

Conceptos generales en bioestadística y epidemiología clínica: error aleatorio y error sistemático

General concepts in biostatistics and clinical epidemiology: Random error and systematic error

Felipe Barraza^a, Marcelo Arancibia^{a,b}, Eva Madrid^{a,b}, Cristian Papuzinski^{a,b,*}

^a Cátedra de Metodología de la Investigación Científica, Escuela de Medicina, Universidad de Valparaíso, Valparaíso, Chile

^b Centro Interdisciplinario de Estudios en Salud (CIESAL), Universidad de Valparaíso, Valparaíso, Chile

*Autor correspondiente cristian.papuzinski@uv.cl

Citación Barraza F, Arancibia M, Madrid E, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Random error and systematic error. *Medwave* 2019;19(7):e7687

Doi 10.5867/medwave.2019.07.7687

Fecha de envío 27/6/2019

Fecha de aceptación 5/8/2019

Fecha de publicación 23/8/2019

Origen Este artículo es parte de una colección de "Notas metodológicas" acordada entre Medwave y la Cátedra de Metodología de la Investigación Científica de la Escuela de Medicina de la Universidad de Valparaíso.

Tipo de revisión Con revisión por pares externa con cuatro revisores, a doble ciego

Palabras clave bias, probability, confidence interval, epidemiology, biostatistics

Resumen

La investigación biomédica, particularmente la que involucra a seres humanos, está siempre sometida a fuentes de error que deben ser reconocidas. El error sistemático o sesgo, se asocia con debilidades en el diseño metodológico o de la fase de ejecución del estudio. Éste afecta su validez y se valora cualitativamente. Por su parte, el error aleatorio se relaciona con las variaciones producidas por el azar, el cual puede expresarse cuantitativamente, pero nunca eliminarse. Esta revisión es la primera entrega de una serie metodológica sobre conceptos generales en bioestadística y epidemiología clínica desarrollada por la Cátedra de Metodología de la Investigación Científica de la Universidad de Valparaíso, Chile. En este artículo se abordan los conceptos teóricos asociados al error, su evaluación y control. Finalmente, se discuten algunas controversias actuales en cuanto a su conceptualización, de relevancia para estudiantes de pre y posgrado de ciencias de la salud.

Abstract

Biomedical research, particularly when it involves human beings, is always subjected to sources of error that must be recognized. Systematic error or bias is associated with problems in the methodological design or during the execution phase of a research project. It affects its validity and is qualitatively appraised. On the other hand, random error is related to variations due to chance. It may be quantitatively expressed, but never removed. This review is the first of a methodological series on general concepts in biostatistics and clinical epidemiology developed by the Chair of Scientific Research Methodology at the School of Medicine, University of Valparaíso, Chile. In this article, we address the theoretical concepts of error, its evaluation, and control. Finally, we discuss some current controversies in its conceptualization that are relevant to undergraduate and graduate students of health sciences.

Ideas claves

- El error es inherente a la investigación biomédica.
- El error sistemático o sesgo se asocia con debilidades del diseño metodológico o la ejecución del estudio, afectando su validez. Puede valorarse cualitativamente y ser evitado.
- El error aleatorio se relaciona con variaciones debidas al azar y compromete la confiabilidad de la investigación. Puede expresarse cuantitativamente con el valor p y los intervalos de confianza. No puede eliminarse, pero sí controlarse mediante el aumento del tamaño muestral y un análisis estadístico eficiente.
- Las conclusiones de un estudio deben interpretarse teniendo en cuenta el efecto del error y el contexto investigativo.

Introducción

La investigación biomédica, sobre todo aquella realizada en seres humanos, está constantemente sometida a errores debido a las características propias de su objeto de estudio, como también a limitaciones prácticas y bioéticas. La evaluación del error es fundamental en el análisis de los datos, pero mayormente durante el diseño del estudio, lo que permitiría anticiparse a la ocurrencia de errores sistemáticos. Por otra parte, aunque la teoría de probabilidades nos permite estimar el efecto del azar en el resultado de una medición, el error aleatorio puede afectar la presunta representatividad de una muestra con respecto a la población de la cual proviene, aportando incertidumbre e imprecisión a la estimación del verdadero valor poblacional. No obstante, un requerimiento de un proceso de medición, mas no de su resultado, es que sea exacto. La exactitud involucra que este proceso sea válido (libre de sesgos), en donde la diferencia entre el estimador y el verdadero valor del parámetro poblacional es baja, y confiable; es decir, reproducible, consistente o preciso, generando datos con una escasa variabilidad entre mediciones sucesivas^{1,2,3}.

El error aleatorio (que afecta la confiabilidad) y el error sistemático (que afecta la validez), son dos de los principales elementos evaluados durante el desarrollo de la investigación científica y la posterior evaluación crítica de los lectores del artículo publicado. Debido a que se asume, desde un punto de vista complejo, que los fenómenos estudiados son siempre multicausales y multivariados, el considerar como verdadera una asociación y, aun más, declararla como “causal”, implica la conjunción de resultados desde distintas disciplinas y siempre requiere de la integración de su contexto.

Este artículo corresponde a la primera entrega de una serie metodológica de seis revisiones narrativas acerca de tópicos generales en bioestadística y epidemiología clínica, las que explorarán publicaciones disponibles en las principales bases de datos y textos de consulta especializados. Esta serie está orientada a la formación de estudiantes de pre y posgrado. Es realizada por la Cátedra de Metodología de la Investigación Científica de la Escuela de Medicina de la Universidad de Valparaíso, Chile. En esta revisión se analizan los distintos elementos teóricos y prácticos asociados al error en investigación biomédica, haciendo hincapié en su evaluación y su control. Finalmente, se revisan las perspectivas actuales en cuanto a su teoría y algunas controversias respecto a su conceptualización.

Conceptos preliminares

Una hipótesis implica una explicación tentativa para una pregunta de investigación. En el caso de la hipótesis estadística, la construcción se basa en dos supuestos: la hipótesis nula (H_0) y la hipótesis alternativa o alterna (H_1). H_0 supone que no existen diferencias entre los fenómenos en el mundo y que las inferencias inductivas (generalización) sobre sus relaciones están equivocadas; es decir, que los factores de exposición y de desenlace no se relacionan entre sí. Se trata de una hipótesis conservadora que se plantea en contraposición a H_1 , la hipótesis del investigador o de trabajo, que afirma que las asociaciones observadas entre distintas observaciones no se explican por el azar⁴.

La declaración de una hipótesis nula y una alternativa es esencial en la estadística inferencial, en donde se aplican pruebas de contraste de hipótesis que buscan encontrar evidencia suficiente para rechazar la hipótesis nula y darle cabida a la hipótesis investigada. No obstante, siempre debe tenerse en mente que el resultado de una prueba de hipótesis es solo un elemento más en la toma de decisiones⁵ (Ejemplo 1⁶).

Ejemplo 1. Se ha estudiado la relación entre el consumo de chocolate y el funcionamiento cognitivo. En este contexto, Messerli plantea que los países que tienen un mayor consumo de chocolate presentan una mayor cantidad de galardonados con el Premio Nobel, pues el chocolate podría asociarse al rendimiento cognitivo. En este caso, la hipótesis estadística del investigador (H_1) sería que el consumo de chocolate se correlaciona con la obtención de un Premio Nobel. Por lo tanto, la hipótesis nula (H_0) sería que el consumo de chocolate no se correlaciona con la obtención de un Premio Nobel.

Error sistemático (sesgo)

El error sistemático o sesgo, puede entenderse como la tendencia sistemática a subestimar o sobrestimar el estimador de interés a causa de una deficiencia en el diseño o en la ejecución de un estudio⁷. Ello atenta contra la validez de este, la que puede ser interna, entendida como el grado de concordancia que existe entre los resultados del estudio y el valor real del parámetro en la población, o externa, o grado en que los resultados del estudio de una muestra pueden ex-

trapolarse a otras poblaciones². Los sesgos pueden asociarse a cualquier fase de la ejecución de una investigación, por lo que tienden a desviar los resultados de la verdad en un mismo sentido².

Existen sesgos que generan una sobreestimación de la magnitud de asociación entre variables, conocidos como sesgos positivos (“en contra” de la hipótesis nula); mientras que aquellos que aminoran la magnitud se conocen como sesgos negativos (“a favor” de la hipótesis nula). En el caso extremo, un sesgo puede desencadenar una inversión en el sentido de asociación, haciendo que un factor protector aparezca como un factor de riesgo, lo que se denomina *switch-over bias*^{3,8}.

Cuando se trata de investigación con seres humanos, la disciplina encargada de controlar el error sistemático es la epidemiología, mediante la utilización de los diseños metodológicos adecuados y de las estrategias de recolección de datos apropiadas². Si bien existen sesgos en múltiples categorías (<https://catalogofbias.org/>)^{9,10}, se agrupan fundamentalmente en tres grandes tipos: de selección, de medición (o sesgo de información) y de confusión⁴. El primero, ocurre debido a que la relación entre la exposición y el desenlace difiere entre los distintos grupos de participantes del estudio (diferencias sistemáticas entre las características de los participantes)⁸ (Ejemplo 2).

Ejemplo 2. Algunas investigaciones han señalado al consumo de carne como factor de riesgo para el desarrollo de cáncer gástrico. Para analizarlo, se diseña un estudio prospectivo de cohorte para comparar la supervivencia a cinco años de cáncer gástrico desde su diagnóstico entre consumidores de carne (grupo A) y no consumidores de carne (grupo B). El grupo A proviene de un país en donde no existe una pesquisa sistemática de la neoplasia ni tampoco un sistema de salud que la regule. El grupo B se selecciona desde un país en donde se realizan endoscopías digestivas de rutina, debido a que el sistema de salud reconoce una alta incidencia en la región. Se concluye que el grupo B presenta una supervivencia significativamente mayor. Sin embargo, es probable que la mayor supervivencia del grupo B sea explicada por un diagnóstico temprano y un tratamiento precoz y no por no consumir carne. En este caso, las muestras se han seleccionado de manera sesgada, ya que sus características basales son diferentes (oportunidades en salud), cuando idealmente solo deberían haber diferido en cuanto a si consumían carne o no.

En cuanto a la medición, existen fundamentalmente tres fuentes de error: el fenómeno medido (por ejemplo, sesgo de memoria por el recuerdo diferencial a la exposición en un estudio de casos y controles), el instrumento de medición (por ejemplo, cambios en los criterios diagnósticos para evaluar una patología a lo largo del tiempo) y el observador quien realiza la medición¹. Debido a su complejidad y ubicuidad, nos referiremos más extensamente al sesgo de confusión. De manera específica, los tipos de sesgos serán tratados en los siguientes artículos de esta serie metodológica en función de los diseños metodológicos a los cuales comprometen.

Una variable de confusión es aquella que se asocia tanto a la variable de exposición (sin ser un resultado de ella) como a la de desenlace (es decir, independientemente de su asociación con la exposición de interés, en sujetos no expuestos debe estar asociada al desenlace),

pero que no se encuentra en la vía causal de la asociación^{11,12}. Esto “confunde” la asociación aparente entre exposición y desenlace¹. Las variables de confusión no deben confundirse con las variables de interacción que operan como “modificadores del efecto”, las que interactúan con la variable de exposición modificando la magnitud de sus efectos sobre el desenlace, pero no son la causa del desenlace en sí mismo¹.

El sesgo de confusión es complejo, prominente y multifactorial, por lo que siempre tiene que estar en la mente del investigador⁴. Puede prevenirse a nivel del diseño de estudio (por ejemplo, aleatorización en ensayos clínicos aleatorizados) o controlarse durante el análisis de los datos (por ejemplo, análisis estratificado, modelos de regresión estadística)⁴. En el Ejemplo 3 se demuestra el efecto del sesgo de confusión en un estudio observacional realizado en Noruega por Strand y colaboradores^{12,13}.

Ejemplo 3. Se desarrolló un estudio de cohorte que comparó 849 niños con parálisis cerebral con 615.668 niños sin la patología, concluyendo que la *Odds ratio* de haber tenido una madre con preeclampsia fue de 2,5 con un intervalo de confianza al 95%: 2,0 a 3,2; para quienes tenían parálisis cerebral. Es decir, tenían una chance 2,5 veces mayor de haber presentado la condición en el caso de que su madre hubiese padecido la enfermedad. Sin embargo, esta *Odds ratio* fue levemente atenuada cuando en el análisis estadístico la asociación se ajustó por la variable “niños que fueron pequeños para la edad gestacional”, resultando en 2,1, con un intervalo de confianza de 95%: 1,7 a 2,7. Incluso, ajustes adicionales que consideraron la variable “recién nacido de pretérmino” invirtieron la asociación a favor de la preeclampsia, demostrando que podría ser un factor protector para el desarrollo de parálisis cerebral en el caso de que los niños nacieran antes de las 32 semanas y no fueran pequeños para la edad gestacional, con una *Odds ratio* de 0,5 con un intervalo de confianza al 95%: 0,5 a 0,8. En este ejemplo se señala un estudio observacional, vale decir, las variables son observadas y no controladas por el investigador, por lo que la probabilidad de incurrir en sesgos es mayor. Los autores controlaron el sesgo de confusión a nivel del análisis estadístico mediante un modelo de regresión estadística conocido como regresión logística, utilizado para evaluar variables de confusión, que de otra manera pudieron no haber sido halladas.

Error aleatorio (azar)

El error aleatorio se asocia a las variaciones explicadas por el azar que está inherentemente involucrado en cada proceso de investigación, por lo que no puede eliminarse. Esto significa que influye en los resultados incluso cuando se han controlado debidamente los sesgos⁷ y compromete la confiabilidad de la investigación. Los factores que se asocian al error por azar en los resultados son esencialmente tres¹⁴: el grado de variabilidad individual e interindividual, el tamaño muestral y la magnitud de las diferencias (a medida que la diferencia hallada en la comparación sea mayor, la probabilidad de que ésta se deba al azar será menor).

El error aleatorio produce observaciones desviadas del verdadero valor en cualquier sentido². Es impredecible, pero puede disminuirse al incrementar el tamaño muestral y al realizar un análisis estadístico

eficiente¹⁴. Ello implica que la estadística controla el error aleatorio² indicando la probabilidad de que ocurra el azar⁷. En función de esto, una adecuada estimación del tamaño muestral contrarrestará el efecto del azar en el estudio. A propósito de esta idea, cabe indicar que el tamaño muestral de un estudio no sería indicador de la validez interna del mismo. O sea, no se asocia directamente al nivel de sesgo que tenga la investigación¹⁴. La estimación del error aleatorio se realiza mediante dos procedimientos: las pruebas de contraste de hipótesis (valor p) y los intervalos de confianza¹⁵.

Valor p

El valor p (de “probabilidad”) puede definirse como la probabilidad de observar los datos observados asumiendo que la hipótesis nula es verdadera. Esto es, la probabilidad de que ocurra el azar en el caso de que la hipótesis nula es cierta, vale decir, asumiendo que los fenómenos en estudio no se relacionan⁴. Puede responder a la pregunta ¿cuál es la probabilidad de haber encontrado esa asociación solo por azar? El valor p es una medida de la fuerza de la evidencia contra la hipótesis nula¹⁴, ya que puede comprenderse como la probabilidad de obtener ciertos resultados dado que se cumple la hipótesis que “se quiere” rechazar. Por lo tanto, intuitivamente, si ese valor p es muy pequeño, se rechaza la hipótesis nula y se logra lo buscado¹⁶.

Tradicionalmente, su valor ha sido fijado en 0,05, lo que implica que al aplicar métodos estadísticos que analicen la probabilidad de ocurrencia de las asociaciones, éstas hayan sucedido debido al azar no más de una de cada veinte ocasiones o el 5% de las veces o menos⁴; de otro modo, conlleva que la hipótesis nula será rechazada por error el 5% de las veces. No obstante, no existe un razonamiento científico detrás del valor 0,05 en sí mismo, sino que más bien se trata de una convención arbitraria¹⁷. A este punto de corte se le ha denominado significancia estadística, vale decir, al valor en el que la H_0 puede ser rechazada. Esto no estriba que algo necesariamente importante (“significativo”), ha sucedido, sino que debe comprenderse como un cálculo que mostraba algo o que “significaba” algo^{4,18}. En esta línea, algunas iniciativas actuales han propuesto disminuir el umbral del nivel de significancia de 0,05 a 0,005¹⁹.

Existen distintas pruebas de hipótesis que luego se vincularán a un valor p. La elección de cada una de ellas se realizará en función del diseño utilizado y de las variables aleatorias en estudio, pero presentan en común el ser una función de la diferencia entre los valores observados en el estudio y aquellos que se observarían asumiendo que la H_0 es verdadera, considerando la variabilidad de la muestra¹⁵. Otra manera de representarlos es como una fracción, cuyo denominador (variabilidad del resultado) disminuye a medida que el tamaño muestral aumenta, y un numerador, que se incrementa cuando la diferencia entre los valores observados y esperados es mayor¹⁴.

A partir de lo anteriormente expuesto, podemos señalar dos tipos de error asociados al azar. El primero de ellos es el error tipo I, que se conceptualiza como la probabilidad de rechazar H_0 dado que H_0 es verdadera. Esto es cuando, pese a que no existe en realidad una asociación entre las variables estudiadas, el estudio muestra que sí la hay. Así, la antes señalada significancia estadística constituye el límite del error tipo I, cuyo valor numérico se denomina α ²⁰. Este tipo de error

se encuentra más frecuentemente en estudios clínicos que buscan analizar un gran número de asociaciones simultáneamente. Algunos ejemplos son un estudio de cohorte que analice múltiples variables para una misma exposición, un ensayo clínico con diversos análisis de subgrupo o un estudio de casos y controles que explore un sinnúmero de factores de riesgo conjuntamente²⁰.

Por otra parte, en el caso de que exista una asociación en la realidad y esta diferencia no sea pesquisada por el estudio, se incurre en un error tipo II. Esto es, la probabilidad de no rechazar H_0 dado que H_0 es falsa. Al valor del error tipo II se le simboliza con β . El complemento de β ($1-\beta$) corresponde al poder del estudio o potencia estadística, vale decir, la probabilidad de encontrar una diferencia si ésta realmente existe, o dicho de otro modo, comprobar la hipótesis del investigador. Comúnmente, el poder del estudio se establece entre 0,8 y 0,9, lo que significaría que el estudio tiene entre 80 y 90% de probabilidades de detectar la diferencia propuesta y que el resultado alcance significación estadística²⁰. En el Ejemplo 4 se indica la interpretación del valor p en virtud de un estudio de biomarcadores en trastornos mentales severos²¹.

Ejemplo 4. Diversas investigaciones han reportado un nivel aumentado de citoquinas proinflamatorias en trastornos psicóticos y trastornos afectivos severos, así como también en personas que han sufrido traumas psicológicos durante la infancia. Al respecto, se estudiaron marcadores de inflamación y antecedentes de trauma infantil en personas con esquizofrenia, trastorno esquizoafectivo, trastorno bipolar con fases psicóticas y personas sanas. Los autores concluyeron que las personas con trastornos esquizofrénicos tenían niveles significativamente mayores de interleucina-6, factor de necrosis tumoral α y proteína C reactiva (citoquinas proinflamatorias) al compararse con personas sanas, así como niveles significativamente mayores de factor de necrosis tumoral α que las personas con trastornos bipolares (todas las comparaciones resultaron en un valor $p < 0,05$). Asimismo, se encontró que la exposición a abuso sexual infantil tuvo una asociación significativa ($p = 0,018$) con los niveles de proteína C reactiva en las personas con esquizofrenia. En este estudio, los autores establecieron un nivel de significancia del 5% (0,05). Es decir, si al aplicar una prueba estadística una asociación presenta un valor p menor a este umbral, la hipótesis nula (H_0) puede rechazarse prevaleciendo la hipótesis del investigador o alternativa (H_1), que en este caso es que los marcadores inflamatorios se asocian a trastornos mentales severos y a la presencia de trauma infantil. Así, se comprobó que existen asociaciones significativas entre los parámetros estudiados. Esto es, que con una probabilidad del 95% las asociaciones observadas no se deben al azar, sino que se explicarían por un mecanismo subyacente distinto.

El que el valor p dependa del tamaño de muestra, de modo que las muestras pequeñas son más vulnerables al error aleatorio que las de mayor tamaño, ha sido objeto de extenso debate. Aquí reside la crítica más profunda formulada contra el uso de las pruebas de hipótesis. Ella señala que el rechazo o no de una hipótesis depende del tamaño muestral²², puesto que si se evalúa una reducida parte de la realidad las conclusiones serán también restringidas, pero si se evalúa

una parte extensa, el rechazo de la hipótesis nula quedaría virtualmente asegurado²². No obstante, existen autores que se oponen a este juicio, indicando que, al estudiar las discrepancias en los resultados entre ensayos clínicos aleatorizados con tamaños de muestra grandes y pequeños, las diferencias halladas no se explican por su tamaño sino por el control de sesgos, sobre todo el de confusión. Por esta razón, la regla general sería mayoritariamente la concordancia entre los resultados y no su diferencia^{14,23}. Debido a la limitada cantidad de información que puede llegar a aportar el valor p por sí solo, otra forma de cuantificar el azar son los intervalos de confianza.

Intervalos de confianza

Consisten en un rango de valores dentro del cual se encuentra el valor real del parámetro con determinada probabilidad⁷, por lo que reflejan el grado de incertidumbre. Como ya ha sido referido, una muestra de mayor tamaño presentará un intervalo de confianza más preciso en cuanto a la estimación de un parámetro poblacional. Esto es, un rango de valores más estrecho, lo que indica un menor efecto del azar sobre la estimación. Así como el valor p, la estimación de los intervalos de confianza da lugar a la inferencia estadística, pues excluye un valor crítico en el intervalo que indica la falta de asociación entre dos variables. En el caso de los indicadores cuya fórmula es un cociente (por ejemplo, riesgo relativo y *Odds ratio*), este valor corresponde a 1, mientras que cuando lo expresado corresponde a una resta entre el riesgo de un grupo y de otro (por ejemplo, reducción del riesgo absoluto), el valor es 0. En ambos casos, estos valores representarían los puntos en que el evento es igualmente probable en ambos grupos¹⁸. A ciencia cierta, no podría asegurarse que los valores que excedan los límites del intervalo de confianza están del todo excluidos, pero sería razonable pensar que es altamente improbable hallar el valor real del parámetro más allá de estos límites²⁴. En el Ejemplo 5 se señala la interpretación del intervalo de confianza a partir de los resultados del estudio de Strand y colaboradores citados en el Ejemplo 3¹³.

Ejemplo 5. En un comienzo, los investigadores hallaron que el haber tenido una madre con preeclampsia aumentaba 2,5 veces la chance de presentar parálisis cerebral, asumiendo que esta asociación era significativa, pues su intervalo de confianza al 95% era de 2,0 a 3,2. Es decir, su rango no incluía al valor 1, el que indica que no existe asociación entre las variables estudiadas. Este detalle es importante, pues la *Odds ratio* se calcula mediante un cociente. Análisis estadísticos posteriores demostraron que la preeclampsia sería un factor protector para el desarrollo de parálisis cerebral en niños nacidos antes de las 32 semanas que no fueran pequeños para la edad gestacional, ya que la *Odds ratio* fue de 0,5 y su intervalo de confianza de 0,5 a 0,8, cuya interpretación es la misma: con una probabilidad de 95%, la asociación entre madres con preeclampsia e hijos sin parálisis cerebral se explica por un mecanismo subyacente distinto del azar.

Perspectivas y consideraciones finales

El hecho de considerar a una asociación como “verdadera”, es decir, no explicada ni por un sesgo ni por el azar, implica pensar en la causalidad²⁵, integrando lo que se sabe hasta el minuto acerca del mecanismo subyacente al fenómeno. Esto nos llama a pensar en ideas y no solo en interpretar asociaciones estadísticas de manera superficial, lo que conlleva avanzar en el nivel de complejidad comprensivo y asumir algunas problemáticas que se puntualizan a continuación.

Para extrapolar el resultado encontrado en la muestra a un nivel poblacional, es decir, para generalizar una verdad, las conclusiones no solo deben basarse en un procedimiento estadístico o en el nivel de representatividad de la muestra respecto de la población, sino que deben incorporar el conocimiento existente en torno al fenómeno en estudio²⁶. La causalidad debe estudiarse considerando los hallazgos previos de otros estudios en el ámbito. Sin embargo, la posibilidad de integrar estos hallazgos en el propio análisis no puede medirse de una forma precisa.

El proceso analítico asociado a la inferencia estadística mediante pruebas de hipótesis carece de la incorporación de algunos factores de gran relieve como la plausibilidad biológica y el cuerpo de la evidencia ya existente. Frente a esto han surgido distintos enfoques, entre ellos los métodos bayesianos²⁷, los que integran las experiencias previas en el proceso inferencial, ya que se asume que la experiencia acumulada sobre una hipótesis puede y debe contribuir a su verificación¹⁶. Aquí, el investigador expresa probabilísticamente sus puntos de vista apriorísticos y estos son añadidos al análisis formal de los datos²⁸.

Pese a que el enfoque bayesiano aún no es utilizado ampliamente por la investigación biomédica¹⁶, existe ingente investigación al respecto, lo que proyecta positivamente su aplicación. En el intertanto, y aunque provienen desde la misma plataforma teórica (el enfoque frecuentista) que el valor p, desde hace por lo menos tres décadas la publicación científica promueve el que el valor p siempre sea acompañado por intervalos de confianza^{29,30}, que si bien se sostienen sobre la misma base estadística que el valor p, otorgan adicionalmente una noción de la diferencia entre los efectos y de la incerteza en la medición²².

La lectura y la escritura de artículos científicos se ha centrado en el valor p que, como ha sido mencionado anteriormente, es un mecanismo cuantitativo de evaluación del azar. En efecto, de acuerdo con algunos autores, el quehacer científico se ha centrado en la casualidad⁴. Esto ha ido en desmedro de la lectura y la escritura crítica orientada a la evaluación de los sesgos, cuya valoración es cualitativa. Por lo tanto, teniendo una base teórica que, por una parte entregue luces respecto de los sesgos en que con mayor probabilidad puede incurrirse en un diseño metodológico, y que por otra sitúe el hallazgo en lo ya conocido sobre el fenómeno; cabe preguntarse, ¿en qué sentido y en qué magnitud pudo haber afectado el sesgo? ¿Puedo creer en lo que veo? Los números no se sostienen sobre sí mismos, sino que son en virtud de su circunstancia.

Figura 1. Esquema de síntesis del error aleatorio y sistemático.



Referencias

1. Ruiz J. Variables y fuentes de error en la medición. Confusión, sesgo y azar. In: Ruiz Á, Gómez-Restrepo C, editors. *Epidemiología clínica: investigación clínica aplicada*. 2nd ed. Bogotá: Editorial Médica Panamericana; 2015. | Link |
2. Martínez-González M, Toledo J, López-Fidalgo J. Intervalos de confianza y contraste de hipótesis. In: Martínez M, Sánchez-Villegas A, Toledo E, Faulin J, editors. *Bioestadística amigable*. 3rd ed. Barcelona: Elsevier; 2014. | Link |
3. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004 Aug;58(8):635-41. | PubMed |
4. Ghaemi S. Hypothesis-testing: the dreaded p-value and statistical significance. In: *A clinician's guide to statistics and epidemiology in mental health: measuring truth and uncertainty*. Cambridge: Cambridge University Press; 2009.
5. Fisher R. *The design of experiments*. 9th ed. New York: MacMillan Pub Co; 1971. | Link |
6. Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. *N Engl J Med*. 2012 Oct 18;367(16):1562-4. | CrossRef | PubMed |
7. Coggon D, Geoffrey R, Barker D. Measurement error and bias. In: *Epidemiology for the uninitiated*. 5th ed. London: BMJ Books; 2003.
8. Araujo M. General validity criteria in clinical trials. *Medwave* 2012 Mar/Abr;12(3):e5338. | CrossRef |
9. Araujo M. Some biases and limitations present in biomedical literature: Part 1. *Medwave* 2013;13(3):e5647. | CrossRef |
10. Araujo M. Some biases and limitations present in biomedical literature: Part 2. *Medwave* 2013;13(4):e5668. | CrossRef |
11. de Irala J, Martínez-González MA, Guillén Grima F. [What is a confounding variable?]. *Med Clin (Barc)*. 2001 Oct 6;117(10):377-85. | PubMed |
12. Ananth CV, Schisterman EF. Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics. *Am J Obstet Gynecol*. 2017 Aug;217(2):167-175. | CrossRef | PubMed |
13. Strand KM, Heimstad R, Iversen AC, Austgulen R, Lydersen S, Andersen GL, et al. Mediators of the association between pre-eclampsia and cerebral palsy: population based cohort study. *BMJ*. 2013 Jul 9;347:f4089. | CrossRef | PubMed |
14. Araujo M. Random error in clinical trials. *Medwave* 2012 Jun;12(5):e5423. | CrossRef |
15. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016 Apr;31(4):337-50. | CrossRef | PubMed |
16. Jiménez-Paneque R. The questioned p value: clinical, practical and statistical significance. *Medwave* 2016 Sep;16(8):e6534. | CrossRef | PubMed |
17. Salsburg D. *The lady tasting tea: how statistics revolutionized science in the twentieth century*. New York: MacMillan Pub Co; 2001. | Link |
18. Madrid E, Martínez F. Statistics for the faint of heart - how to interpret confidence intervals and p values. *Medwave* 2014;14(1):5892. | CrossRef |
19. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA*. 2018 Apr 10;319(14):1429-1430. | CrossRef | PubMed |
20. Araujo M. Random error in clinical trials (II). *Medwave* 2012 Jul;12(6):e5437. | CrossRef |
21. Quidé Y, Bortolasci CC, Spolding B, Kidnapillai S, Watkeys OJ, Cohen-Woods S, et al. Association between childhood trauma exposure and pro-inflammatory cytokines in schizophrenia and bipolar-I disorder. *Psychol Med*. 2018 Dec 18:1-9. | CrossRef | PubMed |
22. Benavides A, Silva L. Contra la sumisión estadística: un apunte sobre las pruebas de significación. *Metas Enferm*. 2000;3(27):35-40.
23. Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA*. 1996 Oct 23-30;276(16):1332-8. | PubMed |
24. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2009 May;106(19):335-9. | CrossRef | PubMed |
25. Doll R. Proof of causality: deduction from epidemiological observation. *Perspect Biol Med*. 2002 Fall;45(4):499-515. | PubMed |
26. Cohen J. The earth is round ($p < .05$). *Am Psychol*. 1994;49(12):997-1003.
27. Silva L. El advenimiento de la era bayesiana. In: *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica* [Internet]. Madrid: Ediciones Díaz de Santos; 1997. | Link |
28. Silva L, Benavides A. Apuntes sobre subjetividad y estadística en la investigación en salud. *Rev Cuba Salud Pública*. 2003;29(2):170-3. | Link |
29. Evans SJ, Mills P, Dawson J. The end of the p value? *Br Heart J*. 1988 Sep;60(3):177-80. | PubMed |
30. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*. 1986 Mar 15;292(6522):746-50. | Link |

Correspondencia a

Escuela de Medicina
Universidad de Valparaíso
Angamos 655, Oficina 110
Reñaca, Viña del Mar
Chile.



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.