probablemente sería superior a los 2 mil millones de pesos.

Efecto sobre el tamaño de la muestra para comparar dos proporciones o dos promedios independientes

Cuando se comparan dos promedios o dos proporciones independientes, el cálculo de tamaño muestral de cada grupo debe considerar, además del nivel de significación α , la potencia estadística $1-\beta$. En estudios nacionales la potencia usada habitualmente es 80% (es decir, $1-\beta = 0,8$), la cual es la potencia mínima aceptable en estudios científicos. Por otra parte, a nivel internacional es muy usada la potencia 90%.

Si se quiere determinar el tamaño muestral mínimo necesario para encontrar diferencias estadísticamente significativas entre dos promedios μ_1 y μ_2 , asumiendo que la varianza de la variable es igual a σ^2 , con un nivel de significación α y potencia $1-\beta$, el tamaño de cada grupo está dado por:

$$n \geq (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{2 \times \sigma^2}{(\mu_1 - \mu_2)^2}$$

Análogamente, para determinar el tamaño muestral mínimo necesario para encontrar diferencias estadísticamente significativas entre dos proporciones p_1 y p_2 (como por ejemplo, la proporción esperada de pacientes que mejorará con dos tratamientos), con un nivel de significación α y potencia $1-\beta$, el tamaño de cada grupo está dado por:

$$n \geq (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2}$$

En las dos fórmulas anteriores, si se usa significación $\alpha = 0,05$ y potencia estadística $1-\beta = 0,8$, entonces se debe usar $z_{1-\alpha/2} = 1,96$ y $z_{1-\beta} = 0,841$. Al cambiar la significación a $\alpha = 0,005$, se debe usar $z_{1-\alpha/2} = 2,807$. Luego, si el único cambio es la

significación α , el incremento del tamaño muestral con $\alpha = 0,005$ respecto a $\alpha = 0,05$ es:

$$\frac{n_{0,005}}{n_{0,05}} = \frac{(2,807 + 0,841)^2}{(1,96 + 0,841)^2} = 1,696$$

Es decir, al cambiar $\alpha = 0,05$ por $\alpha = 0,005$, asumiendo potencia $1-\beta = 0,8$, el tamaño muestral se incrementa en un 69,6%.

Si hacemos el mismo ejercicio usando potencia 90%, el valor $z_{1-\beta}$ debe cambiarse a $z_{1-\beta} = 1,282$. En este caso, el incremento del tamaño muestral es:

$$\frac{n_{0,005}}{n_{0,05}} = \frac{(2,807 + 1,282)^2}{(1,96 + 1,282)^2} = 1,591$$

Luego, al cambiar $\alpha = 0,05$ por $\alpha = 0,005$ y asumiendo potencia $1-\beta = 0,9$, el tamaño muestral se incrementa en 59,1%.

La Tabla 2 muestra algunos ejemplos del cambio en el tamaño muestral para distintos valores de p_1 y p_2 en un estudio prospectivo (como un ensayo clínico aleatorizado), para encontrar una diferencia entre estas proporciones que corresponda a un riesgo relativo igual a 2, 1,5 y 1,1 (o sea, cuando el riesgo o el beneficio aumenta 100%, 50% y 10%, respectivamente).

Se observa que en todos los cálculos para $\alpha = 0,005$ el aumento del tamaño muestral es de aproximadamente 70% respecto al tamaño con

Tabla 2. Ejemplos de cálculo de tamaño muestral para comparar dos proporciones, con potencia estadística 80%

p_1	p_2	RR	n por grupo ($\alpha = 0,05$)	n por grupo ($\alpha = 0,005$)
0,1	0,05	2,0	435	737
0,1	0,067	1,5	1.078	1.830
0,1	0,09	1,1	13.498	22.892
0,3	0,15	2,0	120	205
0,3	0,20	1,5	293	498
0,3	0,27	1,1	3.554	6.028
0,5	0,25	2,0	58	98
0,5	0,333	1,5	136	231
0,5	0,45	1,1	1.565	2.655

Tabla 3. Situaciones en las cuales el tamaño muestral aumenta entre 60% y 70%, (dependiendo de la potencia elegida) al cambiar la significancia $\alpha = 0,05$ a $\alpha = 0,005$

Situación para cálculo de tamaño muestral mínimo	Tamaño muestral mínimo
Detección de un cambio significativo de magnitud δ en una variable numérica, asumiendo que σ_δ es una estimación de la desviación estándar de este cambio ¹⁹	$(z_{1-\alpha/2} + z_{1-\beta})^2 \left(\frac{\sigma_\delta}{\delta} \right)^2 + 2$
Prueba de no inferioridad para detectar que una proporción p_2 no es inferior a una proporción p_1 en más de una cantidad δ ²⁰	$(z_{1-\alpha/2} + z_{1-\beta})^2 \times \frac{p_1 \times (1 - p_1) + p_2 \times (1 - p_2)}{(p_1 - p_2 - \delta)^2}$
Prueba de no inferioridad para detectar que un promedio μ_2 no es inferior a un promedio μ_1 en más de una cantidad δ , asumiendo que la varianza de la variable en estudio es σ^2 ²⁰	$(z_{1-\alpha/2} + z_{1-\beta})^2 \times \frac{2 \times \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
Detectar si una correlación de Pearson de magnitud r_1 es distinta de una correlación de magnitud r_0 (habitualmente $r_0 = 0$) ²¹	$\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{ z_{r_0}^* - z_{r_1}^* } \right)^2 + 3$
Detectar si un coeficiente Alfa de Cronbach de magnitud ρ_1 es distinto de un valor referencial ρ_0 en una encuesta compuesta de k ítems en escala de Likert ²²	$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \times 2k/(k-1)}{\ln\left(\frac{1 - \rho_0}{1 - \rho_1}\right)} + 2$
Detectar como significativa una pendiente β^* en un modelo de regresión logística simple, asumiendo que la tasa esperada del evento para el promedio de X es igual a P_M ²³	$\frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{P_M(1 - P_M) \times \beta^{*2}}$

$\alpha = 0,05$, que es lo que se calculó teóricamente. Si el diseño es un ensayo clínico aleatorizado, los tamaños muestrales deben considerar además las pérdidas de seguimiento esperadas.

Efecto sobre el tamaño de la muestra para detectar como significativos otros estimadores y tamaños de efecto

Existen varias situaciones donde el cálculo del tamaño muestral depende de la expresión $(z_{1-\alpha/2} + z_{1-\beta})^2$, como en el punto anterior, asumiendo que el resto de los parámetros necesarios para el cálculo se mantienen sin cambios.

La Tabla 3 muestra las situaciones más importantes en este grupo de diseños.

En todas las situaciones anteriores, al cambiar la significación $\alpha = 0,05$ a $\alpha = 0,005$, el tamaño muestral aumenta entre 60% y 70%, dependiendo de la potencia elegida (90% u 80% respectivamente).

Discusión

El valor p , y la regla de decisión de rechazar una hipótesis H_0 si el valor p es menor que $\alpha = 0,05$, son los instrumentos más utilizados en ciencia para decidir si los datos de un estudio son compatibles con esa hipótesis H_0 . Aunque han surgido propuestas como dejar de referirse a “resultados estadísticamente significativos” en términos dicotómicos²⁴, reemplazar el cálculo del valor p por la estimación de tamaños del efecto, o usar intervalos de confianza para mostrar el grado de precisión de las estimaciones, la creciente presión sobre el valor $\alpha = 0,05$, hace que no sea impensable que este límite cambie en el futuro.

En este trabajo se ha mostrado que un cambio en el nivel de significación estadística de $\alpha = 0,05$ a $\alpha = 0,005$, produciría un aumento sustancial en el tamaño de la muestra de varios tipos de diseño, siendo los nuevos tamaños de muestra entre 50%

y 100% mayores que los obtenidos con la significación actual.

Respecto al costo económico que tendría el aumento del tamaño muestral, existen escasas publicaciones que permitan hacer estimaciones al respecto. Sin embargo, es claro que el costo del estudio no aumentaría en la misma proporción que el tamaño muestral, ya que no todos los costos dependen del número de pacientes²⁵.

En un ensayo clínico aleatorizado, los costos relacionados con los pacientes pueden representar 45,5% del costo total del estudio²⁶. Luego, si el cambio de significación aumenta el tamaño muestral de un estudio comparativo entre 60% y 70%, el costo total del estudio podría aumentar entre 27% y 32% como mínimo. En un estudio descriptivo este aumento podría ser más acotado, asumiendo menores costos asociados a los participantes de investigación.

En un país como Chile, que invierte menos de 0,4% del producto interno bruto en ciencia y tecnología²⁷, mientras los países de la OECD, al año 2018, invertían en promedio 2,4% del PIB²⁸, un cambio de nivel de significación estadística sería una mala noticia para la generación de ciencia local, ya que este “simple cambio” haría que la investigación científica chilena se encarezca, distribuyendo los pocos recursos disponibles entre menos proyectos. Esto es complejo, en un ecosistema de financiamiento nacional de la ciencia altamente competitivo, con tasas de adjudicación de concursos como el Fondo Nacional de Investigación en Salud (FONIS) o el Fondo Nacional de Desarrollo Científico y Tecnológico que no superan 12 y 27% respectivamente en el último quinquenio²⁹.

Este cambio del valor p no solo afectaría el número probable de estudios financiados por la Agencia Nacional de Investigación (ANID), sino que también podría impactar en la difusión científica de los hallazgos nacionales. Las revistas científicas de mayor impacto exigirían desde un inicio ajustarse al nuevo estándar, la investigación nacional de vanguardia se vería obligada a sumarse con rapidez al nuevo criterio, porque de otra forma los estudios tendrían que divulgarse solo a nivel local, poniéndonos en desventaja respecto a otros países con igual nivel de desarrollo económico. Es decir, este cambio sería algo que no se podría evitar ni postergar.

En consecuencia, los académicos e investi-

gadores nacionales deben mantenerse en activa observación de este fenómeno internacional que promueve el cambio del nivel de significación estadística. Si en algún momento se logra un consenso sobre este cambio, habría efectos importantes sobre la ciencia chilena. Este sería un elemento más a considerar en el necesario aumento en la inversión en ciencia y tecnología que requiere el país, basando estas decisiones políticas en evidencia científica que muestre el impacto de mantener un presupuesto austero: no hay progreso sin ciencia.

Referencias

1. Fisher R. *Statistical methods for research workers*. Edinburgh, England: Oliver and Boyd; 1925. p 44-5.
2. Neyman J, Pearson ES. The testing of statistical hypotheses in relation to probabilities a priori. *Math Proceed* 1933.
3. Rothman K. Six Persistent Research Misconceptions. *J Gen Intern Med* 2014; 29 (7): 1060-4. DOI: 10.1007/s11606-013-2755-z
4. Goodman S. Of P-Values and Bayes: A Modest Proposal. *Epidemiology May* 2001; 12 (3): 295-7.
5. Wasserstein R, Lazar N. The ASA's Statement on p-Values: Context, Process and Purpose. *The American Statistician* 2016; 70 (2): 129-33.
6. Cleophas T, Zwinderman A. *Modern Meta-Analysis - Review and Update of Methodologies*. Switzerland: Springer 2017.
7. Simonsohn U, Nelson L, Simmons J. P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General* 2014; 143 (2): 534-47.
8. Nuzzo R. Scientific method: Statistical errors 2014; 506 (7487): 150-2. doi: 10.1038/506150a.
9. Head M, Holman L, Lanfear R, Khan A, Jennions M. The extent and consequences of P-Hacking in Science. *PLoS Biol* 2016; 13 (3): e1002106. doi:10.1371/journal.pbio.1002106
10. Johnson V. Revised standards for statistical evidence. *Proc. Natl Acad. Sci. USA* 2013; 110: 19313-7.
11. Greenwald A, Gonzalez R, Harris R, Guthrie D. Effect sizes and p values: What should be reported and what should be replicated? *Psychophysiology* 1996; 33 (2): 175-83.
12. Benjamin D, Berger J, Johannesson M, Nosek B, Wagenmakers E, Berk R, et al. Redefine statistical significance. *Nat Hum Behav* 2018; 2 (1): 6-10. doi:10.1038/s41562-017-0189-z
13. Lakens, D, Adolff F, Albers, C, Anvari F, Apps M, Arga-

- mon S, et al. Justify your alpha. *Nat Hum Behav* 2018; 2 (3): 168-71 doi:10.1038/s41562-018-0311-x
14. Wasserstein R, Schirm A, Lazar N. Moving to a World Beyond “ $p < 0.05$ ”. *The American Statistician* 2019; 73(1): 1-19; Editorial.
 15. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 16. Dupont WD, Plummer WD. "Power and Sample Size Calculations for Studies Involving Linear Regression", *Controlled Clinical Trials* 1998; 19: 589-601.
 17. Encuesta Nacional de Salud 2016-2017. Departamento de Epidemiología, Ministerio de Salud de Chile. Página: <http://epi.minsal.cl/encuesta-ens/> (consultado el 1 de octubre de 2020).
 18. http://www.minsal.cl/wp-content/uploads/2018/01/2-Resultados-ENS_MINSAL_31_01_2018.pdf (consultado el 1 de octubre de 2020).
 19. Dell R, Holleran S, Ramakrishnan R. Sample Size Determination. *ILAR J* 2002; 43 (4): 207-13.
 20. Blackwelder WC. "Proving the Null Hypothesis" in *Clinical Trials*. *Control. Clin. Trials* 1982; 3: 345-53.
 21. Sokal RR, Rohlf FJ, Biometry WH. Freeman and Company, New York, 1995, p. 578.
 22. Bonet D. Sample Size Requirements for Testing and Estimating Coefficient Alpha. *J Educ and Behav Stat* 2002; 27 (4): 335-40.
 23. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Stat in Med* 1998; 7 (14): 1623-34.
 24. Amrhein V, Greenland S, McShane B, Aaby P, Aagaard K, Aavitsland P. Retire statistical significance. *Nature* 2019; 567: 305-7.
 25. Van Asselt T, Ramaekers B, Ramos I, Joore M, Al M, Lesman-Leegte I, et al. Research costs investigated: a study into the budgets of Dutch publicly funded drug-related research. *Pharmacoeconomics* 2018; 36: 105-13.
 26. Sertkaya A, Wong H, Jessup A, Baleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clinical Trials* 2016; 13 (2): 117-16.
 27. Balbontín R, Roeschmann J, Zahler A. Ciencia, Tecnología e Innovación en Chile: un análisis presupuestario. Dirección de Presupuestos. Gobierno de Chile. Enero 2018.
 28. OECD Data. Gross domestic spending on R&D. <https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm> (consultado el 1 de octubre de 2020).
 29. Agencia Nacional de Investigación en Salud ANID. (2020). Compendio estadístico 1025-2019. https://s3.amazonaws.com/documentos.anid.cl/compendio/Compendio_Estadistico.pdf (consultado el 1 de octubre de 2020).