

Bayesian analysis. Basic and practical concepts for its interpretation and use

Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso

Mario Enrique Rendón-Macías,¹ Alberto Riojas-Garza,² Daniela Contreras-Estrada,² José Darío Martínez-Ezquerro³

Abstract

Bayesian statistics is based on subjective probability. It works with evidence updating considering the knowledge acquired prior to an investigation, plus the evidence obtained thereof. Results' interpretation requires for the hypotheses to be tested to be specified and their *a priori* probability to be estimated before the study. Study evidence is measured with the Bayes factor (compatibility ratio of the data under the proposed hypotheses). The conjunction of hypotheses *a priori* probabilities with the Bayes factor allows calculating the *a posteriori* probability of each one of them. The hypothesis with the highest degree of certainty at its update is the one that is accepted for decision making. In this review, three examples of hypothesis to be tested are shown: difference of means, correlation and association.

Keywords: Bayes statistics; Bayesian analysis; Subjective probability

Este artículo debe citarse como: Rendón-Macías ME, Riojas-Garza A, Contreras-Estrada D, Martínez-Ezquerro JD. Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso. Rev Alerg Mex. 2018;65(3):285-298

ORCID

Mario Enrique Rendón-Macías, 0000-0001-7310-6656; Alberto Riojas-Garza, 0000-0002-8880-6484; Daniela Contreras-Estrada, 0000-0001-7907-5870; José Darío Martínez-Ezquerro, 0000-0002-2609-4207

¹Instituto Mexicano del Seguro Social, Coordinación de Investigación en Salud, Unidad de Investigación en Epidemiología Clínica, Ciudad de México, México

²Universidad Panamericana, Escuela de Medicina, Ciudad de México, México

³Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Pediatría, Unidad de investigación en Genética Humana, Ciudad de México, México

Correspondencia. Mario Enrique Rendón-Macías.
drmariorendon@gmail.com

Recibido: 2018-06-24

Aceptado: 2018-07-05

DOI: 10.29262/ram.v65i3.512



Resumen

La estadística bayesiana se basa en la probabilidad subjetiva, trabaja con la actualización de la evidencia considerando los conocimientos adquiridos previos a una investigación, más la evidencia obtenida con esta. La interpretación de los resultados requiere la especificación de las hipótesis por contrastar y su probabilidad *a priori* antes del estudio. La evidencia del estudio se mide con el factor Bayes (razón de la compatibilidad de los datos bajo las hipótesis propuestas). La conjunción de las probabilidades *a priori* de las hipótesis con el factor Bayes permite calcular la probabilidad *a posteriori* de cada una. La hipótesis con mayor grado de certidumbre en su actualización es la aceptada para la toma de la decisión. En esta revisión se muestran tres ejemplos de hipótesis por contrastar: diferencia de promedios, correlación y asociación.

Palabras clave: Estadística de Bayes; Análisis bayesiano; Probabilidad subjetiva

Abreviaturas y siglas

FB, factor Bayes

PSEHN, prueba significación estadística de la hipótesis nula

RM, razón de momios

Antecedentes

Aunque la estadística bayesiana surgió antes que la frecuentista (o clásica), hasta años recientes ha sido retomada como una herramienta más intuitiva en la inferencia estadística para la toma de decisiones en la validez de los resultados de una investigación.^{1,2,3} En esta revisión mostramos cómo trabaja la prueba de hipótesis bayesiana para respaldar cuál hipótesis (nula o alterna) es más creíble con base en la evidencia obtenida en un estudio. Se incluyen tres ejemplos de estudios simulados sobre diferencias de medias, correlación y asociación. En los tres ejemplos se muestra cómo interpretar el factor Bayes (FB) y la probabilidad posterior o *a posteriori* de la veracidad de las hipótesis actualizadas con la información de los resultados, así como el efecto de las expectativas previas (probabilidad *a priori*) en esta probabilidad posterior.

Inferencia estadística

En la investigación médica como en otros campos científicos, el objetivo de un estudio es probar hipótesis de relaciones, causales o no, entre fenómenos

de interés del universo mediante estudios con muestras de este.⁴ La inferencia surge de la extrapolación de los resultados en dicha muestra hacia la población no participante. Un ejemplo frecuente consiste en establecer si un nuevo medicamento es más eficaz sobre el estándar aceptado en la resolución de una dolencia o enfermedad. La evidencia para apoyar esta hipótesis consiste en resumir la información (datos obtenidos empíricamente) de los participantes tratados con una de las dos opciones consideradas. Si la evidencia apoya la mayor eficacia de uno, el investigador sugerirá su uso en otros pacientes no participantes (inferencia). Como puede esperarse, ningún estudio puede controlar todos los factores involucrados en la predicción exacta de los fenómenos biológicos, en consecuencia, los resultados deben ser presentados como estimadores basados en probabilidades y su extrapolación dependerá del cumplimiento de supuestos de selección y medición durante el estudio. La estadística es la ciencia que apoya a la investigación en estos propósitos.⁵

Desde el origen de la estadística, dos perspectivas o posturas surgieron para establecer estas

inferencias: la frecuentista y la bayesiana. Aunque la primera ha dominado en las ciencias, en años recientes y a consecuencia del uso inadecuado e interpretación del valor de p y de la prueba de significación estadística de la hipótesis nula (PSEHN), aunado a varias ventajas de la estadística bayesiana, la Asociación Americana de Estadísticas y otros grupos académicos han considerado retomar la segunda.^{6,7,8,9}

Para entender cómo se genera la inferencia estadística en estos dos postulados es importante aclarar la forma de conceptualizar la probabilidad en cada uno. Para la frecuentista o clásica, la probabilidad se establece como una frecuencia de eventos esperados sobre un número de repeticiones de un procedimiento (por ejemplo, 0.5 o 50 % de las veces);⁵ mientras para la bayesiana, la probabilidad es un grado de creencia de la realidad, es decir, el nivel de certidumbre (0.5 o 50 % de certeza hacia algo).^{8,10,11} Estas dos visiones son importantes porque en las pruebas de hipótesis, en la primera se asume que la realidad desconocida tiene un comportamiento estable (parámetro), el cual puede ser determinado por la repetición de estudios con muestras iguales de esta. En la segunda, la probabilidad del parámetro desconocido se construye con los conocimientos adquiridos previos a un estudio, junto con la evidencia obtenida después de este. La nueva información modifica o no nuestra confianza o certeza sobre estos conocimientos *a priori*.^{2,12}

Pruebas de hipótesis

Una hipótesis es una conjetura sobre la realidad. En general se presenta como una predicción alcanzable con base en los argumentos teóricos disponibles antes de realizar una investigación, sin eliminar la posibilidad de la intuición.^{4,11} Cuando nos referimos a hipótesis estadísticas, estas son las que pueden ser probadas por métodos matemáticos, en particular los probabilísticos. Aunque en una investigación clínica se pueden contrastar más de dos hipótesis estadísticas, lo habitual es considerar solo dos: una hipótesis de nulidad y una alterna. La hipótesis de nulidad se refiere a una conjetura donde se predice que no hay diferencia o asociación entre las variables analizadas; por el contrario, la hipótesis alterna predice la existencia de una diferencia o asociación. La manera de contrastar estas hipótesis depende del tipo de estadística utilizada.

Contrastación por el método frecuentista

Para la estadística frecuentista únicamente se desafía la hipótesis nula, marcada como H_0 . Para ello, los frecuentistas desarrollaron modelos de distribución al azar bajo el supuesto de una hipótesis nula cierta. Esta distribución centraliza su mayor probabilidad hacia el valor nulo. Para determinar cuán compatible es un estimado observado en un estudio con este valor nulo, Fisher¹³ creó el valor de p . Este valor define la probabilidad de observar un dato resumido y sus valores más extremos en un modelo matemático especificado sobre la hipótesis nula. Entre más pequeña es esta probabilidad del dato, menos compatible será con el nulo y hará deducir que algo influye en este resultado tan extremo. Sin embargo, hasta las aportaciones de Neyman y Pearson no había manera de negar que este dato fuera una variación de la distribución del valor nulo.¹⁴ Por ello, ambos crearon la PSEHN. Propusieron que la hipótesis nula podría ser rechazada si el valor de p fuera menor al riesgo de cometer un error tipo I, fijado en 5 % o 0.05 (también expresado como $p < 0.05$). Este error implica rechazar la hipótesis nula (no asociación o no diferencia) y optar por la única alternativa de diferencia o asociación (señalada como H_A) cuando en realidad la nula es cierta.^{9,15} Como se puede inferir, la perspectiva frecuentista no pretende dar evidencia de la hipótesis nula (esta se asume como cierta) y mucho menos apoyar la alterna si se rechaza. Un error frecuente en la interpretación es asumir que el valor de p es la probabilidad del azar en los resultados, o bien, decir que es la probabilidad de certeza de la hipótesis nula.^{9,15} La prueba de hipótesis frecuentista solo indica la compatibilidad del estimado con la distribución de la hipótesis nula y asume que si el estudio se repitiera infinitas veces y mantuviera los valores extremos, pensaríamos que factores ajenos al modelo están causando esta rareza. Dichos factores pudieran ser los de interés al investigar.

Contrastación por el método bayesiano

Para contrastar hipótesis desde esta perspectiva, es importante que los investigadores especifiquen claramente estas antes de realizar sus estudios. En muchas ocasiones solo dos hipótesis son formuladas: una de no diferencia o no asociación, anotada como H_0 , y otra como el evento contrario, señalada como H_1 (para no confundirla con la alterna frecuentista). Sin embargo, una ventaja de este modelo es la posibilidad de con-

trastar más de una hipótesis.^{10,12,16,17} Una vez establecidas las hipótesis, asegurándose de ser mutuamente excluyentes y exhaustivas para las explicaciones plausibles, se determina su probabilidad bayesiana *a priori* (nivel de credibilidad). Esta probabilidad puede ser expresada como una razón (momios preresultados o momios *a priori*):

$$P(H_1)/P(H_0)$$

Donde:

$P(H_1)$ = Probabilidad de veracidad de la hipótesis de diferencia o asociación.

$P(H_0)$ = Probabilidad de no diferencia o no asociación.

Esta razón refleja los momios a favor o en contra de cada hipótesis antes de iniciar el estudio. Un momio de 1 equivale a igualdad de posibilidades, > 1 a favor de la H_1 y < 1 a favor de la H_0 . Con los datos colectados en la investigación, 1 establece la $p(H_1|\text{datos})$ o la compatibilidad de la hipótesis 1 con base en los datos obtenidos y la $p(H_0|\text{datos})$ o compatibilidad de la hipótesis 0 con estos mismos datos.^{8,12} La razón de estas dos probabilidades es la fuerza de la evidencia de un estudio y es conocida como FB,^{18,19} que puede ser informado como FB_{10} para referirse a la comparación entre H_1/H_0 o como FB_{01} para la relación inversa: H_0/H_1 . De esta forma, un $FB_{10} = 6$ se interpretaría como la existencia de seis veces más compatibilidad de los datos con la H_1 contra la H_0 , y una $FB_{01} = 5$ como cinco veces más compatibilidad de los datos con la H_0 en comparación con la H_1 . El FB, por tanto, es un valor por reportar en un estudio donde se utiliza la estadística bayesiana.^{2,12,16,20}

Otra ventaja de esta perspectiva es que la credibilidad final hacia una hipótesis depende también de nuestros conocimientos o expectativas antes de realizar un estudio, por ello, la necesidad de especificación previa.^{17,21} Es lógico pensar que si la evidencia *a priori* ha sido muy sólida en su teoría o secundaria a observaciones empíricas, los nuevos datos aportarán poco a la credibilidad final, pero si estas han sido muy pocas o mínimamente fundadas, aún con un FB tan grande (fuerza de la evidencia) como para aumentar o disminuir sustancialmente la credibilidad, es posible que no sea suficiente para cambiar la decisión en contra de la hipótesis

que inicialmente se consideró. En esta postura es fundamental calcular la probabilidad posestudio o *a posteriori* como la credibilidad actualizada, es decir, la credibilidad que se obtiene tras la conjunción del nivel *a priori* más la evidencia obtenida del estudio.¹²

Análisis e interpretación

Para ejemplificar cómo se analizaría e informaría un análisis bayesiano contra uno frecuentista se presentan tres escenarios de investigación. Existen más posibilidades, por lo que lectores interesados en profundizar sobre el tema deberán revisar obras clásicas.^{1,3}

Escenario 1

En un estudio se desea conocer la eficacia de dos antihistamínicos (R o C) para controlar el grado de hinchazón cutánea (área) medida en milímetros cuadrados. Los autores determinaron el nivel la reducción al finalizar los tratamientos como el área final menos la inicial. Por el método frecuentista, la hipótesis nula a contrastar puede ser de dos tipos: los promedios de las diferencias en el tamaño de la hinchazón después de los tratamientos son iguales (hipótesis de dos colas) o esta es mayor o igual con el tratamiento R con respecto al C (una cola), de tal forma que tendremos la siguiente información:

- *Tratamiento R*: 40 participantes, promedio de la reducción -9.7 mm^2 , desviación típica de la reducción 1.8, error típico de la reducción 0.29.
- *Tratamiento C*: 40 participantes, promedio de la reducción -8.2 mm^2 , desviación típica de la reducción 2.3, error típico de la reducción 0.36.

Análisis frecuentista de un contraste de medias

Hubo una reducción 0.72 mm mayor con el tratamiento R en comparación con el C y esta diferencia equivale a 3.19 desviaciones del estadístico t (medida de resumen de la diferencia considerando las variancias conjuntas) (Cuadro 1). Como esta desviación es muy extrema al valor nulo de 0 (media del valor nulo), la probabilidad de encontrarla en este modelo teórico es de 0.2 % de las veces ($p = 0.002$). Si se considera el criterio de la PSEHN, el valor rebasa el umbral de rechazo de $p < 0.05$, por lo tanto, rechazamos la hipótesis nula y decidimos considerar la alterna de diferencia.⁵ Sin embargo, lo único que podemos concluir es lo raro de este resultado

bajo la hipótesis nula. Dado que el modelo frecuentista se basa en que la repetitividad de un resultado en muestras independientes se aproximará al valor real (paramétrico desconocido), una estrategia ha sido calcular los intervalos de confianza a 95 %, ^{22,23} Con este cálculo se podría concluir que en 95 % de intervalos calculados de forma independiente y con la misma cantidad de pacientes, la media de la diferencia estaría entre una reducción de 0.29 y 1.2 mm más con el tratamiento R comparado con el C. Como la diferencia siempre es a favor del tratamiento R, tenemos más confianza en decidir su uso. Sin embargo, el impacto del tratamiento (la diferencia esperada) puede tomar cualquier valor del intervalo referido, por eso se dice que entre más pequeño sea, habrá mayor precisión. Un problema es que rara vez se repiten estudios para comprobar estos supuestos.⁷

Análisis bayesiano de un contraste de medias

En el análisis bayesiano (Cuadro 2) lo que se contrasta es la credibilidad, también conocida como verosimilitud, de las hipótesis planteadas una vez actualizada la información con los resultados de un estudio.^{20,21} Como se comentó, aquí sí importa la estimación de nuestras creencias antes de un estudio y los resultados se analizarán con tres supuestos. El primero es con una probabilidad previa no informativa, es decir, los autores no tienen claro si la hipótesis de asociación o diferencia es más plausible que la de no asociación o no diferencia. El segundo es cuando los autores son optimistas a favor de la hipótesis de asociación o diferencia (probabilidad > 50 %) por evidencia previa o alguna teoría a favor. La tercera, cuando se tiene una expectativa pesimista donde se apuesta más a la no

asociación o no diferencia (> 50 %), también por evidencias previas o teorías contrarias.

Para cada ejemplo se considerará, igual que en la frecuentista, la posibilidad de una hipótesis nula (H_0) bilateral de no diferencia y otra donde la H_0 defiende un tratamiento como igual o mejor que el otro.

En un estudio real, los autores deben señalar las probabilidades *a priori* y sus hipótesis de análisis y reportar solo estas. Aquí realizamos todas para fines explicativos.

Para la primera opción de una probabilidad previa no informativa (Cuadro 2, A) o ambas igualmente plausibles (50 % cada una), la evidencia del estudio dio un FB_{10} de 17.8 (Cuadro 2, A.1). Es decir, los datos son 17.8 veces más compatibles con la hipótesis de diferencia (H_1) que con la de no diferencia (H_0). Si se utilizara el FB_{01} diríamos que los datos son 0.05 veces más compatibles con la H_0 que con la H_1 (misma conclusión). Aunque no existe un nivel de corte del FB_{10} para determinar alta evidencia, en general niveles arriba de 5 ya son muy informativos.^{18,19} Más importante, el análisis muestra que con la evidencia observada en el estudio y considerando creencias *a priori* neutras, la probabilidad actualizada es de 98.3 % a favor de un efecto diferencial entre los tratamientos contra 1.7 % a favor de ser iguales. También este análisis permite determinar el estimado más probable de la diferencia (0.64 mm menos) y su posible variación. Este intervalo se conoce como intervalo de credibilidad a 95 %, ^{11,24} el cual a diferencia del intervalo de confianza sí expresa un valor posible del parámetro. Por tanto, la inferencia es directa y no indirecta como intervalos posibles.^{6,7,20,25}

Por otro lado, si la hipótesis 1 hubiera sido unilateral a favor del tratamiento A (Cuadro 2, A.1), los datos muestran un FB_{10} más grande (35.5) que el

Cuadro 1. Contrastación de hipótesis de diferencia de medias por el método frecuentista			
Hipótesis estadística	Valor de t (grados de libertad)	p	Diferencia de reducción (intervalo de confianza a 95 %)
Bilateral			
$H_N: \bar{X}_R = \bar{X}_C$ (tratamiento R = C)	Prueba dos colas 3.219 (78 gl)	0.002	0.72 (0.26 a 1.2)
$H_A: \bar{X}_R \neq \bar{X}_C$ (tratamiento R ≠ C)			
Unilateral			
$H_N: \bar{X}_R \geq \bar{X}_C$ (tratamiento R ≥ C)		< 0.001	Tratamiento R reduce 0.27 (0.59 a ∞)
$H_A: \bar{X}_R < \bar{X}_C$ (tratamiento R < C)			
H_N = hipótesis nula, H_A = hipótesis alterna			

Cuadro 2. Contrastación de hipótesis de diferencia de medias por el método bayesiano

Hipótesis estadística	Factor Bayes	Probabilidad posterior (%)	Mediana de diferencia (IC a 95 %)
A. Probabilidad <i>a priori</i> no informativa			
Bilateral A.1			
$H_0: \bar{X}_R = \bar{X}_C$ (tratamiento R = C)	FB ₀₁ 0.05:1	1.7	0.64 (0.21 a 1.09)
$H_1: \bar{X}_R \neq \bar{X}_C$ (tratamiento R ≠ C)	FB ₁₀ 17.8:1	98.3	
Unilateral A.2			
$H_0: \bar{X}_R \leq \bar{X}_C$ (tratamiento R ≤ C)	FB ₀₁ 0.02:1	2.8	0.65 (0.23 a 1.1)
$H_1: \bar{X}_R > \bar{X}_C$ (tratamiento R > C)	FB ₁₀ 35.5:1	97.2	
Unilateral A.3			
$H_0: \bar{X}_R \geq \bar{X}_C$ (tratamiento R ≥ C)	FB ₀₁ 16.6:1	94.4	-0.08 (-0.09 a -0.008)
$H_1: \bar{X}_R < \bar{X}_C$ (tratamiento R < C)	FB ₁₀ 0.06:1		
B. Probabilidad <i>a priori</i> informativa optimista (diferencia a favor tratamiento A de promedio 1 ± 0.2)			
Bilateral B.1			
$H_0: \bar{X}_R = \bar{X}_C$ (tratamiento R = C)	FB ₀₁ 0.01:1	1.5	0.88 (0.58 a 1.17)
$H_1: \bar{X}_R \neq \bar{X}_C$ (tratamiento R ≠ C)	FB ₁₀ 66.7:1	98.5	
Unilateral B.2			
$H_0: \bar{X}_R \leq \bar{X}_C$ (tratamiento R ≤ C)	FB ₀₁ 0.01:1	1.0	0.88 (0.59 a 1.16)
$H_1: \bar{X}_R > \bar{X}_C$ (tratamiento R > C)	BF ₁₀ 68:1	99.0	
Unilateral B.3			
$H_0: \bar{X}_R \geq \bar{X}_C$ (tratamiento R ≥ C)	FB ₀₁ 1.53:1	60.7	-0.017 (-0.09 a -0.001)
$H_1: \bar{X}_R < \bar{X}_C$ (tratamiento R < C)	FB ₁₀ 0.65:1	39.3	
C. Probabilidad <i>a priori</i> informativa pesimista (diferencia a favor tratamiento B de promedio 1 ± 0.2)			
Bilateral C.1			
$H_0: \bar{X}_R = \bar{X}_C$ (tratamiento R = C)	FB ₀₁ 80052:1	99.99	-0.26 (0.56 a 0.03)
$H_1: \bar{X}_R \neq \bar{X}_C$ (tratamiento R ≠ C)	FB ₁₀ 0.00:1	0.001	
Unilateral C.2			
$H_0: \bar{X}_R \leq \bar{X}_C$ (tratamiento R ≤ C)	FB ₀₁ 0.54:1	35.1	0.04 (0.002 a 0.204)
$H_1: \bar{X}_R > \bar{X}_C$ (tratamiento R > C)	FB ₁₀ 1.83:1	64.9	
Unilateral C.3			
$H_0: \bar{X}_R \geq \bar{X}_C$ (tratamiento R ≥ C)	FB ₀₁ 83572:1	99.99	-0.27 (-0.56 a -0.033)
$H_1: \bar{X}_R < \bar{X}_C$ (tratamiento R < C)	FB ₁₀ 0.0001:1	0.001	

IC = Intervalo de credibilidad

bilateral. Esto traduce una mayor congruencia a la diferencia esperada. Como se observa, la mediana de la diferencia y la probabilidad posterior de apoyo a la H_1 no cambian sustancialmente. Contrariamente, si la hipótesis 1 hubiera apoyado el tratamiento C (Cuadro 2, A.3), la FB_{10} sería a favor de la hipótesis la nula (0.06, o FB_{01} de 16.6) y la credibilidad a favor de la hipótesis de mayor eficiencia con el tratamiento C bajaría de una inicial de 50 a 5.6 %.

Como revisamos, en ocasiones los investigadores tienen argumentos para apoyar una mayor eficacia del tratamiento R comparado con el C desde antes del estudio. Esta información optimista haría pensar en una reducción promedio mayor a 1 mm de este grupo sobre el otro e implicaría considerar una probabilidad *a priori* de esta hipótesis arriba de 50 %. Bajo esta premisa, la evidencia del estudio generará modificaciones en las conclusiones (Cuadro 2). En una hipótesis bilateral (Cuadro 2, B1), el FB_{10} fue de 66.7 o muy compatible con la H_1 (son diferentes) y no con la H_0 . Esta fuerte evidencia en conjunto con la información previa actualiza nuestra confianza en 98.5 % de certidumbre al afirmar que son diferentes. Si se hubiera formulado una hipótesis unilateral a favor del tratamiento R (Cuadro 2, B.2), nuevamente el FB_{10} indica datos más compatibles con la hipótesis de superioridad de R contra la de igual o menor a C. Asimismo, apoyaría más nuestra creencia, hasta 99 % de credibilidad a favor del tratamiento R.

Es de notar que el estimado de la diferencia se hace mayor (0.88 mm) al considerarse el margen propuesto *a priori*. Por otro lado, en el supuesto de que los autores aún con evidencia a favor de R hubieran apostado más a un efecto superior con el tratamiento C (Cuadro 2, B.3), el FB_{01} seguiría apoyando más la hipótesis de efecto igual o mayor de R (hipótesis 0, con FB_{01} de 1.53); su credibilidad dirigida a mayor eficiencia de C no subiría sino hasta 39.3 %. En este contraste, los datos actualizados apoyarían 60.7 % de credibilidad de que el tratamiento R es al menos igual si no es que superior al C.

En el mismo Cuadro 2 (C) se presentan el análisis en la situación donde se tiene baja confianza en la superioridad del tratamiento R (previo pesimista). En el contraste de una hipótesis bilateral, el BF_{01} es muy elevado apoyando la hipótesis de efectos no diferentes entre los tratamientos, con una certeza de esta hipótesis muy cercana a 100 %.

El estimado en la prueba bilateral (diferencia de medias) muestra una mediana negativa, pero con un intervalo de credibilidad de valores tanto a favor del tratamiento C como del tratamiento R, es decir, no son estadísticamente diferentes. En el análisis unilateral a favor del tratamiento R (Cuadro 2, C.2), el FB_{10} fue de 1.83, dada la existencia de evidencia de datos más compatibles con esta hipótesis de superioridad sobre la de igualdad o inferioridad a C. Aunque en este ejemplo los autores tenían una probabilidad *a priori* a favor de C, con los datos su probabilidad se actualizaría aumentando de una inicial menor a 50 % a una en favor de R de 64.9 % (probabilidad *a posteriori*).

En el contraste a favor del tratamiento C, con una probabilidad *a priori* de una mayor eficiencia de este sobre el R (Cuadro 2, C.3), observamos que los datos son prácticamente nada compatibles con esta hipótesis ($FB_{10} = 0.001$), sino altamente compatibles con la superioridad o al menos igualdad con el tratamiento R ($FB_{01} = 83572$). Aun con evidencia previa y con los datos del estudio, la probabilidad o creencia en que C sea más efectivo a R es de 0.001 %.

En conclusión, con la estadística bayesiana podemos aseverar en este estudio que, sin evidencia previa informativa, el tratamiento R tiene 98.3 % de credibilidad de ser diferente o 97.2 % de ser superior a C; mientras C solo tiene 5.6 % de superioridad en comparación con R. Con evidencia *a priori* de mayor eficiencia de R sobre C, una certeza de 98.5 % de no ser igual a C, 99 % de ser superior si apostáramos a R y 60.7 % de superioridad o igual a C (si apostáramos a este tratamiento). Finalmente, en una situación de alta creencia *a priori* de una eficacia superior con el tratamiento C, los resultados llevarían a dudar de ello, apoyando con 99.9 % de certeza la probabilidad de igualdad. En una hipótesis a favor del tratamiento R, su credibilidad subiría a 35.1 % y con una hipótesis a favor del tratamiento C tendría una credibilidad de 0.001 %.

Escenario 2

En un estudio se desea saber si existe correlación entre las semanas de edad gestacional de neonatos pretérmino con los valores de eosinófilos en su sangre periférica. Los resultados son los siguientes:

- *Semanas de edad gestacional*: 130 sujetos, promedio 33, desviación típica 1.9, error típico 0.21.

- *Eosinófilos*: 130 sujetos, promedio 2.28 %, desviación típica 2.3, error típico 0.19.

Análisis frecuentista de una correlación

Los resultados del Cuadro 3 muestran un estadístico de Pearson de $r = 0.10$, el cual se interpreta como una correlación baja, donde solo 3.16 % de la variabilidad de los eosinófilos puede estar relacionada con la edad gestacional. Nuevamente para la estadística frecuentista solo se analiza la hipótesis nula (H_N), donde se afirma ausencia de correlación o $r = 0$. Una gran repetición de estudios iguales en $N = 130$ tendrían un promedio de r cercanos a 0.10 y pocas veces con un valor de 0. En este estudio dicha probabilidad es hasta de 25 % (valor de p); como es mayor al valor crítico de 5 %, la PSEHN indicaría que no podemos rechazar la hipótesis. Otra forma de expresarlo es una posibilidad de 25 % de cometer un error tipo I (rechazar la hipótesis nula cuando es cierta). Bajo esta perspectiva en realidad nunca se rechaza la hipótesis nula, solo se afirma la probabilidad del estimado en esta distribución.^{9,11} Para dar más información sobre esta correlación, el intervalo de confianza a 95 % indica que existe la posibilidad de encontrar en 95 de 100 estudios semejantes en N , intervalos de valores de R entre -0.07 y 0.26 . Como se incluye el valor 0, no hay confianza en aseverar una correlación.

Por otro lado, una hipótesis de correlación pudiera ser unilateral al aseverar relaciones positivas o negativas contra su ausencia o dirección contraria. En el Cuadro 3 se muestra que ante una hipótesis nula de ausencia de correlación positiva, nuevamente la p es superior a 0.05. Por la PSEHN no se debe

rechazar esta hipótesis, porque tendríamos una posibilidad de 12.8 % de cometer un error tipo I. Si la propuesta era una hipótesis nula de ausencia de correlación negativa, la probabilidad de cometer el error I al rechazarla sería mayor de 87.2 %.

Análisis bayesiano de una correlación

En este análisis, el valor de r de Pearson no cambia dado que es calculado igual que el frecuentista.^{1,2} La diferencia estriba en la forma de contrastar las hipótesis. Una vez más es necesario que estas sean establecidas antes del estudio. En el Cuadro 4 se muestran los resultados considerando el tipo de hipótesis y la probabilidad o credibilidad *a priori* de la correlación establecida por los investigadores antes de realizar el estudio. Para una evidencia no informativa previa con una hipótesis bidireccional (Cuadro 4, A.1), las dos hipótesis a considerar son $H_0 =$ no hay correlación o r es igual a 0, contra $H_1 =$ hay correlación o r es diferente a 0. El FB_{01} al ser de 4.8 traduce casi cinco veces más compatibles los datos con la H_0 que con la H_1 , apoyando a la H_0 .

De esta manera, con la evidencia del estudio y en un marco de falta de información previa, donde ambas hipótesis son igualmente posibles, la probabilidad actualizada es de 82.7 % a favor de la no encontrado es el más posible dentro de los 95 % creíbles. Por eso, aunque en el intervalo r pudiera ir desde un valor negativo a uno positivo, el más posible es 0.10. Esta información es valiosa porque permite hacer conjeturas aun sin la significación estadística. En la perspectiva frecuentista, el valor de r puede ser cualquiera en el intervalo de confianza con la misma probabilidad.^{7,9}

Cuadro 3. Contrastación de hipótesis de correlación por el método frecuentista

Hipótesis estadística	Prueba estadística r de Pearson	p	Intervalo de confianza a 95 %
Correlación			
H_N : $R = 0$ (no correlación)	0.10	0.25	-0.07 a 0.26
H_A : $R \neq 0$ (hay correlación)			
Unilateral positiva			
H_N : $R \leq 0$ (no hay o es negativa)	0.10	0.128	-0.04 a 1.00
H_A : $R > 0$ (es positiva)			
Unilateral negativa			
H_N : $R \geq 0$ (no hay o es positiva)	0.10	0.872	-1.00 a 0.24
H_A : $R < 0$ (es negativa)			
H_N = hipótesis nula, H_A = hipótesis alterna			

Siguiendo con una probabilidad *a priori* no informativa, pero ahora si las hipótesis fueran unilaterales, con opciones hacia una positiva o negativa contra su ausencia o en dirección contraria (Cuadro 4, A.2 y A.3), los resultados serían los siguientes: para una hipótesis a favor de una correlación positiva esta tiene un FB_{10} es de 0.36, es decir, con datos más compatibles con la hipótesis contraria; después del estudio, la certeza de esta hipótesis es de 26 %. Para la propuesta inversa de una correlación negativa, el FB_{10} es de 0.5, nuevamente hay datos más compatibles con no correlación o posiblemente positiva. La plausibilidad de que la hipótesis de correlación negativa sea cierta es de 5.1 %.

Ahora bien, en el supuesto de disponer de evidencia previa a favor de una correlación positiva, porque teóricamente sería congruente con los conocimientos existentes o existen estudios anteriores que así lo sugieran, uno puede ver en la sección B del Cuadro 4 cómo bajó la evidencia en el estudio de la no correlación ($FB_{01} = 3.1$), por tanto, la probabilidad *a posteriori* actualizada de esta hipótesis ahora es de 76.2 %. Es evidente que aún no hay suficiente evidencia como para cambiar la conclusión a que existe correlación ($H_1 = 23.8$ %). Si se hubiera apostado por una correlación positiva, la credibilidad en esta hipótesis sería de 35.7 % (Cuadro 4, B.2), sin embargo, nuevamente los datos no dan suficiente evidencia para apoyarla. Pero como se observa para la posibilidad contraria de una correlación negativa (Cuadro 4, B.3), la hipótesis a favor de esta después de recolectados los datos es de 7.5 % (probabilidad *a posteriori*). Por tanto, aun con altas posibilidades iniciales de la existencia de una correlación, en congruencia con los datos obtenidos del estudio no hay suficiente evidencia para creer la posibilidad de una correlación de ningún tipo (positiva o negativa); por el contrario, los datos apoyan una hipótesis de no correlación.

En la parte C del Cuadro 4 se analizan las evidencias con una probabilidad *a priori* de una correlación negativa. Bajo este supuesto, los datos nuevamente en un análisis bilateral son más compatibles con la no correlación ($FB_{01} = 7.5$ y probabilidad *a posteriori* de 88.2 %). Si se apostara a una correlación positiva, la evidencia disminuiría la plausibilidad de la hipótesis nula ($FB_{01} = 4.3$ y probabilidad *a posteriori* de 81.1 %), pero aún a favor de su aceptación. En una apuesta contraria a favor de una correlación negativa,

acorde con su postura inicial, los datos obtenidos del estudio van más en contra de esta conjetura ($FB_{10} = 0.03$, probabilidad *a posteriori* de 3.3 %).

Podemos concluir en este ejercicio que independientemente de la probabilidad *a priori* considerada, la evidencia de los datos en todo momento era más compatible con la falta de correlación, sin embargo, la dirección de las hipótesis (bilateral, positiva o negativa) aunada a la credibilidad previa al estudio sin cambiar el grado de confianza en la hipótesis nula (probabilidad *a posteriori*) al final del estudio.

Escenario 3

En un estudio de cohortes se desea saber el impacto en los niños preescolares de nadar en albercas cerradas cloradas con el desarrollo de dermatitis alérgica. Los resultados fueron los siguientes:

- *Si nadar en albercas cerradas cloradas:* desarrollaron dermatitis 35 niños (39.3 %), no desarrollaron dermatitis 54 (60.7 %), total 89.
- *No nadar en albercas cerradas cloradas:* desarrollaron dermatitis 31 (24.8 %), no desarrollaron dermatitis 94 (75.2 %), total 125.

Análisis frecuentista de una asociación

En el análisis frecuentista de una asociación, la fuerza de esta se obtiene principalmente calculando el riesgo relativo o la razón de momios (RM). Se afirma la ausencia de asociación cuando estos toman el valor de 1. Cuando es < 1 se habla de un efecto protector y cuando es > 1 de riesgo o daño. En el Cuadro 5 se muestra el análisis estadístico del comportamiento de los datos de este estudio. Bajo una hipótesis nula de probabilidades (frecuencias) iguales de dermatitis entre los niños nadadores en albercas cloradas contra los controles, se obtuvo chi cuadrado de 5.14 (medida de resumen entre las diferencias de las proporciones de ambos grupos). Este valor bajo una distribución teórica al azar se aleja del valor nulo de 0, por lo cual la probabilidad de encontrarlo en esta distribución teórica es de 2.3 % de las veces, es decir, en muy rara vez. Aplicando la PSEHN, este valor está más allá de 5 % del usualmente prefijado, por lo cual se rechazaría la hipótesis de no asociación y se tomaría una posible alternativa de alguna asociación.²⁶

Insistimos en que en realidad no se rechaza la H_N , está en el modelo, es cierta; en realidad se

Cuadro 4. Contrastación de hipótesis de correlación por el método frecuentista				
Hipótesis estadística	Prueba r de Pearson	FB	Probabilidad posterior (%)	IC a 95 %
A. Probabilidad previa (<i>a priori</i>) no informativa				
Correlación A.1				
$H_0: r = 0$ (no correlación)	0.10	$FB_{01} 4.8:1$	82.7	-0.07 a 0.26
$H_1: r \neq 0$ (no correlación)		$FB_{10} 0.21:1$	17.3	
Unilateral positiva A.2				
$H_0: r \leq 0$ (no hay o es negativa)	0.10	$FB_{01} 4.8:1$	82.7	-0.07 a 0.26
$H_1: r > 0$ (es positiva)		$FB_{10} 0.21:1$	17.3	
Unilateral negativa A.3				
$H_0: r \geq 0$ (no hay o es positiva)	0.10	$FB_{01} 18.5:1$	94.8	-0.001 a 0.13
$H_1: r < 0$ (es negativa)		$FB_{10} 0.5:1$	5.1	
B. Probabilidad previa (<i>a priori</i>) informativa correlación positiva esperada de $R = 0.50$				
Correlación B.1				
$H_0: r = 0$ (no correlación)	0.10	$FB_{01} 3.2:1$	76.2	-0.07 a 0.26
$H_1: r \neq 0$ (hay correlación)		$FB_{10} 0.31:1$	23.8	
Unilateral positiva B.2				
$H_0: r \leq 0$ (no hay o es negativa)	0.10	$FB_{01} 1.8:1$	64.3	-0.008 a 0.26
$H_1: r > 0$ (es positiva)		$FB_{10} 0.53:1$	35.7	
Unilateral negativa A.3				
$H_0: r \geq 0$ (no hay o es positiva)	0.10	$FB_{01} 12.4:1$	92.5	-0.0001 a 0.13
$H_1: r < 0$ (es negativa)		$FB_{10} 0.08:1$	7.5	
C. Probabilidad previa (<i>a priori</i>) informativa correlación negativa esperada de $R = -0.50$				
Correlación C.1				
$H_0: r = 0$ (no correlación)	0.10	$FB_{01} 7.5:1$	88.2	-0.07 a 0.26
$H_1: r \neq 0$ (hay correlación)		$FB_{10} 0.13:1$	11.8	
Unilateral positiva C.2				
$H_0: r \leq 0$ (no hay o es negativa)	0.10	$FB_{01} 4.3:1$	81.1	-0.008 a 0.26
$H_1: r > 0$ (es positiva)		$FB_{10} 0.32:1$	18.8	
Unilateral negativa C.3				
$H_0: r \geq 0$ (no hay o es positiva)	0.10	$FB_{01} 29.1:1$	96.1	-0.0001 a 0.13
$H_1: r < 0$ (es negativa)		$FB_{10} 0.03:1$	3.3	
IC = Intervalo de credibilidad				

Cuadro 5. Contrastación de hipótesis de asociación por el método frecuentista		
Hipótesis estadística	Chi cuadrado	p
Diferencias proporciones		
H_N : $p(A) = p(B)$ (No diferentes)	5.14 (1 gl)	0.023
H_A : $p(A) \neq p(B)$ (Diferentes)		
Intervalo de confianza a 95 %		
Razón de momios		
H_N : $R = 1$ (no asociación)	1.96	1.09 a 3.54
H_A : $R \neq 1$ (hay asociación)		
H_N = Hipótesis nula, H_A = Hipótesis alterna		

acepta que de repetirse el estudio muchas veces este valor aparecería en 2.3 % veces. Sin embargo, con esta prueba estadística no se apoya la probabilidad de certeza de la H_A de la existencia de una asociación. En el análisis del estimador, su intervalo de confianza a 95 % fue siempre superior a 1, lo cual informa que en 95 % de los intervalos calculados con muestras del mismo tamaño, el estimado estaría por arriba del valor 1, entre 1.09 y 3.54 y siempre cada número con la misma posibilidad de ser el valor paramétrico.^{9,27}

Análisis bayesiano de una asociación

Como se ha indicado, el análisis se puede realizar bajo las tres posibles premisas: una no informativa o sin compromiso para alguna hipótesis (no asociación contra asociación), otra con evidencia previa a favor de una posible asociación y otra en contra de cualquier asociación. Bajo la primera (Cuadro 6, A), los datos arrojaron un BF_{10} de 2, es decir, hay dos veces más compatibilidad de los datos con la hipótesis de asociación contra la de no asociación. Con la evidencia de una probabilidad *a priori* neutra (0.5 o 50 %) al final del estudio se tiene 80 % de certeza (probabilidad *a posteriori*) sobre la hipótesis de asociación.

Bajo el supuesto inicial del estudio sobre la alta posibilidad de una asociación de riesgo (> 1) (Cuadro 4, B), los resultados dan mayor apoyo a esta hipótesis. La FB_{10} subió a 3.9, casi cuatro veces más compatibilidad de los datos con la hipótesis de RM mayor a 1 contra ser de igual o menor a 1. La credibilidad de esta hipótesis *a posteriori* ahora es ahora de 80 %. Finalmente, para una postura inicial contraria, donde los autores pensaban más en que la exposición fuera protectora (Cuadro 4, C), la evidencia

de los datos fue menos compatible (H_0). El FB_{01} fue 22.5 veces más compatible con la no asociación o de riesgo (daño). La credibilidad en la hipótesis de protección es de solo 4.3 % (probabilidad *a posteriori*).

En cualquiera de los análisis se observa la alta posibilidad de concluir sobre el efecto de riesgo de la exposición. Este riesgo es bajo con un estimador más probable de 1.94; en esta perspectiva podemos aseverar que este valor tiene esa cualidad. Además, podemos decir que la variabilidad va de 1.08 a 3.53, con una credibilidad de 95 %. El impacto de este riesgo en la decisión clínica depende del umbral de significación para esta. Los lectores interesados en conocer cómo interpretar la significación clínica y la estadística pueden leer una revisión previa en esta revista.²⁸

Comentarios finales

Mediante los ejemplos anteriores hemos mostrado algunas ventajas de la estadística bayesiana sobre la frecuentista. La primera es que constituye un método más intuitivo para comprender cómo una investigación no modifica, aumenta o disminuye los conocimientos sobre un problema científico.^{8,29,30,31} La segunda es que permite incluir la información previa a un estudio en la evaluación y conclusiones del mismo, situación no tomada en cuenta en el análisis clásico o frecuentista.^{2,7} Tercero, el método frecuentista se basa en el supuesto de repetir los estudios varias veces, situación que nunca sucede. Por su parte, el bayesiano asume que un estudio puede brindar evidencia para apoyar o no las hipótesis; se enfoca en la confianza en nuestras observaciones.^{8,32} Cuarto, dado que el método bayesiano es inductivo y no deductivo, es factible llegar a conclusiones con menos ensayos o pacientes. Este problema es muy

Cuadro 6. Contrastación de hipótesis de asociación por el método bayesiano		
Razón de momios bayesiano	1.94	IC a 95 % = 1.08 a 3.52
	Factor de Bayes	Probabilidad posterior (%)
A. Previo no informativo		
H ₁ : RM ≠ 1 (hay asociación)	FB ₁₀ 2.0	66.8
H ₀ : RM = 1 (no asociación)	FB ₀₁ 0.4	31.2
B. Previo informativo a favor de una asociación de riesgo		
H ₁ : RM > 1 (hay asociación de daño)	FB ₁₀ 3.9	80
H ₀ : RM ≥ 1 (no asociación)	FB ₀₁ 0.2	20
C. Previo informativo a favor de una asociación de protección		
H ₀ : RM ≥ 1 (no asociación o protectora)	FB ₀₁ 22.5	95.7
H ₁ : RM < 1 (hay asociación)	FB ₁₀ 0.04	4.3

IC = Intervalo de credibilidad

frecuente en la investigación clínica de enfermedades con baja incidencia y prevalencia.¹

Una ventaja final es poder calcular los estimados (promedios, diferencias, riesgos, etcétera) con mayor precisión, dado que el valor calculado será el más probable. El intervalo de credibilidad traduce mejor nuestra expectativa de cuáles pudieran ser las variaciones del estimado buscado, pero siempre con la certeza sobre el obtenido como el más probable.

Una pregunta esperada es, ¿por qué con todas estas ventajas no se realizaba o realiza en todos los trabajos publicados? Varios autores,^{7,15,19,33,34} consideramos cuatro principales explicaciones:

- La falta de difusión, dado el predominio de la estadística frecuentista en el mundo, incluso su exigencia en el reporte de trabajos enviados a publicación, situación que ha estado cambiando en los últimos años.
- El escaso conocimiento en su interpretación y ejecución. Pocos profesores en pregrado y posgrado conocen y enseñan este tipo de estadística.
- La complejidad de su cálculo en algunos procedimientos. Actualmente se han desarrollado programas estadísticos específicamente para análisis bayesianos y otros, ya disponibles, han agregado aplicaciones para su ejecución.
- La dificultad de varios investigadores y clínicos en comprender la probabilidad subjetiva. Al ser subjetiva se piensa que no tiene fundamento matemático, que está a criterio del investigador y, por lo tanto, no es reproducible. Sin embargo, el ser subjetiva no significa ser azarosa. Los investigadores pueden apoyar sus probabilidades claramente en la evidencia disponible en estudios previos, puede ser consensuada por expertos o puede ser admisible a otros colegas bajo argumentos teóricos bien fundados.

Referencias

1. Kruschke JK. Introduction. En: Doing Bayesian data analysis. A tutorial with R, JAGS, and Stan. EE. UU.: Elsevier; 2015. DOI: 10.1016/C2012-0-00477-2
2. Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. Clin Trials. 2005;2(4):282-290. DOI: 10.1191/1740774505cn098oa

3. Peter M Lee. Bayesian statistic: an introduction. EE. UU.: John Wiley & Sons, Ltd; 2000.
4. Indrayan A, Kumar-Malhotra R. Medical biostatistics. EE. UU.: Chapman & Hall/Chemical Rubber Company Press; 2008.
5. Krzywinski M, Altman N. Significance, P values and t-tests. *Nat Methods*. 2013;10(11):1041-1042.
6. Buchinsky FJ, Chadha NK. To P or not to P: backing Bayesian statistics. *Otolaryngol Head Neck Surg*. 2017;157(6):915-918. DOI: 10.1177/0194599817739260
7. Dienes Z, Mclatchie N. Four reasons to prefer Bayesian analyses over significance testing. *Psychon Bull Rev*. 2018;25(1):207-218. DOI: 10.3758/s13423-017-1266-z
8. Austin PC, Brunner LJ, Hux JE. Bayeswatch: an overview of Bayesian statistics. *J Eval Clin Pract*. 2002;8(2):277-286. DOI: 10.1046/j.1365-2753.2002.00338.x
9. 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;31(4):337-350. DOI: 10.1007/s10654-016-0149-3
10. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. *Br J Math Stat Psychol*. 2013;66(1):8-38. DOI: 10.1111/j.2044-8317.2011.02037.x
11. Wagenmakers EJ, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychon Bull Rev*. 2018;25(1):35-57. DOI: 10.3758/s13423-017-1343-3
12. Etz A, Gronau QF, Dablander F, Edelsbrunner PA, Baribault B. How to become a Bayesian in eight easy steps: an annotated reading list. *Psychon Bull Rev*. 2018;25(1):219-234. DOI: 10.3758/s13423-017-1317-5
13. *Statistical Methods for Research Workers*. Reino Unido: Olivier & Boyd; 1932.
14. Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc Math Phys Eng Sci*. 1933;231(694-706):289–337.
15. Nurminen M. Statistical significance—a misconstrued notion in medical research. *Scand J Work Environ Health*. 1997;23(3):232-235.
16. Masson ME. A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav Res Methods*. 2011;43(3):679-690. DOI: 10.3758/s13428-010-0049-5
17. Cohen AL, Sidlowski S, Staub A. Beliefs and Bayesian reasoning. *Psychon Bull Rev*. 2017;24(3):972-978. Disponible en: <https://link.springer.com/article/10.3758/s13423-016-1161-z>
18. Goodman SN. Toward evidence-based medical statistics. 2: the Bayes factor. *Ann Intern Med*. 1999;130(12):1005-1013.
19. Stern HS. A test by any other name: P Values, Bayes factors, and statistical inference. *Multivariate Behav Res*. 2016;51(1):23-29. DOI: 10.1080/00273171.2015.1099032
20. Miller G, Martz H, Little T, Bertelli L. Bayesian hypothesis testing—use in interpretation of measurements. *Health Phys*. 2008;94(3):248-254. DOI: 10.1097/01.HP.0000290624.35701.00
21. Bittl JA, He Y. Bayesian analysis: a practical approach to interpret clinical trials and create clinical practice guidelines. *Circ Cardiovasc Qual Outcomes*. 2017;10(8):e003563. DOI: 10.1161/CIRCOUTCOMES.117.003563
22. Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers EJ. The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev*. 2016;23(1):103-123. DOI: 10.3758/s13423-015-0947-8
23. Van-Rijn MHC, Bech A, Bouyer J, Van-Den-Brand JA JG. Statistical significance versus clinical relevance. *Nephrol Dial Transplant*. 2017;32(Suppl 2):ii6–ii12. DOI: 10.1093/ndt/gfw385
24. Matthews RAJ. Beyond “significance”: principles and practice of the analysis of credibility. *R Soc Open Sci*. 2018;5(1):171047. DOI: 10.1098/rsos.171047
25. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev*. 2009;16(2):225-237. DOI: 10.3758/PBR.16.2.225
26. Lecoutre B, Derzko G, Grouin JM. Bayesian predictive approach for inference about proportions. *Stat Med*. 1995;14(9-10):1057-1063.

27. Zaslavsky BG. Bayesian versus frequentist hypotheses testing in clinical trials with dichotomous and countable outcomes. *J Biopharm Stat.* 2010;20(5):985-997. DOI: 10.1080/10543401003619023
28. Martínez-Ezquerro JD, Riojas-Garza A, Rendón-Macías ME. Clinical significance vs statistical significance. How to interpret the confidence interval at 95. *Rev Alerg Mex.* 2017;64(4):477-486. DOI: 10.29262/ram.v64i4.334
29. Rottman BM. Physician Bayesian updating from personal beliefs about the base rate and likelihood ratio. *Mem Cognit.* 2017;45(2):270-280. DOI: 10.3758/s13421-016-0658-z
30. López-Puga J, Krzywinski M, Altman N. Points of significance: Bayesian statistics. *Nat Methods.* 2015;12(5):377-378.
31. Gurrin LC, Kurinczuk JJ, Burton PR. Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. *J Eval Clin Pract.* 2000;6(2):193-204. DOI: 10.1046/j.1365-2753.2000.00216.x
32. Wagenmakers EJ, Beek TF, Rotteveel M, Gierholz A, Matzke D, Steingroever H, et al. Turning the hands of time again: a purely confirmatory replication study and a Bayesian analysis. *Front Psychol.* 2015;6:494. DOI: 10.3389/fpsyg.2015.00494
33. Greenland S, Poole C. Living with p values: resurrecting a Bayesian perspective on frequentist statistics. *Epidemiology.* 2013;24(1):62-68. DOI: 10.1097/EDE.0b013e3182785741
34. Bertram T. Significance testing—*are we ready yet to abandon its use?* *Curr Med Res Opin.* 2011;27(11):2087-2089.