

Análisis Bayesiano de un ensayo controlado aleatorizado: aspectos matemáticos y clínicos

Bayesian analysis of a randomized controlled trial: mathematical and clinical aspects

Jaime Cerda¹, Ricardo Castro², Claudia Bambs^{1,3,4} y María Elvira Balcells⁵

¹Escuela de Salud Pública, Pontificia Universidad Católica de Chile, Santiago, Chile.

²Departamento de Medicina Intensiva. Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile.

³Centro Avanzado de Enfermedades Crónicas (ACCDiS), Fondap 1513001, Santiago, Chile.

⁴Centro para la Prevención y el Control del Cáncer (CECAN), Fondap 152220002, Santiago, Chile.

⁵Departamento de Enfermedades Infecciosas del Adulto. Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile.

Conflictos de interés: Los autores declaran no presentar conflictos de interés.

Financiamiento: CB declara recibir financiamiento de los Centros de Investigación en Áreas Prioritarias Fondap 1513001 y 152220002.

Recibido: 6 de marzo de 2025

Artículo por invitación

Resumen

El proceso clínico de diagnóstico y la selección de tratamientos se basa en razonamientos que emplean inferencias lógicas y científicas para llegar a conclusiones y tomar decisiones. Estos razonamientos se fundamentan en bases estadísticas sólidas, dentro de las cuales operan dos enfoques principales: el frecuentista y el Bayesiano. El enfoque frecuentista es el más utilizado y se centra en la prueba de significancia de la hipótesis nula y el valor-p. En los últimos años, la evaluación de la evidencia científica sobre la eficacia de las intervenciones mediante este enfoque ha generado conclusiones controvertidas. Como complemento, el análisis Bayesiano de ensayos controlados aleatorizados, basado en la actualización de hipótesis mediante los resultados de estudios clínicos, ofrece una perspectiva más intuitiva y acorde con la práctica clínica, lo que puede dar lugar a conclusiones distintas a las obtenidas con el enfoque frecuentista. Este artículo educativo examina los fundamentos matemáticos y clínicos del análisis Bayesiano en ensayos controlados aleatorizados, ilustrando su aplicación mediante un ejemplo real.

Palabras clave: ensayo controlado aleatorizado; análisis Bayesiano; teorema de Bayes; bioestadística.

Abstract

The clinical process of diagnosis and treatment selection is based on reasoning that employs logical and scientific inferences to reach conclusions and make decisions. These reasonings are grounded in solid statistical foundations, within which two main approaches operate: the frequentist and the Bayesian. The frequentist approach is the most commonly used and focuses on null hypothesis significance testing and the p-value. In recent years, the evaluation of scientific evidence on the efficacy of interventions using this approach has led to controversial conclusions. As a complement, Bayesian analysis of randomized controlled trials, based on hypothesis updating through the results of clinical studies, offers a more intuitive perspective that aligns with clinical practice, potentially leading to different conclusions from those derived from the frequentist approach. This educational article examines the mathematical and clinical foundations of Bayesian analysis in randomized controlled trials, illustrating its application with a real-world example.

Keywords: randomized controlled trial; Bayesian analysis; Bayes theorem; biostatistics.

Correspondencia a:

Jaime Cerda Lorca

jcerdal@uc.cl

Introducción

*La vida es corta, el arte largo, la ocasión fugaz,
la experiencia insegura, el juicio difícil.*

(Hipócrates. Aforismos, número 1).

Durante su formación profesional, los médicos adquieren una amplia gama de conocimientos y habilidades, siendo especialmente relevantes la capacidad de diagnosticar enfermedades y administrar tratamientos. El éxito en ambas depende en gran medida de la destreza técnica del profesional, pero también de su experiencia, intuición y sentido común. No es casualidad que, hace más de 24 siglos, Hipócrates considerara la medicina como una combinación de ciencia y arte.

Al profundizar en los aspectos clínicos del diagnóstico y la terapéutica, los médicos suelen centrarse en procedimientos e intervenciones, pero rara vez reflexionan sobre los procesos mentales que guían dicho razonamiento ni sobre los fundamentos estadísticos que sustentan ambas actividades. Podemos afirmar que, para diagnosticar, los médicos operan bajo un paradigma estadístico típicamente Bayesiano, mientras que, para evaluar la evidencia sobre la eficacia de intervenciones, especialmente en ensayos controlados aleatorizados, adoptan un enfoque estadístico clásicamente frecuentista.

En los últimos años hemos sido testigos de un fenómeno interesante: la evaluación de la evidencia científica sobre la eficacia de intervenciones bajo el paradigma frecuentista ha dado lugar a conclusiones controvertidas. Esto se debe en gran medida a su dependencia de la prueba de significancia de la hipótesis nula (*null hypothesis significance testing*) y del valor-p, a menudo malinterpretado y criticado por su excesiva dependencia del tamaño muestral, así como el *p-hacking* o práctica de manipular los datos o el análisis estadístico de una investigación para obtener un valor-p estadísticamente significativo, entre otros argumentos¹. De este modo, muchos ensayos controlados aleatorizados logran demostrar el efecto de una intervención, pero no logran alcanzar significancia estadística, lo que lleva a descartar su uso en la práctica clínica, privándonos de alternativas terapéuticas potencialmente beneficiosas. Lo mismo puede ocurrir a la inversa, cuando se demuestra el efecto pequeño y clínicamente irrelevante de una intervención, el cual logra alcanzar significancia estadística debido al uso de una muestra de gran tamaño, siendo la intervención ampliamente recomendada. Ante este escenario cobra relevancia el análisis Bayesiano de los ensayos controlados aleatorizados, el cual adopta una perspectiva más intuitiva y alineada con la práctica clínica, permitiendo evaluar la eficacia de las intervenciones de manera distinta y, con frecuencia, conduce a conclusiones diferentes a

las obtenidas a través del análisis frecuentista²⁻⁴. Existen varios casos emblemáticos en la literatura científica que ilustran esta diferencia, así como voces que han señalado los problemas del uso de la estadística frecuentista en la ciencia contemporánea⁵.

Este artículo, de carácter educativo, explora conceptos estadísticos clave y explica el proceso matemático subyacente al análisis Bayesiano de un ensayo controlado aleatorizado, ilustrando su aplicación mediante un ejemplo real.

¿De qué manera la actividad clínica se relaciona con la estadística frecuentista y Bayesiana?

Casi todos los médicos, durante su formación, realizan algún curso de bioestadística, generalmente impartido por docentes del área de las matemáticas. La bioestadística que se enseña en estos cursos es de tipo frecuentista; de hecho, para muchos médicos puede resultar sorprendente descubrir, a través de estas líneas, la existencia de un enfoque estadístico diferente, conocido como estadística Bayesiana. Este enfoque se basa en el teorema propuesto por el reverendo presbiteriano y matemático inglés Thomas Bayes (1702-1761), también conocido como teorema de la probabilidad condicional.

En términos generales, la estadística frecuentista se enfoca en la frecuencia o proporción de los eventos; considera la existencia de parámetros fijos y desconocidos (los parámetros son valores utilizados para caracterizar una población, pero que no se conocen directamente) y descansa en el valor-p para la toma de decisiones, el cual indica la probabilidad de observar los resultados de un estudio bajo el supuesto de hipótesis nula verdadera. Por el contrario, la estadística Bayesiana se enfoca en la actualización de creencias previas acerca de un parámetro a medida que se obtiene nueva información; trata los parámetros como variables aleatorias con distribuciones de probabilidad y actualiza las creencias previas a medida que se obtiene nueva evidencia (Tabla 1). El debate sobre la investigación clínica frecuentista y Bayesiana puede remontarse a la antigüedad grecorromana y al discurso entre los racionalistas (o frecuentistas) y los empiristas (o Bayesianos)⁶.

La inferencia Bayesiana se encuentra plenamente vigente y está en sintonía con la actividad clínica, por cuanto se basa en la actualización de nuestras creencias o hipótesis utilizando la evidencia que aportan estudios clínicos. Su mayor uso ocurre en el proceso de diagnóstico clínico, operando de la siguiente manera: en un primer paso, frente a un paciente, basándose en la anamnesis y la información obtenida del examen físico, el médico estima la probabilidad de que el paciente presente el diagnóstico X. Esta probabilidad se denomina probabilidad pre-test, ya que se establece antes de conocer los resultados de una prueba diagnóstica. En un segundo paso, estando en

Tabla 1. Diferencias conceptuales entre los paradigmas frecuentista y Bayesiano (adaptado de (2)).

	Paradigma frecuentista	Paradigma Bayesiano
Enfoque principal	Se centra en la frecuencia esperada de observar los datos bajo la hipótesis nula, para basar la inferencia científica en datos supuestamente objetivos.	Se centra en la actualización de creencias sobre un parámetro conforme se obtiene nueva información.
Parámetros	Los parámetros son fijos y desconocidos.	Los parámetros son tratados como variables aleatorias con distribuciones de probabilidad.
Probabilidad	Se interpreta como la frecuencia de eventos esperados en experimentos repetidos.	Se interpreta como el grado de creencia sobre una hipótesis o parámetro.
Pregunta para inferencia estadística	¿Cuál es la probabilidad de observar los datos, dada la hipótesis nula?	¿Cuál es la probabilidad de la hipótesis nula, dado los datos observados?
Inferencia intervalar	Utiliza intervalos de confianza generados a partir de los datos de una muestra, basándose en repeticiones hipotéticas de un mismo estudio.	Utiliza intervalos de credibilidad, definidos como un rango plausible de valores para el parámetro de interés.
Valor-p	Utiliza el valor-p, que evalúa la evidencia asumiendo la veracidad de la hipótesis nula.	No se utiliza el valor-p.
Muestra y tamaño	El tamaño de la muestra influye en la precisión de los resultados.	El tamaño de la muestra influye, pero se combina con la información previa (<i>prior</i>) para mejorar la estimación.

una situación de incertidumbre respecto al diagnóstico, el médico solicita una prueba diagnóstica (examen de laboratorio, estudio de imagen, cuestionario, etc). Por último, en un tercer paso, recalcula la probabilidad de que el paciente presente el diagnóstico X, actualizándola a partir de la información que aporta la prueba diagnóstica. Esta nueva probabilidad se conoce como probabilidad post-test y es utilizada para confirmar o descartar el diagnóstico inicialmente planteado. Matemáticamente, la prueba diagnóstica aporta el llamado Factor de Bayes (*o likelihood ratio*), correspondiente a un cociente entre la probabilidad condicional de obtener un determinado resultado en la prueba diagnóstica dada la certeza que se está enfermo y la probabilidad condicional de obtener el mismo resultado en la prueba diagnóstica dada la certeza que se está sano. El Factor de Bayes permite calcular la probabilidad post-test de diagnóstico, correspondiente a la probabilidad condicional de estar enfermo o sano, dado que se obtuvo un determinado resultado en la prueba diagnóstica. Para mayor abundamiento, sugerimos revisar el artículo de Cerda⁷, publicado en esta misma revista.

A diferencia del proceso clínico diagnóstico, la evaluación de la evidencia científica sobre la eficacia de intervenciones se basa en el análisis crítico de ensayos controlados aleatorizados, utilizando tradicionalmente un enfoque frecuentista. Supongamos que se compara la eficacia de una intervención frente a un placebo. Las medidas de efecto más comúnmente empleadas son el riesgo relativo (RR), el *odds ratio* (OR) y la reducción absoluta de riesgo (RAR). El fundamento probatorio se basa en la prueba de significancia de la hipótesis nula, que involucra la formulación de una hipótesis nula (H_0)

y una hipótesis alternativa (H_1) respecto al parámetro que se desea estimar. La hipótesis nula plantea que no existe efecto o relación entre las variables estudiadas: en un ensayo controlado aleatorizado podría postular que un tratamiento no tiene efecto sobre una enfermedad, es decir, que no hay diferencia entre el tratamiento y el placebo (H_0 : RR = 1,0). Por otro lado, la hipótesis alternativa sostiene que sí existe tal efecto, diferencia o relación (H_1 : RR $\neq$ 1,0). La prueba de significancia de la hipótesis nula tiene como objetivo determinar si los resultados del estudio son lo suficientemente improbables asumiendo la veracidad de H_0 como para justificar su rechazo. Si la probabilidad de obtener los resultados del estudio asumiendo la veracidad de la hipótesis nula es baja (generalmente inferior al 5%, por consenso de la comunidad científica) se rechaza dicha hipótesis y se acepta la hipótesis alternativa. Si no se rechaza la hipótesis nula, se concluye que no hay evidencia suficiente para aceptar la hipótesis alternativa.

Una de las principales limitaciones de la prueba de significancia de la hipótesis nula es que el valor-p no nos entrega información sobre la probabilidad de que una hipótesis acerca del valor del parámetro sea verdadera, lo cual es precisamente lo que nos interesa conocer. Este concepto cobra especial relevancia al examinar la diferencia entre el intervalo de confianza frecuentista y su equivalente Bayesiano, conocido como intervalo de credibilidad. En estadística frecuentista, las medidas de efecto se acompañan generalmente de un intervalo de confianza, típicamente de 95%. Este intervalo de confianza representa el rango de valores, calculado a partir de los datos de una muestra, que podría adoptar la medida de efecto si repitiésemos hipotéticamente el

estudio múltiples veces bajo las mismas condiciones. Por ejemplo, si una muestra presenta un riesgo relativo de 4, con un intervalo de confianza de 95% entre 2 y 6, significa que, en múltiples repeticiones hipotéticas del estudio, 95 de cada 100 presentarían un riesgo relativo entre 2 y 6. No es raro encontrar la siguiente explicación para el intervalo de confianza de 95%: *existe una probabilidad de 95% de que dicho intervalo contenga el verdadero valor del parámetro poblacional*. En estricto rigor, el intervalo de confianza no garantiza lo que se sugiere en esa explicación, ya que el intervalo de confianza frecuentista solo refleja la variabilidad o dispersión de los resultados cuando se obtienen diferentes muestras de una misma población, debido a la aleatoriedad inherente al proceso de muestreo.

A diferencia del intervalo de confianza frecuentista, el intervalo de credibilidad Bayesiano (ICr) sí contiene el valor del parámetro que se desea conocer, con una probabilidad específica. Un intervalo de credibilidad de 95% significa que, considerando la información previa y los datos observados a partir de un estudio, existe un 95% de probabilidad de que el parámetro efectivamente se encuentra dentro de dicho intervalo. La diferencia de interpretación entre el intervalo de confianza frecuentista y el intervalo de credibilidad Bayesiano puede parecer sutil; sin embargo, es de máxima importancia y es uno de los aspectos más atractivos de la estadística Bayesiana.

¿Cómo se realiza el análisis Bayesiano de un ensayo controlado aleatorizado?

Comenzaremos por describir un ensayo controlado aleatorizado real. Con el objetivo de evaluar si la terapia intravenosa con vitamina C reduce la mortalidad a los 28 días en pacientes con *shock* séptico, Wacker⁸ llevó a cabo un estudio multicéntrico, doble ciego, aleatorizado y controlado en cinco unidades de cuidados intensivos. Un total de 124 pacientes adultos, en las primeras 24 horas de iniciada la administración de vasopresores por *shock* séptico, participaron en el estudio: 60 recibieron vitamina C intravenosa (10 mg/mL en solución salina normal, administrada como un bolo de 1.000 mg durante 30 minutos seguido de una infusión continua de 250 mg/h

durante 96 horas) y 64 recibieron placebo. El resultado principal fue la mortalidad a los 28 días por todas las causas, presente en (16/60) = 26,7% de los pacientes que recibieron vitamina C y en (26/64) = 40,6% de los pacientes que recibieron placebo, siendo el valor-p = 0,10. Al ser este valor mayor a 5%, se afirma que no alcanzó significancia estadística. Los autores concluyeron lo siguiente, traducido literalmente: *la monoterapia con vitamina C no logró reducir significativamente la mortalidad en pacientes con shock séptico, como se había hipotetizado. Nuestros hallazgos no apoyan su uso clínico rutinario para este propósito*.

Para efectuar el análisis Bayesiano de un ensayo controlado aleatorizado sugerimos comenzar por expresar los resultados del estudio en forma de OR, medida de efecto que será necesaria para una serie de cálculos posteriores, siendo equivalente a $(16/44)/(26/38) = 0,53$ (nótese que dicho OR es compatible con un efecto favorable a la vitamina C, en comparación al placebo). Toda vez que analizamos un desenlace adverso (e.g. muerte), dado que el OR es una medida de efecto asimétrica, es decir, adopta valores entre 0 y 1 cuando la intervención es protectora y entre 1 e infinito cuando es perjudicial, resulta conveniente su transformación logarítmica, ya que $\ln(\text{OR})$ adopta valores entre infinito negativo y cero cuando la intervención es protectora y entre 0 e infinito positivo cuando es perjudicial, siendo su valor nulo igual a cero. En el estudio analizado, $\ln(\text{OR})$ equivale a -0,63. De manera complementaria, calculamos la RAR, siendo equivalente a $(26/64)-(16/44) = -14,0\%$ (Tabla 2).

Conceptualmente, el análisis Bayesiano de un ensayo controlado aleatorizado consta de tres partes: un *prior* (i.e. distribución previa de tamaños de efecto creíbles), una verosimilitud (i.e. grado en que los resultados de un estudio actualizan el *prior*) y una distribución posterior. El *prior* se combina con la verosimilitud para generar la distribución posterior, la cual representa una estimación actualizada de la distribución de tamaños de efecto creíbles⁹. Estas tres partes están representadas en la siguiente ecuación, derivada del teorema de Bayes:

Odds a priori de la hipótesis alternativa	x	Factor de Bayes (FB_{10})	=	Odds posterior de la hipótesis alternativa
$\frac{P(H1)}{P(H0)}$	x	$\frac{P(data H1)}{P(data H0)}$	=	$\frac{P(H1 data)}{P(H0 data)}$

El *odds a priori* de la hipótesis alternativa corresponde al *odds* de que la hipótesis alternativa (generalmente H_1 : $OR \neq 1,0$) sea verdadera previo a realizar el estudio y conocer sus resultados, con respecto a la hipótesis nula (H_0 : $OR = 1,0$). Refleja la creencia inicial del investigador sobre la veracidad de la hipótesis alternativa,

Tabla 2. Resultados del ensayo controlado aleatorizado de Wacker y cols (8).

Intervención	Mortalidad a los 28 días, por todas las causas		Total
	Presente	Ausente	
Vitamina C intravenosa	16 (26,7%)	44 (73,3%)	60 (100%)
Placebo	26 (40,6%)	38 (58,4%)	64 (100%)

El valor-p de la tabla equivale a 0,10. El *odds ratio* (OR) corresponde a $(16/44)/(26/38) = 0,53$ siendo su IC95% 0,25 a 1,14. El $\ln(\text{OR})$ equivale a -0,63 y su IC95% -1,39 a 0,13. La reducción absoluta de riesgo corresponde a $(26/38)-(16/44) = -14,0\%$ siendo su IC95% -30,4% a 2,47%.

planteada en base a experiencia previa, conocimiento experto, fuentes de información relevantes y/o racionalidad. Por su parte, los resultados del estudio influyen en el cálculo del Factor de Bayes (FB), el cual indica cuántas veces son más probables los resultados del estudio bajo la hipótesis alternativa en lugar de la hipótesis nula. Este factor actúa como un mecanismo para actualizar las creencias previas. Por ejemplo, si el $FB_{10} = 7$ significa que los resultados observados son 7 veces más probables bajo la hipótesis alternativa que bajo la hipótesis nula, lo que sugiere una mayor evidencia a favor de H_1 en lugar de H_0 . La fuerza con que el Factor de Bayes apoya las hipótesis motivó a Jeffreys, en 1961, a clasificar distintos puntos de corte; por ejemplo, un FB_{10} mayor a 10 aporta evidencia fuerte a favor de H_1 , un FB_{10} menor a 0,1 aporta evidencia fuerte a favor de H_0 y un FB_{10} igual a 1,0 no aporta evidencia. Finalmente, el *odds* posterior de la hipótesis alternativa corresponde a la actualización de la creencia inicial, teniendo en consideración los resultados del estudio.

El planteamiento del *odds a priori* de la hipótesis alternativa requiere definir distintos *priors*. En el análisis Bayesiano, un *prior* (o distribución *a priori*) es una distribución de probabilidad que refleja el conocimiento o las creencias previas sobre un parámetro antes de observar los datos. Este componente es esencial en la inferencia Bayesiana, ya que permite incorporar la información previa disponible sobre un parámetro, basándose en experiencia previa, conocimiento experto, fuentes relevantes y/o racionalidad. Un *prior* se representa como una distribución de valores de $\ln(OR)$ con una media (μ) y una desviación estándar (σ), asumiendo una distribución normal.

En este ejemplo, planteamos un total de seis *priors*, intentando cubrir un amplio espectro de creencias previas, desde muy favorables a la intervención hasta muy desfavorables a ésta. Es importante señalar que estos *priors* no estuvieron basados en experiencia previa, conocimiento experto, fuentes relevantes y/o racionalidad, sino más bien intentaron cubrir un amplio rango de posibles OR. Los *priors* utilizados fueron los siguientes:

1. *Prior no informativo*: su valor de $\ln(OR)$ tiene una media de 0 y desviación estándar de 2, correspondiente a un OR de 1,0 (ICr 95% 0,02-50,40). Este *prior* considera que todos los posibles valores de OR son igualmente plausibles.
2. *Prior neutral*: su valor de $\ln(OR)$ tiene una media de 0 y desviación estándar de 0,4, correspondiente a un OR de 1,0 (ICr 95% 0,46 a 2,19). Este *prior* plantea una ausencia de efecto, con una misma probabilidad de beneficio (50%) o perjuicio (50%) de la intervención. La desviación estándar de 0,4 corresponde al valor aproximado del error estándar obtenido en el estudio de Wacker⁸, siendo utilizada en todos los siguientes *priors*.

3. *Prior muy optimista, favorable a la intervención*: su valor de $\ln(OR)$ tiene una media de -0,693 y desviación estándar de 0,4, correspondiente a un OR de 0,50 (ICr 95% 0,34 a 1,64). Su valor de OR es cercano al OR reportado por el estudio de Wacker⁸.
4. *Prior levemente optimista, favorable a la intervención*: su valor de $\ln(OR)$ tiene una media de -0,288 y desviación estándar de 0,4, correspondiente a un OR de 0,75 (ICr 95% 0,34 a 1,64).
5. *Prior levemente pesimista, favorable al placebo*: su valor de $\ln(OR)$ tiene una media de 0,288 y desviación estándar de 0,4, correspondiente a un OR de 1,33 (ICr 95% 0,61 a 2,92).
6. *Prior muy pesimista, favorable al placebo*: su valor de $\ln(OR)$ tiene una media de 0,693 y desviación estándar de 0,4, correspondiente a un OR de 2,00 (ICr 95% 0,91 a 4,38).

El siguiente paso consistió en calcular el Factor de Bayes, el que, a diferencia de los estudios de exactitud diagnóstica, no puede calcularse manualmente debido a su complejidad, requiriendo de programas estadísticos especializados. En este caso, se utilizó el software estadístico JASP¹⁰, el cual ofrece una interfaz gráfica de usuario e incluye el menú *Bayesian A/B Test*. Los datos ingresados fueron los siguientes: 1) número total de pacientes que recibieron la intervención ($n = 60$); 2) número de pacientes que recibieron la intervención y que presentaron resultado exitoso, es decir, no fallecieron en los primeros 28 días ($n = 44$); 3) número total de pacientes que recibieron placebo ($n = 64$); y 4) número de pacientes que recibieron placebo y que presentaron resultado exitoso ($n = 38$).

Para cada *prior* el software estadístico calculó la distribución posterior del $\ln(OR)$, OR, RR y RAR, además de sus respectivos intervalos de credibilidad 95%. Asimismo, proporcionó la probabilidad posterior de beneficio ($OR < 1,0$) (Tabla 3, Figuras 1, 2 y 3). A continuación se presenta el análisis matemático para el *prior levemente optimista*, como ejemplo docente:

- En primer lugar, se planteó una creencia previa respecto al resultado del estudio bajo el *prior levemente optimista*, correspondiente a un OR de 0,75, siendo éste favorable a la intervención. La forma en que se introdujo este valor al software estadístico es bajo el supuesto que el $\ln(OR)$ corresponde a una distribución normal, cuya media (μ) es -0,288 y desviación estándar (σ) es 0,4.
- En base a la información recientemente aportada y los resultados del estudio, el *software* estadístico calculó la mediana posterior del $\ln(OR)$ y OR, y sus respectivos intervalos de credibilidad 95%. La mediana posterior del $\ln(OR)$ fue -0,459, con un ICr 95% de -0,997 a 0,079. En el caso del OR, el resultado del

Tabla 3. Resultados del re-análisis Bayesiano del ensayo controlado aleatorizado de Wacker y cols.⁸

Definición del prior, expresado como ln(OR)	Efecto estimado de la vitamina C intravenosa versus placebo			
	Desenlace: mortalidad 28-d			
	OR a priori (CrI 95%)	Probabilidad a priori de beneficio (OR < 1,0)	OR posterior* (CrI 95%)	Probabilidad posterior de beneficio (OR < 1,0)
No informativo ~N(0 ; 2,0)	1,00 (0,02 a 50,40)	50,0%	0,54 (0,26 a 1,13)	94,7%
Neutral ~N(0 ; 0,4)	1,00 (0,46 a 2,19)	50,0%	0,72 (0,42 a 1,23)	89,0%
Muy optimista ~N(-0,693 ; 0,4)	0,50 (0,23 a 1,10)	95,8%	0,52 (0,30 a 0,88)	99,2%
Levemente optimista ~N(-0,288 ; 0,4)	0,75 (0,34 a 1,64)	76,4%	0,63 (0,37 a 1,08)	95,3%
Levemente pesimista ~N(0,288 ; 0,4)	1,33 (0,61 a 2,92)	23,6%	0,83 (0,48 a 1,42)	75,4%
Muy pesimista ~N(0,693 ; 0,4)	2,00 (0,91 a 4,38)	4,2%	1,00 (0,58 a 1,71)	50,7%

OR: odds ratio. Se consideró que la distribución del ln(OR) sigue una distribución normal, con media (μ) y desviación estándar (σ). *Mediana y su intervalo de credibilidad al 95%

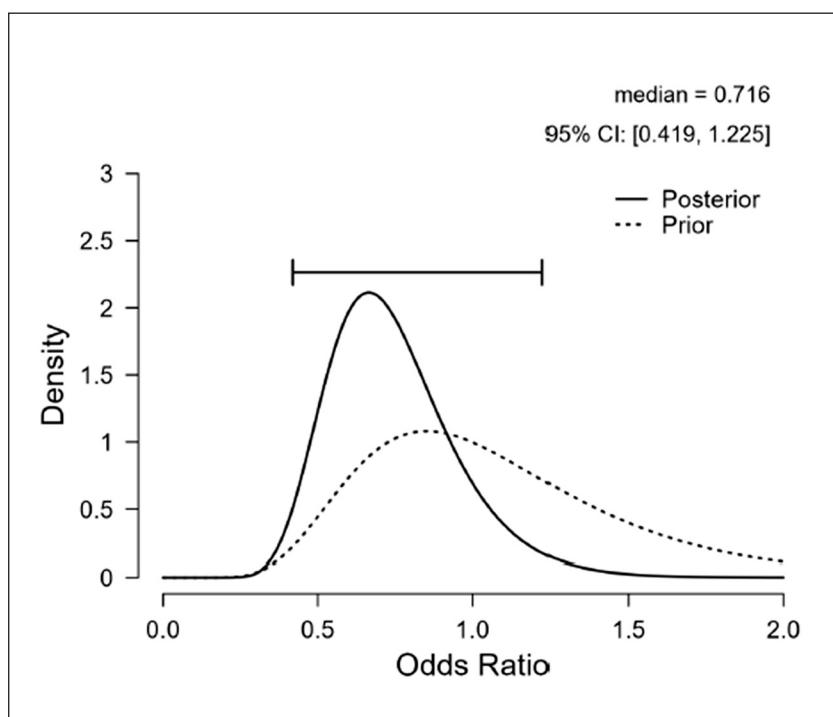


Figura 1. Gráfico de la distribución previa (*prior*, línea punteada) y posterior (línea continua) del OR bajo el *prior neutral*, para el efecto de la vitamina C intravenosa, en comparación con el placebo, sobre la mortalidad a los 28 días. La probabilidad posterior de beneficio (OR < 1,0) es 89,0%. Median: mediana OR posterior; 95%CI: intervalo de credibilidad 95%.

estudio modificó la creencia previa, pasando de un OR *a priori* de 0,75 a un OR posterior de 0,63 (ICr 95% 0,37 a 1,08). Asimismo, el *software* reportó que la probabilidad posterior de beneficio (OR < 1,0) pasó de 76,4 a 95,3%.

Con excepción del *prior muy pesimista*, en todos los demás escenarios planteados el valor del OR posterior fue inferior a 1,0, resultado que favorece la vitamina C por sobre el placebo. La probabilidad posterior de beneficio (OR < 1,0) en estos escenarios fue superior a 75%, alcanzando su valor más alto (como se esperaba) para el *prior muy optimista* (99%). Es importante destacar que incluso en el caso de un escenario que considere un *prior levemente pesimista* (OR 1,33), el OR posterior favoreció la intervención por sobre el placebo (OR 0,83), siendo su probabilidad posterior de beneficio (OR < 1,0) equivalente a 75,4%.

¿Qué conclusiones y reflexiones obtenemos de este análisis Bayesiano?

Lo primero, y posiblemente sea lo más importante, es que la conclusión planteada originalmente por los autores del estudio, basada en un análisis frecuentista, podría cambiar. Con respecto a la intervención, los autores señalaron que “*nuestros hallazgos no apoyan su uso clínico rutinario para este propósito*”; no obstante,

a la luz del análisis Bayesiano, resulta razonable plantear una conclusión que considere el uso de la vitamina C, pues incluso en un escenario levemente pesimista los resultados son alentadores: la probabilidad posterior de beneficio ($OR < 1,0$), excepto para el escenario muy pesimista, es de 75% o más. Para dimensionar el posible beneficio de la intervención resulta conveniente interpretar la RAR obtenida en el estudio (RAR -14%): por cada 100 pacientes con *shock* séptico a quienes se agrega la administración de vitamina C intravenosa, en lugar de placebo, se evitan 14 muertes extras. Este tamaño de efecto resulta muy relevante, toda vez que el desenlace evitado es de máxima importancia (muerte de pacientes) y la intervención pareciera ser inocua, de bajo costo y asequible.

Ahora bien, desde una mirada escéptica o adversarial, es posible argumentar que los resultados obtenidos bajo el *prior muy pesimista* (cuyo OR posterior fue 1,00 y probabilidad posterior de beneficio ($OR < 1,0$) fue 50,7%) están en sintonía con la conclusión originalmente formulada en base al análisis frecuentista, restando apoyo al uso clínico rutinario de la vitamina C. Sin embargo, parece razonable cuestionar la idoneidad de dicho *prior*: la más reciente revisión sistemática con metaanálisis de ensayos controlados aleatorizados que evaluó la eficacia del uso de vitamina C versus placebo en el manejo de pacientes con *shock* séptico (6 estudios, 1.390 pacientes)¹¹ reportó un OR de 0,71 (IC95% 0,45 a 1,12) para el desenlace mortalidad a los 28 días, y ninguno de los seis estudios primarios presentaron valores de OR como el considerado por el *prior muy pesimista*; de hecho, cinco de ellos presentaron OR inferior a 1,0. De especial interés es el estudio LOVIT de Lamontagne y cols.¹², por ser el de mayor tamaño y el único que presentó un OR superior a 1,0 (OR 1,19, IC95% 0,90 a 1,58). Coincidentemente, los resultados originales de dicho estudio, formulados bajo un enfoque frecuentista, fueron reanalizados Bayesianamente por Angriman y cols.¹³. Utilizando diferentes *priors*, los autores de este re-análisis Bayesiano arribaron a una conclusión similar a la planteada en el estudio original, basada en un análisis frecuentista.

Wacker y cols.⁸, condicionaron de manera importante la conclusión de su estudio en la obtención de un valor-p que no alcanzó significancia estadística (es decir, fue mayor a 5%, específicamente 10%). Matemáticamente, esto se explica simplemente por una falta de potencia de la muestra utilizada, equivalente a 37,1% (por convención, la potencia mínima de una muestra debe ser 80%). Para el cálculo de tamaño muestral, los autores anticiparon una reducción absoluta de mortalidad de al menos 20%, basándose en un único estudio previo que mostraba una reducción de 30%. Sin embargo, los resultados obtenidos arrojaron una reducción de 14%, siendo menor a la anticipada; por consiguiente, el tamaño muestral calcu-

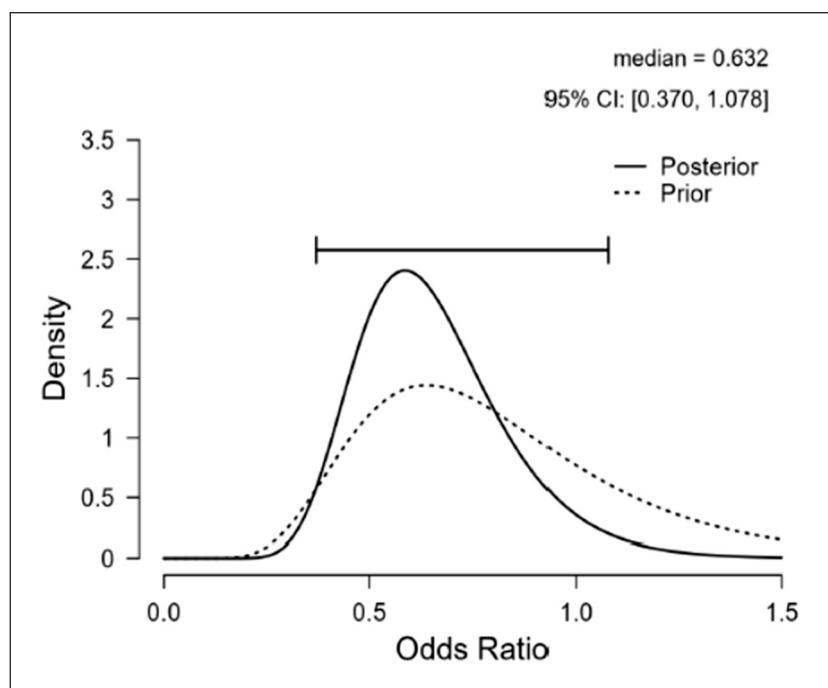


Figura 2. Gráfico de la distribución previa (*prior*, línea punteada) y posterior (línea continua) del OR bajo el *prior levemente optimista*, para el efecto de la vitamina C intravenosa, en comparación con el placebo, sobre la mortalidad a los 28 días. La probabilidad posterior de beneficio ($OR < 1,0$) es 95,3%. Median: mediana OR posterior; 95%CI: intervalo de credibilidad 95%..

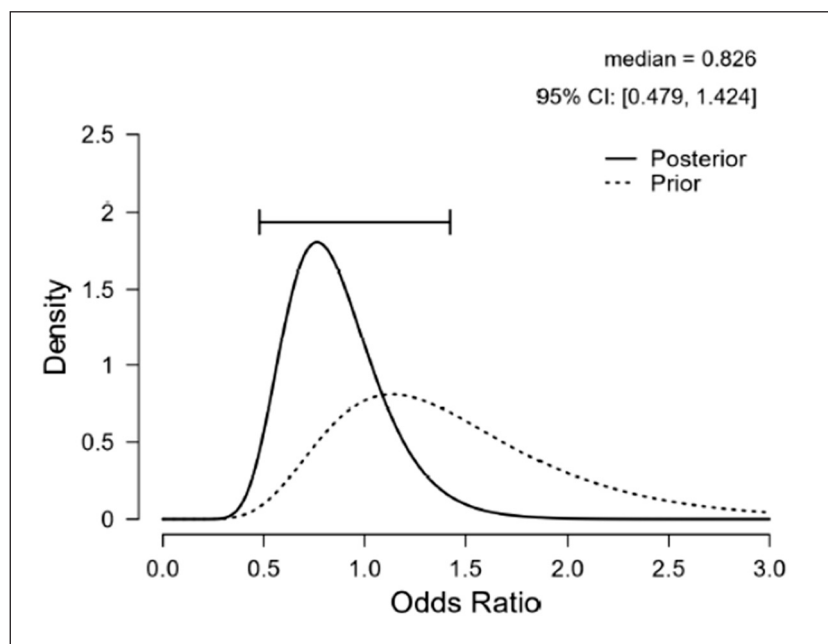


Figura 3. Gráfico de la distribución previa (*prior*, línea punteada) y posterior (línea continua) del OR bajo el *prior levemente pesimista*, para el efecto de la vitamina C intravenosa, en comparación con el placebo, sobre la mortalidad a los 28 días. La probabilidad posterior de beneficio ($OR < 1,0$) es 75,4%. Median: mediana OR posterior; 95%CI: intervalo de credibilidad 95%..

lado fue insuficiente para que esta reducción absoluta de mortalidad alcanzara significancia estadística. La existencia de muestras con potencia estadística insuficiente es una causa muy frecuente de estudios que pese a demostrar la eficacia de una intervención, no logran significancia estadística. Si bien en el análisis Bayesiano el tamaño muestral tiene importancia, éste no condiciona la conclusión del mismo en la medida que ocurre en el análisis frecuentista.

En comparación con su equivalente frecuentista, el análisis Bayesiano ofrece una serie de ventajas clave. En primer lugar, se ajusta mejor al acto clínico, ya que se basa en la experiencia, conocimientos previos, razonamiento y creencias debidamente fundamentadas sobre los pacientes y sus enfermedades, siendo actualizadas a medida que se incorpora la información proveniente de estudios clínicos. Por el contrario, el enfoque frecuentista sigue una lógica diferente: se basa en el supuesto de repetir los estudios múltiples veces, algo que en la práctica no ocurre. Por otro lado, el análisis Bayesiano permite llegar a conclusiones con menos pacientes, hecho de singular importancia para la toma de decisiones en enfermedades poco frecuentes. Desde un punto de vista matemático, el enfoque Bayesiano ofrece una ventaja clara: nos proporciona, con un determinado nivel de credibilidad, la probabilidad de que el parámetro de interés se encuentre dentro de un intervalo específico, algo que el enfoque frecuentista, en términos estrictos, no puede hacer.

Ahora bien, el enfoque Bayesiano no es perfecto. Como hemos explicado, este análisis descansa fuertemente en la idoneidad de los *priors* planteados. Este hecho es fundamental y posiblemente constituya su principal vulnerabilidad y fuente de controversia^{14,15}, pues resulta razonable afirmar que en el análisis Bayesiano de un ensayo controlado aleatorizado se podría incluir deliberadamente valores de *priors* que permitan obtener resultados preconcebidos, conducta que va en sentido contrario de lo que pretende la investigación clínica y el acto médico propiamente tal. En este punto es necesario distinguir dos situaciones diferentes: el análisis Bayesiano primario de un estudio (cuyos *priors* son definidos antes de ejecutar el estudio) versus el re-análisis Bayesiano de un estudio (cuyos resultados y conclusiones originales se basan en estadística frecuentista, y el análisis Bayesiano con sus respectivos *priors* son definidos habiendo conocido dichos resultados y conclusiones). En ambos escenarios, para dotar de validez al análisis Bayesiano, Goligher y Harhay¹⁶ proponen a los investigadores los siguientes consejos: 1) evitar realizar un re-análisis Bayesiano no planificado de los datos; 2) considerar el análisis Bayesiano una parte estándar del plan de análisis preespecificado, incluso si no es el análisis principal; 3) justificar claramente los *priors*, aportando una explicación científica detallada

para el planteamiento de cada uno, en base a datos previos siempre que sea posible y explicando exactamente cómo se derivó y especificó numéricamente el *prior*. El objetivo es asegurar que los *priors* representen el rango real de creencias plausibles en la comunidad experta, especialmente las de los posibles “adversarios” (es decir, escépticos inclinados a dudar de un resultado positivo y entusiastas inclinados a dudar de una interpretación negativa); 4) en general, usar un *prior* neutral para el análisis principal, compatible con la hipótesis de no-efecto (OR 1,0); y 5) evaluar la calidad metodología del estudio utilizando métodos estándar. Por su parte, Heuts y cols.⁴, enfatizan considerar la “diferencia mínima clínicamente importante” (*minimal clinically important difference, MCID*) al momento de plantear *priors*, definiéndose como el umbral de tamaño de efecto considerado como el efecto más pequeño del tratamiento relevante para los pacientes, cuidadores y la sociedad. Este umbral es específico para el paciente, el procedimiento y el resultado, y debería derivarse preferentemente de declaraciones de consenso o cuestionarios, o tener alguna otra base científica. Por último, la interpretación y conclusión de un análisis Bayesiano requiere definir qué valor de probabilidad posterior de beneficio debe ser considerada suficiente para impulsar un cambio en la práctica clínica, sin existir umbrales estándar definidos arbitrariamente, sino más bien depende de factores como el tipo de desenlace, alternativas terapéuticas disponibles, costos, preferencias de los pacientes, etc.².

En relación a los resultados y conclusiones del ensayo controlado aleatorizado de Wacker y cols.⁸, utilizado en nuestro ejemplo docente, tanto en su enfoque frecuentista original como en su reanálisis Bayesiano, nuestro artículo no pretende aconsejar ni desaconsejar el uso de la vitamina C intravenosa en pacientes con *shock* séptico. Su inclusión obedeció a una razón estrictamente docente, sin considerar su validez metodológica. Los últimos años han visto la publicación de numerosos ensayos controlados aleatorizados y revisiones sistemáticas con metanálisis evaluando la eficacia del uso de vitamina C en el manejo de pacientes con *shock* séptico, hecho que refleja el interés y la evidencia inconsistente en torno al tema. Por consiguiente, sugerimos a los lectores examinar el cuerpo de evidencia actualmente disponible, en aras de tomar la mejor decisión clínica.

Al término de este artículo, cabe plantear a la comunidad médica la siguiente pregunta: ¿por qué el análisis Bayesiano de un ensayo controlado aleatorizado no goza de mayor popularidad en nuestro medio? Posiblemente su falta de difusión se deba al predominio de la estadística frecuentista en la formación médica, con el consecuente escaso conocimiento sobre su interpretación y ejecución computacional. Por otra parte, la subjetividad en la selección de los *priors* es un factor que requiere una conceptua-

lización cuidadosa para no restar méritos a este enfoque¹⁵. Finalmente, los autores de este artículo consideramos un error confrontar los enfoques frecuentista y Bayesiano; en lugar de eso, creemos que ambos enfoques son estrategias complementarias que, al combinarse, enriquecen el proce-

so de toma de decisiones. En este sentido, consideramos una buena práctica que el análisis primario de un ensayo controlado aleatorizado incluya ambos enfoques, lo cual brindaría al lector mejores herramientas para fundamentar sus decisiones clínicas.

Referencias bibliográficas

- Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008; 45(3): 135-40. doi: 10.1053/j.seminhematol.2008.04.003.
- Goligher EC, Heath A, Harhay MO. Bayesian statistics for clinical research. *Lancet.* 2024; 404(10457): 1067-76. doi: 10.1016/S0140-6736(24)01295-9.
- Muehleman N, Zhou T, Mukherjee R, Hossain MI, Roychoudhury S, Russek-Cohen E. A tutorial on modern Bayesian methods in clinical trials. *Ther Innov Regul Sci.* 2023; 57(3): 402-16. doi: 10.1007/s43441-023-00515-3
- Heuts S, Kawczynski MJ, Sayed A, Urbut SM, Albuquerque AM, Mandrolia JM, et al. Bayesian analytical methods in cardiovascular clinical trials: why, when, and how. *Can J Cardiol.* 2024; S0828-282X(24)01130-9. doi: 10.1016/j.cjca.2024.11.002.
- Yarnell CJ, Granton JT, Tomlinson G. Bayesian analysis in critical care medicine. *Am J Respir Crit Care Med.* 2020; 201(4): 396-8. doi: 10.1164/rccm.201910-2019ED.
- Tsagkaris C, Papazoglou AS, Moysidis DV, Papadakis M, Trompoukis C. Bayesian versus frequentist clinical research now and then: Lessons from the Greco-Roman medical scholarship. *Ethics Med Public Health.* 2022; 23: 100805. doi: 10.1016/j.jemep.2022.100805
- Cerda J, Cifuentes L. Inferencia Bayesiana en el proceso de diagnóstico clínico: un enfoque docente para la toma de decisiones. *Rev Chilena Infectol* 2024; 41(6): 762-8. doi: 10.4067/s0716-10182024000600168
- Wacker DA, Burton SL, Berger JP, Hegg AJ, Heisdorffer J, Wang Q, et al. Evaluating vitamin C in septic shock: a randomized controlled trial of vitamin C monotherapy. *Crit Care Med.* 2022; 50(5): e458-e467. doi: 10.1097/CCM.00000000000005427.
- Malone H, Coyne I. Complementing the P-value from null-hypothesis significance testing with a Bayes factor from null-hypothesis Bayesian testing. *Nurse Res.* 2020. doi: 10.7748/nr.2020.e1756.
- van Doorn J, van den Bergh D, Böhm U, Dablander F, Derks K, Draws T, et al. The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon Bull Rev.* 2021; 28(3): 813-26. doi: 10.3758/s13423-020-01798-5.
- Chen CY, Chiu CT, Lee HS, Lai CC. The impact of vitamin C-containing treatment on the mortality of patients with sepsis: a systematic review and meta-analysis of randomized controlled trials. *J Infect Public Health.* 2022; 15(12): 1514-20. doi: 10.1016/j.jiph.2022.11.015.
- Lamontagne F, Masse MH, Menard J, Sprague S, Pinto R, Heyland DK, et al; LOVIT Investigators and the Canadian Critical Care Trials Group. Intravenous vitamin C in adults with sepsis in the intensive care unit. *N Engl J Med.* 2022; 386(25): 2387-98. doi: 10.1056/NEJMoa2200644.
- Angriman F, Muttalib F, Lamontagne F, Adhikari NKJ, LOVIT Investigators. IV vitamin C in adults with sepsis: a Bayesian reanalysis of a randomized controlled trial. *Crit Care Med.* 2023; 51(8): e152-e156. doi: 10.1097/CCM.00000000000005871.
- de Grooth HJ, Cremer OL. Bayes and the evidence base: reanalyzing trials using many priors does not contribute to consensus. *Am J Respir Crit Care Med.* 2024; 209(5): 483-4. doi: 10.1164/rccm.202308-1455VP.
- Rendón-Macias M, Riojas-Garza A, Contreras-Estrada D, Martínez-Ezquerro J. Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso. *Rev. alerg. Méx.* [online]. 2018; 65(3): 285-98. https://doi.org/10.29262/ram.v65i3.512
- Goligher EC, Harhay MO. What is the point of Bayesian analysis? *Am J Respir Crit Care Med.* 2024; 209(5): 485-7. doi: 10.1164/rccm.202310-1757VP.