



EDITORIAL



Cuando la respuesta es correcta, pero la pregunta no: el caso de la sífilis congénita

When the answer is correct, but the question is not: the case of congenital syphilis

Hernando Guillermo Gaitán-Duarte^{1,2}

Recibido: 9 diciembre, 2025 Aceptado: 9 diciembre, 2025

En el presente número de la Revista Colombiana de Obstetricia y Ginecología (RCOG) presentamos una revisión sistemática (RS) sobre los efectos, en términos de los resultados perinatales, del tratamiento con penicilina benzatínica (PB) de la sífilis gestacional durante las últimas 12 semanas de embarazo. En esta revisión identificamos que la recomendación actual de considerar posiblemente infectados a los hijos de las mujeres con sífilis gestacional que recibieron tratamiento durante menos de 30 días antes del parto, está soportada en evidencia de muy baja certeza, y que podría haber alguna alternativa para prevenir o curar la sífilis congénita al final del embarazo, que merece ser explorada.

No es frecuente que la RCOG publique estudios realizados por miembros del equipo editorial, sin embargo, tomando en cuenta que la revista es consultada por una numerosa audiencia en Latinoamérica, y que esta revisión identifica un vacío del conocimiento para una condición que afecta de manera importante a nuestra región, consideramos

importante su publicación en la RCOG. Por otra parte, en el proceso de evaluación de la revisión sistemática, los pares hicieron varias observaciones acerca del uso de los estimadores de efecto en estudios de evaluación de intervenciones, que muestran posibles limitaciones en la adecuada interpretación de estos, por lo que consideramos que es un tema que amerita su reflexión en el editorial de su publicación. Los pares invocaron, como debilidad de la revisión, que no se calcularan los estimadores comparativos de riesgo, ya fuera mediante las medidas de efecto relativo como la razón de riesgos (RR), o la diferencia de riesgos (DR). Así, en este editorial dirigido a los tomadores de decisiones y epidemiólogos, presentamos el problema planteado y nuestro punto de vista sobre cuál es la opción más adecuada para evaluar la efectividad de una intervención en escenarios alternativos, donde pueden surgir las diferentes preguntas por contestar.

Comenzaremos por definir una intervención como una acción realizada con el objetivo de mejorar la salud humana mediante la prevención de enfermedades, o cambiar la historia natural de una condición patológica (1). Pueden ser vacunas, tratamientos o dispositivos médicos, intervenciones quirúrgicas o diagnósticas, e inclusive educativas. Es importante reconocer que, además del resultado respecto al beneficio esperado de las intervenciones, su aplicación conlleva el riesgo de resultar en un posible daño, por lo que siempre

* **Correspondencia:** Federación Colombiana de Obstetricia y Ginecología (FECOLSOG), Carrera 15 n.º 98-42 / Oficina 204-205. (Bogotá, Colombia). hggaitand@gmail.com

1. Departamento de Obstetricia y Ginecología, Facultad de Medicina, Universidad Nacional de Colombia. (Bogotá, Colombia).
2. Editor Jefe, Revista Colombiana de Obstetricia y Ginecología, Federación Colombiana de Obstetricia y Ginecología (FECOLSOG). (Bogotá, Colombia).

Cómo citar este artículo: Gaitán-Duarte HG. Cuando la respuesta es correcta, pero la pregunta no: el caso de la sífilis congénita. *Rev Colomb Obstet Ginecol*. 2025;76:4613. <https://doi.org/10.18597/rcog.4613>

se deben evaluar las dos caras de la moneda (2). Los beneficios se evalúan por medio de la eficacia o efectividad (3), así como los riesgos, mediante la presencia de eventos adversos del medicamento (4). Respecto a la eficacia/efectividad existen dos enfoques principales para su evaluación, los cuales dependen de la pregunta de investigación que se quiere contestar y que describimos a continuación.

Primer enfoque. En este escenario, se quiere definir cuál es la mejor alternativa por escoger; esta pregunta implica una comparación, y requiere una hipótesis a priori susceptible de ser rechazada, usualmente “de no diferencia”, entre dos o más tratamientos. Esta es pertinente cuando ya hay un tratamiento disponible de comparación que tiene unos beneficios ya probados, o cuando no se dispone de un tratamiento específico y se diseña una nueva intervención que, se propone, podría ser útil para la prevención o el manejo definido para esa condición. Como ejemplo de la primera situación tenemos la comparación que se hizo entre la heparina no fraccionada –manejo estándar en su momento– con las heparinas de bajo peso molecular en la profilaxis de la enfermedad tromboembólica en pacientes con cáncer (5). Como ejemplo de la segunda opción tenemos la evaluación de una nueva vacuna para prevenir una infección, que se compara con la no aplicación de la vacuna por medio de un placebo, como sucedió con la vacuna del virus del papiloma humano (6). En estas situaciones, el diseño óptimo para su evaluación es el ensayo controlado aleatorizado (ECA) (4), que permite evaluar, de una manera ideal, la asociación causal entre el tratamiento y el resultado. Esta aproximación metodológica permite el control de los factores de confusión conocidos y no conocidos, mediante la asignación aleatoria (7), aunque tiene limitaciones ya que no es inmune a otro tipo de sesgos (8).

Como se quiere rechazar la hipótesis nula, de no diferencia, es importante proteger el estudio del error resultado del azar; hay dos tipos de error aleatorio. El primero, está dado porque el estudio señala diferencias que en realidad no existen (error

tipo I), y el segundo, porque se informa que no hay diferencias cuando realmente sí existen (error tipo II). Ante el error tipo I o α se utiliza como protección el nivel de significancia o probabilidad de que las diferencias observadas sean determinadas por el azar; se ha establecido en una probabilidad con un punto de corte del 5 % ($p < 0,05$) como suficiente para limitar un posible efecto aleatorio y considerar que el efecto es real hasta no tener otra mejor explicación. Para el error tipo II o β , se estableció como punto de corte aceptable una probabilidad de 20 % o 0,2 de que ese resultado sea por azar (también denominado poder del estudio con un valor de $1 - \beta = 0,8$ u 80 %) (9).

En este escenario, para hacer la evaluación de la eficacia / efectividad y seguridad se utilizan los estimadores epidemiológicos que comparan dos grupos, ya sea mediante las medidas de efecto relativo o de asociación, tales como: la razón de incidencias acumuladas o razón de riesgos (RR), o la razón de tasas de incidencia. También se pueden hacer por medio de los estimadores de impacto, entre ellos la diferencia de incidencias acumuladas o de tasas de incidencia que, para nuestro ejemplo, nos daría la reducción absoluta del riesgo (RAR) o, a partir de este, el cálculo del número necesario para tratar (NNT) (10). Se considera que hay asociación cuando los valores del intervalo de confianza del 95 % no cruzan el valor de no efecto: para el RR, el valor de 1 y para la diferencia de riesgos el valor de 0. Si se utilizan los estimadores estadísticos se considera que hay asociación cuando se tiene un valor de $p < 0,05$, y se dice que es estadísticamente significativo. Si se aplica esta aproximación de una manera rígida, se corre el riesgo de que todo se vea como blanco o negro, sin que se miren los tonos de grises existentes entre ellos (11), lo que limita la capacidad de tomar decisiones razonadas en escenarios más extremos como el siguiente.

Segunda opción. Supongamos que enfrentamos una condición que amenaza la vida del paciente o puede dejar importantes secuelas y no tenemos una intervención preventiva ni un tratamiento específico disponible. Como ejemplo recordemos la situación

presentada en marzo de 2020 cuando se declaró la pandemia por el virus Cov 2, en la que los pacientes sufrían frecuentemente compromiso pulmonar severo (20%), quienes tenían una mortalidad cercana al 25% (12). No había ningún tratamiento específico y el único recurso disponible era dar soporte vital con ventilación asistida o dirigido a mantener la función de otros órganos comprometidos por la infección, o por la respuesta inflamatoria sistémica del huésped, hasta su recuperación espontánea o su muerte. En ese momento, la pregunta pertinente no era ¿cuál es el mejor tratamiento?, sino ¿cuál tratamiento puede ser útil para reducir la severidad de la infección y la mortalidad subsecuente? En situaciones como esta no interesan los errores tipo I ni tipo II, es relevante evitar el error tipo III, que se define como “dar la respuesta correcta a la pregunta equivocada”, o decir que un tratamiento es o no mejor cuando lo que interesa es utilizar una alternativa que sea útil en alguna medida para lograr la reducción de la mortalidad, así sea en una pequeña cantidad, en nuestro ejemplo, del 2 o 3%. Para este escenario se sugiere utilizar un ECA pragmático (9) y no evaluar la probabilidad de realmente ser diferentes, sino la de ser la mejor alternativa disponible, al comparar el tratamiento disponible o nuevo con el manejo estándar disponible (de soporte). La efectividad se estima a través de la probabilidad de ser la alternativa más útil entre dos o más opciones (*ranking*) se puede calcular por estadística bayesiana (13) o frecuentista (14). Se utilizan el Factor Bayesiano, la superficie bajo la curva de clasificación acumulada (Sucra - Surface Under the Cumulative Ranking Curve) (15) y su equivalente P-score (14) respectivamente. Si se utiliza la aproximación clásica de no diferencia para una reducción del 2%, se requeriría un tamaño de muestra muy grande (15.500 sujetos) para demostrar el beneficio, lo que sería muy difícil de lograr en muchos escenarios. La alternativa de buscar el tratamiento con la mayor probabilidad de ser el más útil en la crisis, que también requiere un tamaño de muestra mínimo (16), permite hacer una aproximación de más valor para la toma de decisiones en estas situaciones

críticas, y permite ver tonos de grises más allá del blanco y el negro.

Ahora planteemos un tercer escenario, en el cual surge un tratamiento específico de una innovación, pero, no es posible ensamblar un grupo control concurrente. Como ejemplo pensemos en que se requiere la evaluación de un dispositivo inteligente para el tratamiento de la sección medular crónica con compromiso neurológico total desde C5, con cuadriplejía resultante (17). Sería muy difícil lograr que alguien aceptara ser asignado al grupo control conociendo que no va a haber cambio en la historia natural de la condición y seguirá cuadripléjico por siempre, y que, en cambio, el grupo experimental ofrece una oportunidad real de mejora. En estas situaciones, la única opción viable es una cohorte de expuestos a una intervención sin grupo control, que también se podría llamar serie de expuestos o, mal denominada, serie de casos (18). En esta situación, la efectividad de la intervención se basa en la estimación del riesgo absoluto de presentar el evento (mejorar o curar), por medio de la incidencia acumulada. Se utiliza cuando la pregunta de investigación se orienta a determinar la probabilidad de lograr un desenlace beneficioso (p. ej., prevención o curación) o de reducir el riesgo de un desenlace adverso (como fallo en la prevención o persistencia de la enfermedad) dentro de una cohorte de individuos expuestos a un tratamiento, y no es factible tener un grupo control, o cuando el efecto del beneficio de la intervención es muy grande (19).

Ahora aplicaremos estos conceptos en la evaluación del tratamiento de la sífilis gestacional en cuanto al resultado de la prevención de la sífilis congénita. Es importante conocer un poco la historia de cómo se definió el mejor tratamiento para esta condición. Desde 1998, el tratamiento recomendado por los Centers for Disease Control and Prevention (CDC), de los Estados Unidos, para la sífilis es la PB, una dosis para sífilis temprana y tres dosis si es tardía o de duración desconocida (20). Posteriormente, en un estudio de cohorte, sin grupo control, de mujeres gestantes con sífilis (SG), quienes fueron tratadas

bajo este esquema, hubo cura clínica o serológica en el 99 % de ellas, además, se observó una incidencia de sífilis congénita (SC) del 2 % (6/340) en los neonatos de estas mujeres (21). De los seis neonatos con SC, en cuatro de ellos la madre recibió tratamiento en los últimos 15 días antes el parto. Estas observaciones posiblemente originaron las recomendaciones específicas para el tratamiento de la SG y SC publicadas en la guía del CDC en 2006 (22). Para la SG se recomienda desde ese entonces el mismo tratamiento utilizado para la sífilis en personas no gestantes y se hace una segunda recomendación, según la cual, los neonatos nacidos de madres cuyo tratamiento para la SG fue dado menos de 30 días antes del parto, deben considerarse infectados o con SC ; lo anterior lo asimila al estado de un neonato cuya madre no fue tratada y recomienda el tratamiento de estos bebés con penicilina cristalina 100.000 a 150.000 U /kg de peso/día por 10 días, recomendación que está vigente hasta hoy, y ha sido adoptada por otras guías nacionales (23, 24). Vemos cómo a partir de un estudio de cohorte de expuestos a la PB que presenta una frecuencia absoluta con un efecto importante, se tomaron decisiones de salud pública.

En la revisión publicada en este número de nuestra revista, los autores no identificaron ECA, posiblemente por preponderancia de los antiguos conceptos éticos donde se consideraba inadecuado incluir gestantes en estos estudios (25). Se identificaron diez estudios de cohorte que evaluaron los efectos del tratamiento en las últimas 12 semanas de gestación, y de estos dos estudios evaluaron el efecto cuando se aplicaron en las últimas 4 semanas. Para el desenlace de SC se identificaron 9 estudios, cinco fueron cohortes sin grupo control, tres cohortes con grupo control de no tratamiento y otro estudio con un grupo de comparación que mezcló pacientes no tratadas con pacientes tratadas luego de la semana 37. Por lo tanto, los autores de la RS se ubican en el último escenario antes mencionado, esto es, no disponer ECA, no tener un grupo de comparación o tener un grupo control de no tratamiento tomado de registros del sistema de notificación en tres estudios de cohortes, y no haber un tratamiento alternativo disponible para la

SG diagnosticada menos de 30 días antes del parto.

Ahora apliquemos el primer enfoque de establecer riesgos comparados en los tres estudios de cohorte con grupo control disponibles, que no recibieron tratamiento. Recordemos que esta es la aproximación adecuada cuando la hipótesis a priori asume que no existe diferencia entre dos opciones, en este caso comparar tratamiento versus no tratamiento en las últimas 12 y 4 semanas de gestación. Bajo este enfoque, obtendríamos los siguientes resultados, para la estimación de la efectividad en las últimas 12 semanas:

Hong (2017): 1.704 mujeres con SG (RR = 0,16; IC 95 %: 0,09-0,30). Interpretación: es mejor tratar que no tratar.

Wan (2016): 2.598 mujeres con SG (RR = 1,25; IC 95 % 0,77-2,03). Interpretación: no hay diferencia entre tratar y no tratar; importante imprecisión.

Zhang (2016): 2.472 mujeres gestantes (RR = 0,69; IC 95 %: 0,39-1,12). Interpretación: no hay diferencia entre tratar y no tratar; importante imprecisión.

Bajo la aproximación clásica, un estudio indica que hay una diferencia estadísticamente significativa y dos estudios indican que no hay diferencias estadísticamente significativas (resultados heterogéneos e imprecisos) entre los dos grupos; los tres estudios con alto riesgo de sesgos, lo que resulta en evidencia con muy baja certeza o alta incertidumbre. Esta aproximación podría llevar a no tomar una decisión sobre el tratamiento a la madre en las últimas 12 semanas (considerar a todos estos recién nacidos como infectados y darles tratamiento intravenoso), contradiciendo lo identificado en el estudio de Alexander antes mencionado, que mostró un riesgo absoluto de SC con tratamiento del 2 %, aun en mujeres tratadas en un periodo mayor o igual a 30 días antes del parto. Esto representa el error tipo III: una posible respuesta correcta (hay incertidumbre) a la pregunta equivocada (cuál es la mejor alternativa a escoger). Por lo tanto, ese enfoque no es pertinente para el análisis de los resultados de esta RS.

La pregunta de investigación apropiada sería: ¿cuál es el riesgo de las últimas doce semanas, específicamente, de las últimas cuatro semanas del embarazo, de tener neonatos con SC en mujeres que reciben el tratamiento estándar con penicilina G benzatínica o regímenes alternativos?

Tenemos que el cuerpo de la evidencia identificado resulta en una frecuencia agrupada absoluta del 8% (Rango 2,1 al 27,8%) para el tratamiento con PB las últimas 12 semanas (7 estudios) y para las últimas 4 semanas un rango de efectos del 6,5 al 23,4% (2 estudios), ambos inferiores a los malos resultados perinatales (49%) informados en el curso natural de los resultados neonatales en mujeres con SG no tratadas (26). Estos resultados muestran un efecto que podría ser importante en la prevención de la SC tanto en las 12 como en las últimas 4 semanas de gestación, que debería ser considerado. Como consecuencia de lo anterior, es más adecuado emplear la estimación de la incidencia acumulada —o riesgo absoluto— de sífilis congénita como medida principal de desenlace de efectividad. En este contexto, el uso de la razón de riesgo o de la diferencia absoluta de riesgo puede no ser informativo.

Como conclusión, resaltamos la importancia de definir la pregunta adecuada en circunstancias donde no hay una intervención disponible que tenga algún efecto sobre la historia natural de la enfermedad y surge una alternativa; de esta manera, el método analítico siempre debe ser el apropiado según el contexto dentro del cual surge la pregunta de investigación, y no debe ser inflexible por defecto. Además, para la toma de decisiones en salud se debe reivindicar el rol de la cohorte de expuestos a una intervención sin grupo control, que informa sobre el riesgo absoluto en la evaluación de efectividad cuando no es posible ensamblar un grupo control.

REFERENCIAS

- Smith P, Morrow R, Ross D. Types of intervention and their development. En *Field Trials of Health Interventions*, 3 ed. Oxford: Oxford Medicine Online; 2015. <https://doi.org/10.1093/med/9780198732860.003.0002>
- Chou R, Aronson N, Atkins D, et al. Assessing harms when comparing medical interventions. En: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* [Internet]. Rockville: Agency for Healthcare Research and Quality; 2008. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK47098/>
- Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA*. 1994;271(20):1615-9. <https://doi.org/10.1001/jama.271.20.1615>.
- Fletcher R, Fletcher S, Wagner E. Treatment. En: *Clinical Epidemiology. The Essentials*, 3 ed. Baltimore: Williams & Wilkins; 1996.
- Matar C, Kahale L, Hakoum M, Tsolakian I, Etxeandia-Ikobaltzeta I, Yosuico V, et al. Anticoagulation for perioperative thromboprophylaxis in people with cancer. *Cochrane Database Syst Rev*. 2018;7(7):CD009447. <https://doi.org/10.1002/14651858.CD009447>
- Muñoz N, Kjaer S, Sigurdsson K, Iversen O, Hernandez-Avila M, Wheeler C, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. *J Natl Cancer Inst*. 2010;102(5):325-39. <https://doi.org/10.1093/jnci/djp534>
- Kleijnen J, Gøtzsche P, Kunz R, Oxman A, Chalmers I. So what's so special about randomisation? En: Maynard A, Chalmers I, editors. *Non-random reflections on health services research: On the 25th anniversary of Archie Cochrane's Effectiveness and Efficiency*. BMJ Books; 1997. p. 93-106.
- Higgins J, Savović J, Page M, Elbers R, Sterne J. Assessing risk of bias in a randomized. En: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors [Internet]. *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*. Cochrane; 2024. Disponible en: cochrane.org/handbook.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20(8):637-48. [https://doi.org/10.1016/0021-9681\(67\)90041-0](https://doi.org/10.1016/0021-9681(67)90041-0).
- Gordis L. Ensayos clínicos con distribución aleatoria. Algunos aspectos adicionales. En Gordis L. *Epidemiología*, 3 ed. Madrid: Elsevier; 2005.
- Schmidt M, Parner E. The conclusion generator. *Ann Epidemiol*. 2024;97:33-7. <https://doi.org/10.1016/j.annepidem.2024.06.008>.
- Richardson S, Hirsch J, Narasimhan M, Crawford J, McGinn T, Davidson K, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA*. 2020;323(20):2052-9. <https://doi.org/10.1001/jama.2020.6775>

13. Rendón-Macías M, Riojas-Garza A, Contreras-Estrada D, Martínez-Ezquerro J. Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso. *Rev Alerg Mex.* 2018;65(3):285-98. <https://doi.org/10.29262/ram.v65i3.512>.
14. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol.* 2015;15:58. <https://doi.org/10.1186/s12874-015-0060-8>
15. Chaimani A, Higgins J, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS ONE.* 2013;8(10):76654. <https://doi.org/10.1371/journal.pone.0076654>.
16. US Food & Drug Administration. Division of Biostatistics Office of Surveillance and Biometrics. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials 2010 [Internet]. White Oak: FDA; 2010. Disponible en: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials>
17. Lorach H, Galvez A, Spagnolo V, Martel F, Karakas S, et al. Walking naturally after spinal cord injury using a brain-spine interface. *Nature.* 2023;618(7963):126-33. <https://doi.org/10.1038/s41586-023-06094-5>.
18. Dekkers O, Egger M, Altman D, Vandembroucke J. Distinguishing case series from cohort studies. *Ann Intern Med.* 2012;156(1 Pt 1):37-40. <https://doi.org/10.7326/0003-4819-156-1-201201030-00006>.
19. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ.* 2007;334(7589):349-51. <https://doi.org/10.1136/bmj.39070.527986.68>.
20. Workowski K. The 1998 CDC Sexually Transmitted Diseases Treatment Guidelines. *Curr Infect Dis Rep.* 2000;2(1):44-50. <https://doi.org/10.1007/s11908-000-0086-9>. PMID:
21. Alexander J, Sheffield J, Sanchez P, Mayfield J, Wendel G Jr. Efficacy of treatment for syphilis in pregnancy. *Obstet Gynecol.* 1999;93(1):5-8. [https://doi.org/10.1016/s0029-7844\(98\)00338-x](https://doi.org/10.1016/s0029-7844(98)00338-x).
22. Workowski K, Berman S. Sexually transmitted diseases treatment guidelines, 2006. *MMWR Recomm Rep.* 2006;55(RR-11):1-94.
23. Janier M, Unemo M, Dupin N, Tiplica G, Potočnik M, Patel R. 2020 European guideline on the management of syphilis. *J Europ Acad Dermatol Venereol.* 2021;35(3):574-88. <https://doi.org/10.1111/jdv.16946>
24. Ministerio de Salud y Protección Social. Fondo de Población de las Naciones Unidas (UNFPA). Guía de práctica clínica basada en la evidencia para la atención integral de la sífilis gestacional y congénita [Internet]. GPC; 2014. Disponible en: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IETS/gpc-guia-completa-sifilis.pdf>
25. Hunt A, Banner N, Littler K. The global forum on bioethics in research meeting, “ethics of research in pregnancy”: Emerging consensus themes and outputs. *Reprod Health.* 2017;14(Suppl 3):158. <https://doi.org/10.1186/s12978-017-0431-1>.
26. Watson-Jones D, Changalucha J, Gumodoka B, Weiss H, Rusizoka M, Ndeki L, et al. Syphilis in pregnancy in Tanzania. I. Impact of maternal syphilis on outcome of pregnancy. *J Infect Dis.* 2002;186(7):940-7. <https://doi.org/10.1086/342952>.