

[Home](#)[Contenido](#)jameslindlibrary.org

Pruebas auténticas de los tratamientos en la atención médica

Durante el último medio siglo, la atención médica ha tenido un impacto substancial sobre las oportunidades que la gente tiene para vivir más tiempo sin padecer graves problemas de salud. Se ha calculado que la atención médica ha sido responsable de un tercio a la mitad del aumento producido en la expectativa de vida y de cinco años más de vida libres de problemas crónicos de salud (Bunker et al. 1994). Aún así, el público podría haber obtenido —y aún puede obtener— un valor muy superior por los recursos substanciales invertidos en la investigación orientada a mejorar la salud. Agreguemos, también, que podrían haberse prevenido algunos de los desastres causados por los tratamientos en el pasado y que otros podrían prevenirse en el futuro.

The James Lind Library ha sido creada para mejorar el entendimiento general de las pruebas auténticas de los tratamientos que atañen a la atención médica y de cómo estos tratamientos han evolucionado con el tiempo.

Son frecuentes las reivindicaciones engañosas respecto de los efectos de los tratamientos, por lo tanto, todos deberíamos entender de qué manera se generan afirmaciones válidas respecto de esos efectos. Sin este conocimiento, corremos el riesgo de llegar a la conclusión de que los tratamientos inútiles son provechosos o que los tratamientos útiles son infructuosos.

Las pruebas auténticas de los tratamientos son pruebas que, al reducir las influencias engañosas de los [prejuicios](#) y de [la obra de la casualidad](#), toman las medidas necesarias para obtener información confiable sobre los efectos del tratamiento. Cuando esta necesidad de pruebas auténticas de los tratamientos se deja de lado, las personas sufren y mueren innecesariamente.

Los ensayos explicativos con que cuenta *The James Lind Library* se han escrito para que se comprenda mejor el por qué de la necesidad de pruebas auténticas de los tratamientos y en qué consisten dichas pruebas. Usted puede acceder a cada ensayo haciendo clic en las palabras subrayadas, abajo, o puede seleccionarlas de la pantalla de [Contenido](#). Si desea bajar todos los ensayos, de manera que pueda imprimirlos para leerlos fuera de la pantalla, [haga clic aquí](#).

[Es necesario realizar pruebas auténticas](#) porque existen muchos ejemplos de personas que han resultado dañadas, sin intención, cuando las decisiones que atañen al tratamiento no tomaron en cuenta evidencia confiable.

Los [principios que gobiernan las pruebas auténticas](#) han ido evolucionando durante al menos un milenio, y continúan haciéndolo.

Para trabajar sobre una [genuina incertidumbre](#) sobre los efectos de un tratamiento es esencial hacer [comparaciones](#). Las comparaciones auténticas de los tratamientos deben evitar los prejuicios o [sesgos](#), ya sea que provengan de [diferencias entre las personas que se comparan](#) o de las [diferencias en la forma en que se evalúan los resultados del tratamiento](#). La [identificación confiable de efectos no previstos de los tratamientos](#) presenta particulares desafíos.

[Interpretar las comparaciones no prejuiciosas](#) no siempre es sencillo. En ocasiones los efectos del tratamiento se pasan por alto porque existen [diferencias entre el tratamiento que se pretende administrar y el tratamiento recibido](#). [La obra de la casualidad](#) también puede engañarnos.

Las pruebas auténticas de los tratamientos deben tener en cuenta toda la evidencia pertinente. Preparar revisiones sistemáticas de toda la evidencia pertinente significa minimizar el impacto de [los informes prejuiciosos](#) y minimizar también [los sesgos en la selección de la evidencia disponible](#). Un proceso estadístico denominado [metanálisis](#) puede ayudarnos a impedir que [la obra de la casualidad](#) nos engañe al hacer las revisiones sistemáticas .

Para realizar pruebas auténticas de los tratamientos en la atención médica se necesitan [revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable](#). No obstante, aún con revisiones sistemáticas actualizadas, es importante estar atento a las tendencias y a los “giros” que pueden hacer que distintas revisiones del mismo tema lleguen a conclusiones conflictivas.

En resumen, *The James Lind Library* contiene los siguientes ensayos:

Por qué se necesitan pruebas auténticas

Por qué son esenciales las comparaciones

Por qué las comparaciones deben abordar una incertidumbre genuina

Cómo evitar los sesgos en las comparaciones

Diferencias entre las personas comparadas

Diferencias en la manera en que se evalúan los resultados de los tratamientos

Interpretación de comparaciones no prejuiciosas

Diferencias entre el tratamiento que se pretende administrar y el que realmente se recibe

Cómo considerar la obra de la casualidad

Identificación de los efectos no previstos de los tratamientos

Revisiones sistemáticas de toda la evidencia pertinente

Cómo manejar los sesgos en los informes de la evidencia disponible

Cómo evitar los sesgos en la selección de la evidencia disponible

Cómo reducir la obra de la casualidad a través del metanálisis

Revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable

Estos ensayos explicativos se basan en un valioso material ilustrativo incluido en The James Lind Library, al que se puede acceder haciendo clic en los enlaces subrayados o en las imágenes de los ensayos.

El texto incluido en estos ensayos puede copiarse y utilizarse con fines no comerciales, con la condición de que se haga un reconocimiento explícito a "**The James Lind Library (www.jameslindlibrary.org)**".

El material que contienen estos ensayos también se ha incorporado en *Testing Treatments: better research for better health care* (Prueba de los tratamientos: mejor investigación para una salud mejor), de Evans, Thornton y Chalmers, un libro de 100 páginas que la British Library publicó en 2006. ISBN 0-7123-4909-X.

TESTING TREATMENTS BETTER RESEARCH FOR BETTER HEALTHCARE

How do we know whether a particular drug, therapy or operation really works, and how well? How reliable is the clinical evidence? Are clinical trials truly unbiased? And is current research fully focussed on the real needs of patients? Such timely and pressing questions are raised and resolved in this probing inquiry into modern clinical research, with far-reaching implications for daily medical practice and patient care. What emerges is the surprising truth that clinical research is neither as unbiased, nor as relevant as patients have every right to expect, but that everyone – patients, doctors and researchers – can do much to change current practice and achieve better healthcare.

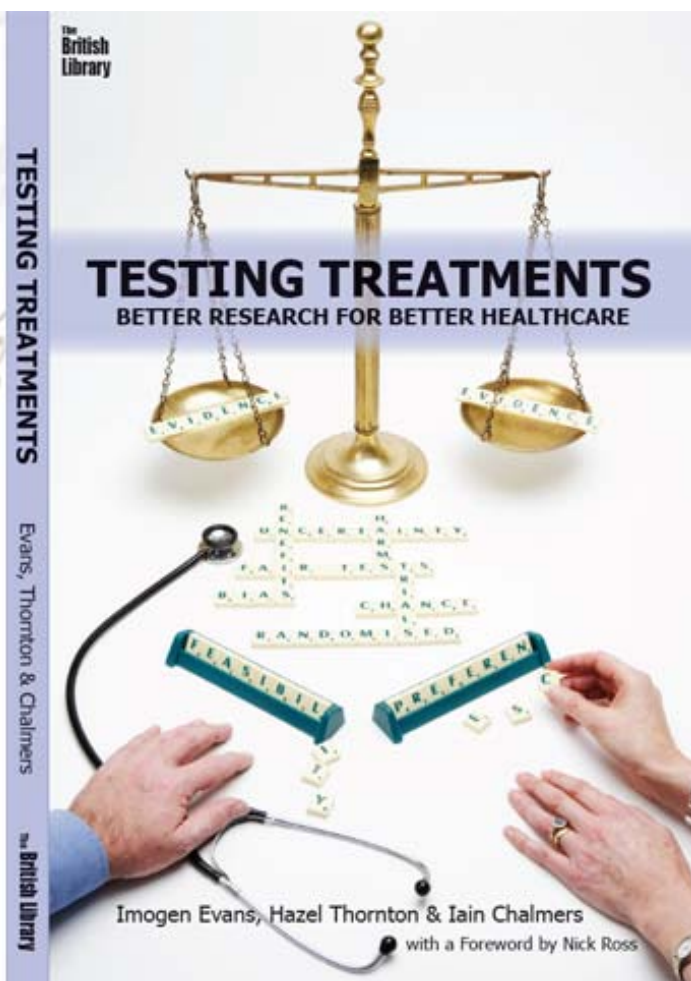
Aimed at both patients and professionals, *Testing Treatments* builds a lively and thought-provoking argument for better, more reliable, more relevant research, with unbiased or 'fair' trials, and explains how patients can work with doctors to achieve this vital goal.

Expertly and thoroughly researched, but never dry or dull, the fast-moving commentary, spanning the gamut of illness and therapy – from mastectomy to thalidomide – explores a vast range of revealing case-studies, enlivened throughout by entertaining anecdotes and vivid eyewitness accounts drawn from the direct experience of patients, practitioners and researchers.

EVIDENCE

The
British Library

Cover image & design by
I. RO CHALMERS
www.thefolio.com



Cite as: Editorial commentary (2007). Pruebas auténticas de los tratamientos en la atención médica. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Por qué se necesitan pruebas auténticas](#)

Select essay:

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

Por qué se necesitan pruebas auténticas

Tratar de ayudar más que dañar

¿Por qué necesitamos pruebas auténticas de los tratamientos en la atención médica? ¿No han hecho los médicos, durante siglos, "lo mejor que pudieron" por sus pacientes? Lamentablemente, hay muchos ejemplos de médicos y de otros profesionales de la salud que han hecho mal a sus pacientes a partir de decisiones respecto del tratamiento que no se basaban en información proveniente de lo que ahora consideramos evidencia confiable sobre los efectos de los tratamientos. Si hacemos una visión retrospectiva, profesionales de la salud de la mayoría —si no de todas— las esferas han dañado a sus pacientes sin intención de hacerlo, a veces en una escala demasiado grande ([si desea ejemplos, haga clic aquí](#)). De hecho, en ocasiones los pacientes mismos han dañado a otros pacientes cada vez que, en función de teorías no probadas y experiencias personales limitadas, los incentivaron a utilizar tratamientos que terminaron siendo dañinos. El tema no es si debemos o no culpar a estas personas, sino si es posible reducir los efectos dañinos de tratamientos no probados adecuadamente. Y, en gran medida, es posible.

Reconocer que los tratamientos pueden, a veces, hacer más daño que bien es un prerrequisito para reducir el daño no intencional (Gregory 1772; [Haygarth 1800](#); Fordyce 1802; [Behring 1893](#)). Entonces, necesitamos estar listos para admitir que existen dudas sobre los efectos de los tratamientos y para promover pruebas de esos tratamientos de manera tal que se reduzca la incertidumbre. Esas son pruebas auténticas.

Por qué las teorías sobre los efectos de los tratamientos deben probarse en la práctica

A menudo las personas han resultado dañadas porque los tratamientos sólo se basaban en teorías sobre cómo debería tratarse la enfermedad, sin una prueba de cómo esas teorías funcionaban en la práctica. Por ejemplo, durante siglos las personas creyeron que las enfermedades eran causadas por "desequilibrios humorales" y, en consecuencia, se purgaba a los pacientes, se les extraía sangre, se los hacía vomitar y tomar rapé, con la convicción de que esto terminaría con los supuestos desequilibrios. Sin embargo, en el siglo XVII, un solitario médico flamenco fue suficientemente impertinente como para desafiar a las autoridades médicas de la época a que evaluaran la validez de sus teorías proponiéndoles una prueba auténtica de los resultados de esos tratamientos tan desagradables ([Van Helmont 1662](#)).



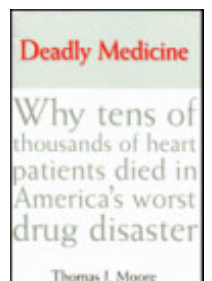
Para comienzos del siglo XIX, los cirujanos militares británicos habían comenzado a mostrar los efectos perjudiciales de la sangría para el tratamiento de las "fiebres" (Robertson 1804; [Hamilton 1816](#)). Unas pocas décadas más tarde, un médico parisino volvía a desafiar estas prácticas ([Louis 1835](#)). Sin embargo, a comienzos del siglo XX, aún se acusaba de negligentes a algunos médicos ortodoxos de Boston, Estados Unidos, que no utilizaban la sangría para tratar la neumonía (Silverman 1980). De hecho, Sir William Osler, una de las autoridades médicas mundiales más influyentes y generalmente cautelosa en la recomendación de tratamientos no probados, decía a sus lectores: "Por cierto, durante las últimas décadas hemos extraído demasiado poca sangre. La neumonía es una de las enfermedades en las que una venesección [sangría] puede salvar vidas. Para que sea útil, debe hacerse a tiempo. En un hombre sano y vigoroso con fiebre alta y pulso saltón, la extracción de veinte a treinta onzas de sangre es en todo sentido beneficiosa." (Osler 1892)



Si bien la necesidad de comprobar la validez de las teorías en la práctica fue reconocida por algunos hace al menos mil años ([Ibn Hindu siglos X-XI](#)), este importante principio es aún ignorado con demasiada frecuencia. Benjamin Spock, el influyente pediatra estadounidense, por ejemplo, se basó en teorías no probadas para informar a los lectores de su libro mejor vendido *Baby and Child Care* (Cuidado de los bebés y de los niños) que la desventaja de que los bebés durmieran boca arriba era que, si vomitaban, corrían mayor riesgo de ahogarse. En consecuencia, el Dr. Spock aconsejó a los millones de lectores que incentivaran a sus bebés a dormir boca abajo (Spock, 1966). Sabemos que esta recomendación, aparentemente racional en teoría, provocó que decenas de miles de bebés murieran de muerte súbita (Gilbert et al. 2004).



La utilización de medicamentos para prevenir anomalías en el ritmo cardíaco en personas con infartos es otro ejemplo del riesgo que representa llevar a la práctica una teoría no probada. Dado que las anomalías en el ritmo cardíaco están asociadas con un mayor riesgo de muerte temprana después de un infarto, la teoría sostenía que estos medicamentos reducirían la cantidad de esas muertes tempranas. No obstante, que una teoría parezca razonable no necesariamente significa que sea acertada. Años después de que los medicamentos obtuvieran la licencia y se adoptaran en la práctica, se descubrió que, de hecho, aumentaban el riesgo de muerte súbita después del infarto. Se ha calculado que, en el momento culminante de su utilización a fines de la



década de 1980, pueden haber matado hasta 70.000 personas por año solamente en los Estados Unidos (Moore 1995), muchas más personas que la cantidad de estadounidenses muertos en la guerra de Vietnam.



Por el otro lado, la confianza erróneamente depositada en el pensamiento teórico como guía de la práctica también provocó que algunos tratamientos se rechazaran de manera inapropiada porque los investigadores no creyeron que pudieran funcionar. Las teorías basadas en los resultados de la investigación realizada en animales, por ejemplo, a veces predicen correctamente los resultados de las pruebas del tratamiento en humanos. Pero éste no siempre es el caso. Basándose en los resultados de experimentos realizados en ratas, algunos investigadores se convencieron de que no tenía sentido administrar medicamentos para disolver los coágulos (trombolíticos) a pacientes que habían sufrido infartos más de seis horas antes. Si estos pacientes no hubieran participado en algunas de las pruebas auténticas de estos medicamentos, no sabríamos ahora que pueden beneficiarse con dicho tratamiento (Fibrinolytic Therapy Trialists' Collaborative Group 1994).

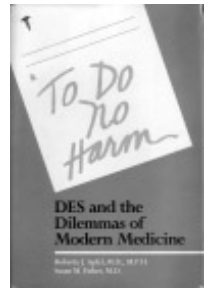


Las observaciones en la práctica clínica o en la investigación en laboratorio y en animales sugieren que determinados tratamientos beneficiarán o no beneficiarán a los pacientes, pero estos y muchos otros ejemplos dejan en claro que es fundamental utilizar pruebas auténticas para averiguar si, en la práctica, estos tratamientos ayudan más que lo que dañan o viceversa.

Por qué las pruebas de los tratamientos médicos deben ser pruebas auténticas

No probar en la práctica las teorías sobre los tratamientos no es la única causa prevenible de las tragedias ocasionadas por los tratamientos. Estas tragedias también se han producido debido a que las pruebas utilizadas para evaluar los efectos de los tratamientos no fueron confiables sino engañosas. Las pruebas auténticas requieren tomar medidas para reducir la probabilidad de que los efectos de los [prejuicios](#) y de [la casualidad nos engañen](#).

En la década de 1950, por ejemplo, la teoría y pruebas mal controladas que generaron evidencia no confiable sugirieron que, si se administraba una hormona sexual sintética —el dietilstilbestrol (DES)— a mujeres embarazadas con antecedentes de abortos espontáneos y nacimientos muertos, aumentaría la probabilidad de buenos resultados en embarazos posteriores. Si bien pruebas auténticas habían sugerido que el DES no daba resultado, la teoría y la evidencia no confiable — además de una comercialización agresiva— lograron que el DES se recetara a millones de mujeres embarazadas durante las décadas siguientes. Las consecuencias fueron desastrosas: algunas de las hijas de mujeres que habían recibido DES desarrollaron cánceres de vagina y otros niños padecieron otros problemas de salud, incluidas malformaciones de sus órganos reproductores e infertilidad (Apfel y Fisher 1984).



Los problemas generados por pruebas inadecuadas de los tratamientos continúan presentándose. Una vez más, como resultado de evidencia no confiable y de una comercialización agresiva, se convenció a millones de mujeres de que utilizaran la terapia de reemplazo hormonal (TRH) no sólo porque podía reducir los molestos síntomas de la menopausia sino también porque se afirmaba que reduciría las posibilidades de que sufrieran infartos y accidentes cerebrovasculares. Cuando estas reivindicaciones sobre el tratamiento se evaluaron en pruebas auténticas, los resultados demostraron que, lejos de reducir los riesgos de infarto y accidente cerebrovascular, la TRH no sólo aumenta los riesgos de estas afecciones a veces fatales, sino que también tiene otros efectos secundarios no deseados (McPherson 2004).



Estos ejemplos de la necesidad de realizar pruebas auténticas de los tratamientos son sólo unos pocos de los muchos que muestran la manera en que los tratamientos pueden hacer más daño que bien. Es necesario que aumentemos nuestro conocimiento sobre las pruebas auténticas de los tratamientos para que —incluida una dosis saludable de escepticismo— podamos evaluar las reivindicaciones sobre los efectos de los tratamientos con mayor sentido crítico. De esa forma, todos estaremos en mejores condiciones de juzgar qué tratamientos pueden generar más beneficios que perjuicios.

Los principios que gobiernan las pruebas auténticas de los tratamientos han ido evolucionando durante siglos, y continúan haciéndolo.

Cite as: Editorial commentary (2007). Por qué se necesitan pruebas auténticas. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Por qué son esenciales las comparaciones](#)

Select essay:

Referencias

- Apfel RJ, Fisher SM (1984). *To do no harm: DES and the dilemmas of modern medicine*. New Haven, Ct: Yale University Press.
- Behring, Boer, Kossel H (1893). Zur Behandlung diphtheriekranker Menschen mit Diphtherieheilserum. *Deutsche Medicinische Wochenschrift* 17: 389-393.
- Fibrinolytic Therapy Trialists' Collaborative Group (1994). Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. *Lancet* 1994;343:311-322.
- Fordyce G (1802). *A second dissertation on fever*. London: J Johnson
- Gilbert R, Salanti G, Harden M, See S (2005). Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology* 34:874-87.
- Gregory J (1772). *Lectures on the duties and qualifications of a physician*. London: Strahan and Cadell.
- Hamilton AL (1816). *Dissertatio Medica Inauguralis De Synocho Castrensi* (Inaugural medical dissertation on camp fever). Edinburgh: J Ballantyne.
- Haygarth J (1800). *Of the imagination, as a cause and as a cure of disorders of the body: exemplified by fictitious tractors, and epidemical convulsions*. Bath: R. Crutwell.
- Ibn Hindu (10th-11th century CE; 4th-5th century AH). *Miftah al-tibb wa-minhaj al-tullab* [The key to the science of medicine and the students' guide].
- Louis PCA (1835). *Recherches sur les effets de la saignée dans quelques maladies inflammatoires et sur l'action de l'émétique et des vésicatoires dans la pneumonie*. Paris: Librairie de l'Académie royale de médecine.
- McPherson K (2004). Where are we now with hormone replacement therapy? *BMJ* 328: 357-358.
- Moore TJ (1995). *Deadly Medicine*. New York: Simon and Schuster.
- Osler W (1892). *Principles and Practice of Medicine*. London: Appleton, p 530.
- Robertson R (1804). *Observations on the diseases incident to seamen*, 2nd edn. Vol. 1, London: for the author.
- Silverman W (1980). In: Chalmers I, McIlwaine G (eds). *Perinatal Audit and Surveillance*. London: Royal College of Obstetricians and Gynaecologists, 1980: 110.
- Spock B (1966). *Baby and Child Care*. 165th printing. New York: Pocket Books, pp 163-164.
- van Helmont JB (1662). *Oriatrike, or physick refined: the common errors therein refuted and the whole are reformed and rectified* [translated by J Chandler]. Lodowick-Loyd: London, p 526.

[Home](#)

[Contenido](#)

[Comments welcome](#)

Home

Contenido

jameslindlibrary.org

Por qué son esenciales las comparaciones

¿Es un tratamiento mejor que la naturaleza y el tiempo?

Tanto los pacientes como los profesionales de la atención médica esperan que los tratamientos resulten útiles. Estas expectativas optimistas pueden tener un efecto muy positivo sobre la satisfacción de todos respecto de la atención médica. Como observó el médico británico Richard Asher en uno de sus ensayos para médicos:

“Si cree fervientemente en su tratamiento, aún cuando las pruebas controladas indiquen que no será muy beneficioso, alcanzará resultados mucho mejores, sus pacientes se sentirán mucho mejor y también mejorarán sus ingresos. Creo que esto explica el notable éxito de algunos de los miembros menos dotados pero más crédulos de nuestra profesión, y también el duro desprecio por las estadísticas y las pruebas controladas que suelen mostrar los médicos exitosos que se ponen de moda.” (Asher 1972)

Con frecuencia, las personas se recuperan de sus enfermedades sin ningún tratamiento específico: la naturaleza y el tiempo son grandes remedios. Como sugirió Oliver Wendell Holmes en el siglo XIX, cuando se disponía de muy pocos tratamientos útiles ([Holmes 1861](#)), “Creo firmemente que si toda la *materia medica*, como se usa ahora, se hundiera en el fondo del mar, eso sería lo mejor que podría pasarle a la raza humana... y lo peor para los peces.”

Es obvio que, al probar los tratamientos, deben tenerse en cuenta el progreso y el desenlace que tendría la enfermedad si no se la tratara: el tratamiento puede mejorar o empeorar los resultados. A través de los siglos, los escritores han llamado la atención sobre la necesidad de mantenerse escéptico frente a la afirmación de que los efectos de los tratamientos pueden superar los efectos de la naturaleza. En otras palabras: “Si deja que la influenza se cure naturalmente, probablemente se recupere en una semana, pero si va al médico, se recuperará en sólo siete días”.

El efecto placebo

Sabiendo que muchas enfermedades son autolimitantes, a veces los médicos recetan tratamientos inertes con la esperanza de que sus pacientes generen un beneficio psicológico: el así llamado efecto placebo. Los pacientes que están convencidos de que un tratamiento dado les ayudará a aliviar sus síntomas —aunque, de hecho, ese tratamiento no tenga efectos físicos— seguramente se sentirán mejor.

Durante siglos, los médicos han reconocido la importancia de utilizar placebos ([enumerar registros pertinentes](#)). William Cullen, por ejemplo, ya se refirió a este uso del placebo en 1772 ([Cullen 1772](#)) y las referencias a los placebos aumentaron durante el siglo XIX ([Cummings 1805](#); [Ministro de Asuntos Internos 1832](#); [Forbes 1846](#)). Dado que creía que el tratamiento ortodoxo con medicamentos estaba usurpando el crédito debido a la “naturaleza”, Austin Flint administró a trece pacientes con reumatismo un “remedio placebo” que consistía en un extracto altamente diluido de corteza del árbol *quassia*. El resultado fue que “la evolución favorable de los casos fue tal que, en general, aseguró la confianza absoluta de los pacientes” ([Flint 1863](#)). En el Guy’s Hospital de Londres, William Withey Gull llegó a conclusiones similares después de tratar a 21 pacientes con fiebre reumática “principalmente con agua de menta” ([Sutton 1865](#)). A comienzos del siglo XX, William Rivers analizó en detalle los efectos de los tratamientos que son mediados psicológicamente ([Rivers 1908](#)).



La necesidad de comparar

De la misma manera en que durante siglos se reconocieron el poder de curación de la naturaleza y el efecto placebo, así también se reconoció la necesidad de comparar para evaluar los efectos de los tratamientos más allá de los efectos mediados natural y psicológicamente. En ocasiones, las comparaciones entre tratamientos tienen lugar en la mente de las personas, cuya impresión es que ellos mismos u otras personas están respondiendo de manera diferente a un nuevo tratamiento en comparación con respuestas anteriores a los tratamientos. Ambroise Paré, un cirujano militar francés, por ejemplo, llegó a la conclusión de que el tratamiento de las heridas de combate con aceite hirviendo (como era la práctica común) probablemente fuera perjudicial. Esta deducción tuvo lugar cuando se agotó la reserva de aceite y sus pacientes se recuperaron más rápido que lo acostumbrado ([Paré 1575](#)).



La mayor parte del tiempo, impresiones como ésta necesitan de un seguimiento con investigaciones formales, quizá, inicialmente, del análisis de los registros de atención médica. Entonces, dichas impresiones pueden dar lugar a comparaciones realizadas minuciosamente. El peligro surge cuando se utilizan sólo las impresiones como guía

para recomendaciones y decisiones relacionadas con el tratamiento.

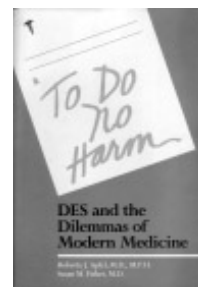
Efectos drásticos y efectos moderados de los tratamientos

Las comparaciones de tratamientos basadas en las impresiones, o en análisis relativamente limitados, sólo proporcionan información confiable en las extrañas ocasiones en que los efectos del tratamiento son drásticos ([si desea ver los registros pertinentes, haga clic aquí](#)). Algunos ejemplos son el uso del opio para aliviar el dolor ([Tibi 2005](#)), la higiene para prevenir el tétanos (trismo) en los recién nacidos (Schleisner 1849), el cloroformo para anestesiar, la insulina para la diabetes ([Banting et al. 1922](#)), la dieta de hígado para la anemia perniciosa ([Minot y Murphy 1926](#)), las drogas sulfa para las infecciones después de dar a luz ([Colebrook y Purdie 1937](#)), la estreptomicina para la meningitis tuberculosa ([MRC 1948](#)), la adrenalina para reacciones alérgicas que ponen en peligro la vida (McLean-Tookey et al. 2003) y los fármacos diseñados genéticamente para algunas formas raras de leucemia (Druker et al. 2001). La mayoría de los tratamientos médicos, sin embargo, no producen esos efectos drásticos y, a menos que se tenga el cuidado de evitar el sesgo en las comparaciones, puede llegarse a conclusiones erróneas y por ende muy peligrosas con respecto a los efectos de un tratamiento.

Comparación de los tratamientos que hoy se administran con los tratamientos del pasado

Fue en parte debido a la confianza en comparaciones prejuiciosas con experiencias anteriores, que los médicos y las mujeres creyeron que el dietilstilbestrol (DES) reduciría el riesgo de abortos y nacimientos muertos. Jamás hubo ninguna evidencia obtenida en pruebas auténticas (no prejuiciosas) de que el DES pudiera generar esos resultados y más tarde se demostró que había provocado cáncer a las hijas de algunas de las mujeres embarazadas a las que había sido recetado. No debería promocionarse ningún tratamiento para el que no se haya demostrado de manera confiable que es beneficioso.

La comparación entre los tratamientos que se administran hoy en día con los que se administraban en el pasado rara vez ofrece una base segura para una prueba auténtica ([Behring et al. 1893](#); [Roux et al. 1894](#)), dado que los factores relevantes, aparte del tratamiento mismo, varían con el tiempo. Los abortos y los nacimientos muertos, por ejemplo, son más comunes en los primeros embarazos que en los siguientes. Comparar, entonces, la frecuencia de abortos y nacimientos muertos acaecidos en embarazos posteriores para los que se había recetado DES con el resultado de los primeros embarazos en los que no se había utilizado ese fármaco será, probablemente, una base que generará serios equívocos al evaluar los efectos del tratamiento. Por lo tanto, siempre que sea posible, las comparaciones deberán hacerse sobre la administración de diferentes tratamientos en aproximadamente el mismo momento.



Comparación de tratamientos en pruebas cruzadas realizadas en los pacientes en forma individual

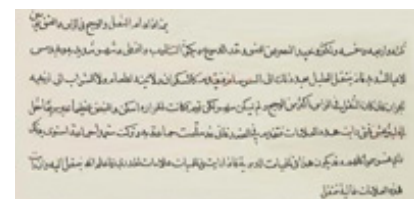
A veces, administrar diferentes tratamientos en aproximadamente el mismo momento puede significar administrar a un mismo paciente diferentes tratamientos, uno después del otro: lo que denominamos una prueba cruzada ([Martini 1932](#); [si desea ver los registros pertinentes, haga clic aquí](#)). En 1786, en Bath, Inglaterra, el Dr. Caleb Parry informó uno de los primeros ejemplos de prueba cruzada. El Dr. Parry deseaba averiguar si existía alguna razón que justificara pagar por el oneroso ruibarbo turco importado para utilizar como purgante en el tratamiento de sus pacientes en vez de utilizar el ruibarbo que se sembraba en Inglaterra. Así que "cruzó" el tipo de ruibarbo que daba a cada paciente individual en diferentes momentos y luego comparó los síntomas que cada paciente tuvo al ingerir cada tipo de ruibarbo ([Parry 1786](#)). (¡Y no encontró ninguna ventaja en el ruibarbo turco!)



Las comparaciones de tratamientos en cada paciente en forma individual se realizan cuando la afección regresa después de suspender el tratamiento. Hay muchas circunstancias en las que esto no puede llevarse a cabo. Por ejemplo, es normalmente imposible comparar de esta manera diferentes operaciones quirúrgicas o tratamientos administrados para enfermedades progresivas.

Comparación de grupos de pacientes que reciben diferentes tratamientos al mismo tiempo

En general, los tratamientos se prueban comparando grupos de personas que reciben distintos tratamientos. Una comparación de dos tratamientos no sería auténtica si uno de los tratamientos se administrara a personas relativamente sanas y el otro, a personas relativamente enfermas; por lo tanto, deben compararse las experiencias de grupos similares de personas que reciben diferentes tratamientos durante el mismo período. Al-Razi reconoció este concepto hace más de mil años cuando, en su deseo de alcanzar una conclusión sobre cómo tratar a pacientes con signos de meningitis en los primeros estadios, trató a un grupo de pacientes e, intencionalmente, no administró el tratamiento a un grupo de comparación ([al-Razi siglo IX](#)).



Para obtener pruebas auténticas de los tratamientos se necesita hacer comparaciones con la naturaleza o con otros tratamientos. Si estas comparaciones han de ser auténticas, [deben abordar una incertidumbre genuina](#), evitar [los sesgos](#) y [la casualidad](#), e [interpretarse con cuidado](#).

Cite as: Editorial commentary (2007). Por qué son esenciales las comparaciones. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Por qué las comparaciones deben abordar una incertidumbre genuina](#)

Select essay:

Referencias

al-Razi (10th century CE; 4th Century AH). Kitab al-Hawi fi al-tibb [The comprehensive book of medicine].

Asher R (1972). Talking sense. London: Pitman Medical.

Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA (1922). Pancreatic extracts in the treatment of diabetes mellitus. Canadian Medical Association Journal 12: 141-146.

Behring, Boer, Kossel H (1893). Zur Behandlung diphtheriekranker Menschen mit Diphtherieheilserum. Deutsche Medicinische Wochenschrift 17: 389-393.

Colebrook L, Purdie AW (1937). Treatment of 106 cases of puerperal fever by sulphanilamide. Lancet 2: 1237-1242 & 1291-1294.

Cullen W (1772). Clinical lectures. Edinburgh, February-April, 218-9.

Cummngs R (1805). Medical and Physical Journal, page 6.

Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, Sawyers CL (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. New England Journal of Medicine 344: 1031-1037.

Flint A (1863). A contribution toward the natural history of articular rheumatism; consisting of a report of thirteen cases treated solely with palliative measures. American Journal of the Medical Sciences 46: 17-36.

Forbes J (1846). Homeopathy, allopathy and 'young physic.' British and Foreign Medical Review 21: 225-265.

Holmes OW (1861). Currents and countercurrents in medical science. In: Works, 1861 Vol ix, p 185.

Martini P (1932). Methodenlehre der Therapeutischen Untersuchung. Berlin: Springer.

McLean-Tooke APC, Bethune CA, Fay AC, Spickett GP (2003). Adrenaline in the treatment of anaphylaxis: what is the evidence? BMJ 327: 1332-1335.

Medical Research Council (1948). Streptomycin treatment of tuberculous meningitis. Lancet 1: 582-596.

Ministry of Internal Affairs (1823). [Conclusion of the Medical Council regarding homeopathic treatment]. Zhurnal Ministerstva Vnutrennih del, 3: 49-63.

Minot GR, Murphy WP (1926). Treatment of pernicious anaemia by a special diet. JAMA 87: 470-476.

Paré A (1575). Les oeuvres de M. Ambroise Paré conseiller, et premier chirugien du Roy avec les figures & portraits tant de l'Anatomie que des instruments de Chirurgie, & de plusieurs Monstres. Paris: Gabriel Buon.

Parry CH (1786). Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs, No. I and No. II made on patients of the Pauper Charity. Letters and Papers of the Bath Society III: 431-453.

Rivers WHR (1908). The influence of alcohol and other drugs on fatigue. London: Edward Arnold.

Roux E, Martin L, Chaillou A (1894). Trois cent cas de diphthérie traité par le serum antidiphthérique. Annales de l'Institut Pasteur 8: 640-661.

Schleisner PA (1849). *Island fra et lægevidenskabeligt Synspunkt*. København: Boghandler Iversen.

Sutton HG (1865). Cases of rheumatic fever, treated for the most part by mint water. Collected from the clinical books of Dr Gull, with some remarks on the natural history of that disease. *Guy's Hospital Report* 11:392-428.

Tibi S (2005). *The medicinal use of opium in ninth-century Baghdad*. Leiden: Brill.

[Home](#)

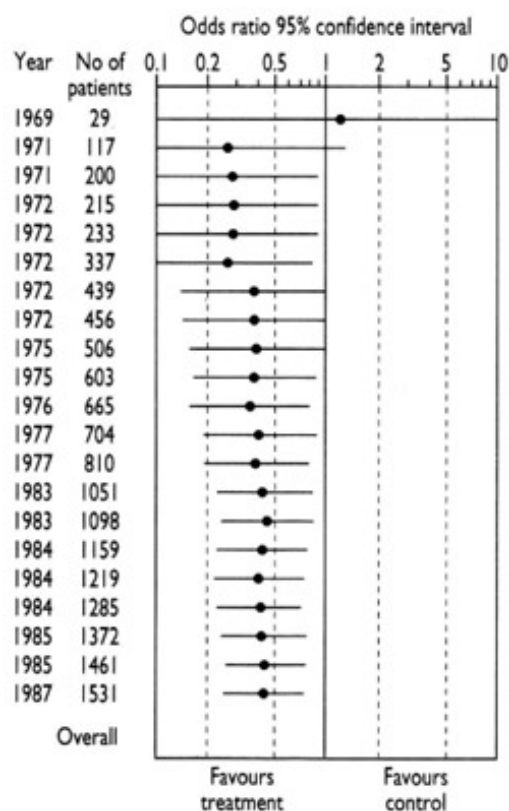
[Contenido](#)

[Comments welcome](#)

Por qué las comparaciones deben abordar una incertidumbre genuina

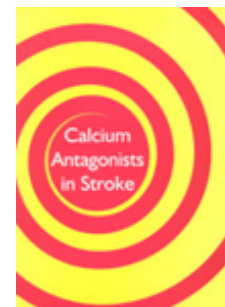
Se sigue llevando a cabo una buena cantidad de investigación aún cuando no exista una incertidumbre genuina. Los investigadores que no realizan revisiones sistemáticas de las pruebas pasadas de los tratamientos antes de embarcarse en nuevos estudios a veces no reconocen (o prefieren ignorar) que las dudas sobre los efectos del tratamiento ya han sido tratadas de manera convincente. Esto significa que, a veces, se les niega a las personas que participan en la investigación un tratamiento que podría ayudarlas o se les da un tratamiento que probablemente las perjudique.

En el diagrama que acompaña a este párrafo y al siguiente se muestra la evidencia acumulada correspondiente a pruebas auténticas realizadas para evaluar si los antibióticos (en comparación con los placebos inactivos) reducían el riesgo de muerte posoperatoria en personas con cirugía de intestino (Lau et al. 1995). La primera prueba auténtica se informó en 1969. Los resultados de ese pequeño estudio dejaron dudas sobre si los antibióticos eran o no útiles: la línea horizontal que representa los resultados cruza la línea vertical que separa los efectos favorables de los antibióticos de los no favorables. Con razón, esta incertidumbre volvió a ser tema de estudio en otras pruebas a principios de la década de 1970.



Sin embargo, a medida que se fue acumulando evidencia, a mediados de la década de 1970 quedó claro que los antibióticos reducen el riesgo de muerte después de la cirugía (la línea horizontal cae francamente del lado de la línea vertical que favorece al tratamiento). Aún así, los investigadores continuaron con los estudios hasta fines de la década de 1980. Por lo tanto, a la mitad de los pacientes, que recibieron placebos en estos últimos estudios, se les negó una forma de atención médica para la que ya se había demostrado que reducía el riesgo de muerte después de sus operaciones. ¿Cómo pudo suceder esto? Probablemente porque los investigadores continuaron embarcándose en nuevas investigaciones sin revisar sistemáticamente la evidencia que ya existía. Este comportamiento continúa siendo demasiado común en la comunidad investigadora, en parte debido a que algunos de los incentivos en el mundo de la investigación —comerciales y académicos— no anteponen los intereses de los pacientes (Chalmers 2000).

Los pacientes y los participantes en la investigación también pueden sufrir debido a que los investigadores no han revisado sistemáticamente la evidencia pertinente obtenida de la investigación en animales antes de comenzar a probar los tratamientos en humanos. Un equipo holandés revisó la experiencia de más de 7000 pacientes que habían participado en pruebas de un nuevo bloqueante de calcio administrado a personas con accidente cerebrovascular y no encontraron ninguna evidencia que respaldara su uso, cada vez mayor, en la práctica (Horn y Limburg 2001). A raíz de esto, se preguntaron acerca de la calidad y los hallazgos de la investigación en animales sobre la que se había basado la investigación en los pacientes. La revisión de los estudios realizados en animales reveló que sus resultados jamás habían sugerido que el fármaco pudiera ser útil en los humanos (Horn et al. 2001).



La razón principal por la que la investigación no aborda incertidumbres genuinas es que los investigadores sencillamente no tienen la disciplina suficiente para revisar la evidencia pertinente de manera sistemática antes de iniciar nuevos estudios. Sin embargo, a veces las razones son más siniestras. Es posible que los investigadores sí

conozcan la evidencia existente pero quieran diseñar estudios para asegurarse de que su propia investigación generará resultados favorables para determinados tratamientos. En general, aunque no siempre, los impulsan razones económicas (Djulgovic et al. 2000; Sackett y Oxman 2003). Estos estudios están deliberadamente diseñados para constituirse en pruebas no auténticas de los tratamientos: esto puede lograrse reteniendo un tratamiento de comparación que, se sabe, ayuda a los pacientes (como en el ejemplo anterior) o administrando tratamientos de comparación en dosis inadecuadamente bajas (de manera que no funcionen tan bien) o en dosis inadecuadamente elevadas (de manera que causen efectos secundarios no deseados) ([vea el comentario de Mann y Djulgovic](#)), y también puede ser el resultado de un seguimiento demasiado breve de los pacientes (que deja de lado los efectos retardados de los tratamientos) y de utilizar medidas de resultados ("sustituciones") que tengan muy poca o ninguna correlación con los resultados que importan a los pacientes.

Quizá sorprenda a los lectores de este ensayo que los comités de ética establecidos durante las últimas décadas para garantizar que la investigación sea ética hayan hecho tan poco para modificar esta mala práctica en la investigación. La mayoría de estos comités han defraudado a las personas que debían haber protegido porque no les han exigido a los investigadores y los patrocinadores que pidieron la autorización para nuevas pruebas que revisaran sistemáticamente la evidencia ya existente (Savulescu et al. 1996; Chalmers 2002). El fracaso de los comités de ética en la investigación en su tarea de proteger eficientemente a los pacientes y al público en su totalidad pone de manifiesto la importancia de aumentar el conocimiento general sobre las características que hacen que una prueba de un tratamiento médico sea auténtica.

Cite as: Editorial commentary (2007). Por qué las comparaciones deben abordar una incertidumbre genuina. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Cómo evitar los sesgos en las comparaciones](#)

Select essay:

Referencias

Chalmers I. Current Controlled Trials: an opportunity to help improve the quality of clinical research. *Current Controlled Trials in Cardiovascular Medicine* 2000; 1: 3-8.
Available: <http://cvm.controlled-trials.com/content/1/1/3>

Chalmers I (2002). Lessons for research ethics committees. *Lancet* 359: 174.

Djulgovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH (2000). The uncertainty principle and industry-sponsored research. *Lancet* 356: 635-638.

Horn J, Limburg M (2001). Calcium antagonists for acute ischemic stroke (Cochrane Review). In: *The Cochrane Library*, Issue 3, Oxford: Update Software.

Horn J, de Haan RJ, Vermeulen M, Luiten PGM, Limburg M (2001). Nimodipine in animal model experiments of focal cerebral ischaemia: a systematic review. *Stroke* 32: 2433-38.

Lau J, Schmid CH, Chalmers TC (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary clinical practice. *Journal of Clinical Epidemiology* 48: 45-57.

Mann H, Djulgovic B. Why comparisons must address genuine uncertainties. James Lind Library (www.jameslindlibrary.org).

Sackett DL, Oxman AD (2003). HARLOT plc: an amalgamation of the world's two oldest professions. *BMJ* 2003; 327: 1442-1445.

Savulescu J, Chalmers I, Blunt J (1996). Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 313: 1390-1393.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)[jameslindlibrary.org](http://www.jameslindlibrary.org)

Cómo evitar los sesgos en las comparaciones

En ocasiones, los tratamientos producen efectos drásticos ([si desea ver los registros pertinentes, haga clic aquí](#)). Es probable que estos efectos sean específicos y no intencionales, como, por ejemplo, cuando una persona tiene una reacción alérgica a un antibiótico. Los tratamientos también pueden tener efectos sorprendentemente beneficiosos, como sucede con la adrenalina en los casos de reacciones alérgicas que ponen la vida en peligro. 2003). Sin embargo, esos efectos impactantes son raros. Normalmente, los efectos de los tratamientos son más modestos, aunque bien vale la pena conocerlos, como sucede con el uso de la aspirina para reducir el riesgo de infarto ([Elwood 2004](#)).

La aspirina, por ejemplo, no previene todas las muertes prematuras después de un ataque cardíaco pero sí reduce la probabilidad de muerte en alrededor de un veinte por ciento, lo cual es importante en una afección tan común. Si debemos detectar de manera confiable estos efectos moderados al tiempo que beneficiosos de la mayoría de los tratamientos, debemos estar atentos a que comparaciones prejuiciosas no nos lleven a creer que los tratamientos son útiles cuando en realidad son inútiles o perjudiciales, o que son inútiles cuando realmente pueden ser beneficiosos.

Los sesgos en las pruebas de los tratamientos son aquellas influencias y factores que pueden conducirnos a conclusiones sobre los efectos de los tratamientos que son sistemáticamente diferentes de la verdad. Si bien existen muchos tipos de sesgos que pueden distorsionar los resultados de la investigación en salud (Sackett 1979), en The James Lind Library nos hemos concentrado en aquéllos que deben minimizarse en las pruebas auténticas de los tratamientos. Ellos son:

- [sesgos relacionados con las diferencias en las personas comparadas](#);
- [sesgos debidos a diferencias en la manera en que se evalúan los resultados de los tratamientos](#);
- [sesgos en los informes de la evidencia disponible](#); y
- [sesgos en la selección de la evidencia disponible](#).

Ignorar esos prejuicios (o, a veces, aprovecharlos sin escrúpulos) puede llevar a la gente a pensar que un tratamiento nuevo es mejor que uno que ya existe cuando no lo es. Esto podría ser el resultado de fundamentar las conclusiones en:

- estudios que comparan el progreso de personas relativamente sanas que reciben un nuevo tratamiento con el progreso de personas relativamente enfermas que reciben un tratamiento estándar ([sesgo en la asignación](#));
- estudios en los que las evaluaciones de los resultados del tratamiento probablemente hayan sido "afectadas" en favor de un tratamiento nuevo, por ejemplo, comparando las opiniones de personas que saben que han utilizado un tratamiento nuevo y caro con las opiniones de aquéllos que quizá se sientan desilusionados porque continúan utilizando un tratamiento estándar nada movilizador ([sesgo en el observador o en la medición](#));
- solamente estudios que muestran un nuevo tratamiento desde un aspecto favorable sin tener en cuenta aquéllos que sugieren que podría ser dañino, estudios que, en general, no se informan ([sesgo en el informe](#));
- la selección e interpretación prejuiciosas de la evidencia disponible para respaldar un punto de vista en particular ([sesgo en la revisión](#)).

En general, no se reconoce a las pruebas no auténticas que resultan de estos prejuicios por lo que son. Sin embargo, las personas con derechos adquiridos a veces explotan estos sesgos de manera tal que los tratamientos se presentan como si fueran mejores de lo que realmente son (Sackett and Oxman 2003).

Ya sea que esos prejuicios sean involuntarios o deliberados, las consecuencias son las mismas: a menos que las pruebas de los tratamientos sean auténticas, algunos tratamientos inútiles o dañinos parecerán beneficiosos mientras que algunos tratamientos beneficiosos parecerán inútiles o dañinos.

Cite as: Editorial commentary (2007). Cómo evitar los sesgos en las comparaciones. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Diferencias entre las personas comparadas](#)

Select essay:

Referencias

Elwood P (2004). The first randomised trial of aspirin for heart attack and the advent of systematic overviews of trials. The James Lind Library (www.jameslindlibrary.org).

McLean-Tooke APC, Bethune CA, Fay AC, Spickett GP (2003). Adrenaline in the treatment of anaphylaxis: what is the evidence? *BMJ* 327: 1332-1335.

Sackett DL (1979). Bias in analytic research. *Journal of Chronic Diseases* 32: 51-63.

Sackett DL, Oxman AD (2003). HARLOT plc: an amalgamation of the world's two oldest professions. *BMJ* 2003; 327: 1442-1445.

[Home](#)

[Contenido](#)

[Comments welcome](#)

Home

Contenido

jameslindlibrary.org

Cómo evitar los sesgos en las comparaciones:

Diferencias entre las personas comparadas

Comparación de diferentes tratamientos administrados a grupos de personas

Normalmente, comparar tratamientos implica comparar las experiencias de grupos de personas que han recibido diferentes tratamientos. Si estas comparaciones han de ser auténticas, la composición de los grupos debe ser similar, de manera que se comparen grupos iguales. Si aquellos que reciben un tratamiento tienen, de alguna manera, mayor probabilidad de tener éxito (o de fracasar) que aquellos que reciben un tratamiento alternativo, este sesgo en la asignación hará imposible pensar que los criterios de valoración reflejarán efectos diferenciales de los tratamientos más de lo que lo harían los efectos de la naturaleza o el paso del tiempo.

En el siglo XVIII, el cirujano William Cheselden ya era consciente del problema que representaban los "grupos disímiles" cuando los cirujanos comparaban sus respectivas tasas de mortalidad después de las operaciones para extraer cálculos de la vejiga. Cheselden señaló que era importante tener en cuenta la edad de las personas que trataban los distintos cirujanos. Destacó el hecho de que las tasas de mortalidad variaban con la edad de los pacientes ([Cheselden 1740](#)): los pacientes mayores tenían más probabilidades de morir que los pacientes más jóvenes. Esto significaba que, si deseaban comparar la frecuencia con que se producían muertes en grupos de pacientes que habían sido sometidos a distintos tipos de operaciones, debían tener en cuenta las diferencias en la edad de los pacientes de los grupos de comparación.



La comparación de las experiencias y los resultados de pacientes que recibieron diferentes tratamientos en el pasado se utiliza aún hoy como una forma de intentar evaluar los efectos de los tratamientos. El desafío es saber si los grupos de comparación eran suficientemente parecidos antes de recibir el tratamiento. Los intentos por evaluar los efectos de la terapia de reemplazo hormonal (TRH) comparando las experiencias relacionadas con la enfermedad de aquellas mujeres que habían recibido TRH con las de otras mujeres que no la habían recibido, es un ejemplo. Como lo demostró un análisis posterior de pruebas auténticas de la TRH, tratar de evaluar de esta manera y en forma retrospectiva los efectos de los tratamientos a veces puede resultar peligrosamente engañoso (McPherson 2004).

Rara vez es posible tener la certeza absoluta de que los grupos de comparación elegidos de personas que recibieron un tratamiento en el pasado sean comparables en todos los sentidos que nos importan con personas que han recibido un tratamiento alternativo más recientemente. Y éste es el caso, aún cuando se disponga de algunos datos (como la edad o sus antecedentes médicos) sobre los pacientes que han recibido diferentes tratamientos. Es posible que, sencillamente, no se disponga de otros datos que pueden ser de gran importancia (como la probabilidad de una recuperación espontánea).

Planificar las comparaciones entre tratamientos antes de comenzar con el tratamiento es un método más acertado. James Lind, por ejemplo, antes de comenzar la comparación de seis tratamientos para el escorbuto a bordo del buque *Salisbury* en 1747, se tomó el trabajo de seleccionar pacientes que se encontraran en una etapa similar de esta enfermedad, a menudo fatal. También se aseguró de que recibieran la misma dieta de base y los acomodó en condiciones similares. Además del tratamiento, estos eran factores que podrían haber influido en sus probabilidades de recuperación ([Lind 1753](#)). Es necesario realizar esfuerzos comparables para intentar asegurarnos de que los grupos de comparación de tratamientos están compuestos por personas similares.



Cómo evitar el sesgo en la formación de los grupos de comparación de tratamientos utilizando la alternación o la aleatorización

Si bien Lind hizo lo posible por asegurarse de que los marineros que formaban sus seis grupos de comparación tuvieran las mismas características, no describió cómo decidió qué marineros recibirían cuál de los seis tratamientos. Existe sólo una manera de asegurarse de que los grupos de comparación de tratamientos estén formados de manera tal que sean similares en todos los sentidos que nos interesan, conocidos y desconocidos: utilizar alguna forma de proceso aleatorio para armar los grupos de comparación y, así, evitar la selección prejuiciosa para los distintos tratamientos antes de iniciar el tratamiento.

Cien años después de Lind, un médico naval, Graham Balfour, mostró cómo podía lograrse esto en una prueba cuyo objetivo era ver si la *belladonna* prevenía la fiebre escarlata en los niños. En el orfanato militar del que era responsable, utilizó la alternación —“para evitar la



imputación de la selección”— para decidir qué niños recibirían y cuáles no recibirían *belladonna* (Balfour 1854). La alternación es uno de los diversos métodos no prejuiciosos para asignar grupos similares para la comparación de tratamientos antes de administrarlos. La primera mitad del siglo XX provee muchos ejemplos de grupos para la comparación de tratamientos que se formaron utilizando la alternación o rotación (como Hamilton 1816; MRC 1944), o por sorteo (Colebrook 1929): utilizando dados (Doull et al. 1931), cuentas de colores (Theobald 1937) o números de muestra tomados al azar (Bell 1941; MRC 1948; MRC 1950; MRC 1951). Esta “asignación al azar” es la única y esencial característica de la categoría de pruebas auténticas a las que nos referimos como “aleatorizadas”. Que una asignación sea aleatoria (a diferencia de la casual) significa que se conocen las probabilidades de que algo ocurra pero que los resultados no pueden preverse en ninguna ocasión en particular. De modo que, por ejemplo, si se utiliza una moneda para escoger al azar, la probabilidad de que salga cara es del 50%, aunque es imposible saber si, al tirar la moneda, ésta será cara o seca.



Como se ilustra en el ensayo al que podrá acceder si hace [clic aquí](#), tirar los dados o hacer un sorteo constituye una manera antigua y respetable de tomar decisiones justas. Estos métodos nos ayudan a asegurarnos de que los grupos de comparación no estén compuestos por distintos tipos de personas. Los factores importantes conocidos y medidos, como la edad, pueden verificarse. Sin embargo, cabe esperar que los factores no medidos que pueden influir en la recuperación de una enfermedad, como la dieta, la ocupación y la ansiedad, se compensen en el promedio. Si desea ver cómo la asignación aleatoria genera grupos similares de personas, [haga clic aquí para ver una demostración](#).

A medida que se fue generalizando la experiencia en el uso de la asignación mediante la alternación y el azar para formar de manera imparcial grupos de pacientes para comparar tratamientos diferentes, quedó claro que, para evitar los sesgos en la creación de los grupos para la comparación de los tratamientos, es esencial adherirse estrictamente a los esquemas de asignación (MRC 1934). Puede eliminarse por completo el riesgo de una asignación prejuiciosa si se ocultan los esquemas de asignación de tratamientos a quienes toman decisiones respecto de la participación en las comparaciones. En resumen, para evitar que hagan trampa y, así, tiñan las comparaciones de parcialidad (MRC 1944; MRC 1948; MRC 1950; MRC 1951).



Cómo evitar pérdidas prejuiciosas en los grupos de comparación de tratamientos

Después de tomarse el trabajo de asegurarse de que los grupos de comparación de tratamientos se formaron de manera que se comparen iguales, es importante evitar que los sesgos se filtren a través de un abandono selectivo de los grupos de comparación por parte de los pacientes. En la medida de lo posible, deberá mantenerse la similitud entre los grupos asegurando que se haga un seguimiento de todas las personas asignadas a los grupos de comparación de los tratamientos y se las incluya en el análisis principal de los resultados de las pruebas, un análisis que llamamos “de intención de tratar” (Bell 1941).

Si esto no se hace, las pruebas de los tratamientos pueden no resultar auténticas. Tomemos, por ejemplo, dos maneras diferentes de tratar a personas con períodos breves de desequilibrios mentales o físicos debido a que los vasos sanguíneos que irrigan su cerebro están parcialmente bloqueados. Es importante tratar esta afección dado que las personas que experimentan estos períodos de desequilibrios por esta causa corren un riesgo mayor de sufrir un accidente cerebrovascular, lo que podría generarles incapacidad o, aún, matarlos. Uno de los tratamientos para los períodos de desequilibrio es tomar aspirinas para evitar que el bloqueo empeore, el otro, una intervención quirúrgica para intentar eliminar el bloqueo en el vaso sanguíneo.

Una comparación auténtica de estos dos métodos para tratar dichos desequilibrios implicaría crear dos grupos de personas utilizando un método de asignación imparcial (como la aleatorización). Así, la comparación se iniciaría cotejando dos grupos de pacientes con características similares y pasaría a comparar las frecuencias respectivas de accidentes cerebrovasculares posteriores. Pero si la frecuencia de accidente cerebrovascular en el grupo tratado con cirugía sólo se registrara entre los pacientes que sobrevivieron a los efectos inmediatos de la operación, el factor importante de que la operación misma puede causar accidente cerebrovascular y muerte sería dejado de lado. La comparación de los dos tratamientos sería, por lo tanto, no auténtica: sus resultados presentarían un cuadro optimista —parcial y engañoso— de los efectos de la cirugía. No se habrían comparado igual con igual.

La comparación principal en los estudios aleatorizados debe basarse, en la medida de lo posible, en todas las personas asignadas para recibir cada uno de los tratamientos comparados, sin excepción, y en los grupos a los que se los asignó originalmente. Si no se cumple con este principio, la información que se reciba sobre los efectos globales de los tratamientos será prejuiciosa.

Cite as: Editorial commentary (2007). Diferencias entre las personas comparadas. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [control of allocation bias](#) and [crossover test](#)

Next essay: [Diferencias en la manera en que se evalúan los resultados de los tratamientos](#)

Select essay:

Referencias

Balfour TG (1854). Quoted in West C. Lectures on the Diseases of Infancy and Childhood. London, Longman, Brown, Green and Longmans, p 600.

Bell JA (1941). Pertussis prophylaxis with two doses of alum-precipitated vaccine. Public Health Reports 56:1535-1546.

Cheselden W (1740). The anatomy of the human body. 5th edition. London: William Bowyer.

Colebrook D (1929). Irradiation and health. Medical Research Council Special Report Series No.131.

Doull JA, Hardy M, Clark JH, Herman NB (1931). The effect of irradiation with ultra-violet light on the frequency of attacks of upper respiratory disease (common colds). American Journal of Hygiene 13:460-77.

Hamilton AL (1816). Dissertatio Medica Inauguralis De Synocho Castrensi (Inaugural medical dissertation on camp fever). Edinburgh: J Ballantyne.

Lind J (1753). A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson.

McPherson K (2004). Where are we now with hormone replacement therapy? BMJ 328:357-358.

Medical Research Council Therapeutic Trials Committee (1934). The serum treatment of lobar pneumonia. BMJ 1:241-245.

Medical Research Council (1944). Clinical trial of patulin in the common cold. Lancet 2:373-5.

Medical Research Council (1948). Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. BMJ 2:769-782.

Medical Research Council (1950). Clinical trials of antihistaminic drugs in the prevention and treatment of the common cold. BMJ 2:425-431.

Medical Research Council (1951). The prevention of whooping-cough by vaccination. BMJ 1:1463-1471

Parry CH (1786). Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs, No. I and No. II made on patients of the Pauper Charity. Letters and Papers of the Bath Society III:407-422.

Silverman WA, Chalmers I. Casting and drawing lots. The James Lind Library (www.jameslindlibrary.org).

Theobald GW (1937). Effect of calcium and vitamin A and D on incidence of pregnancy toxæmia. Lancet 2:1397-1399.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

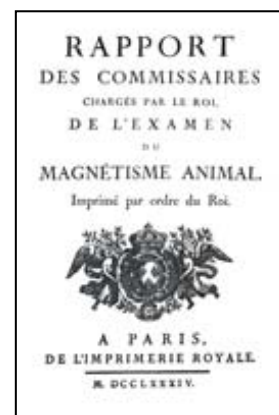
Cómo evitar los sesgos en las comparaciones:

Diferencias en la manera en que se evalúan los resultados de los tratamientos

Utilización del método ciego para reducir los sesgos al evaluar los resultados de los tratamientos

En el caso de algunos criterios de valoración que se utilizan para evaluar un tratamiento —la sobrevida, por ejemplo— es muy poco probable que pueda hacerse una evaluación prejuiciosa dado que no hay mucho espacio para las opiniones. Éste fue el caso en algunas de las pruebas de procedimientos quirúrgicos en el siglo XVIII, en las que la sobrevida era la medida principal del éxito o el fracaso del tratamiento ([Faure 1759](#)). No obstante, la evaluación de la mayoría de los otros resultados o bien siempre da lugar a la subjetividad (como sucede con los síntomas de los pacientes) o bien puede estar relacionada con ella. Los prejuicios que llevan a estas percepciones equívocas se denominan sesgos del observador. Estos sesgos generan un problema particular cuando las personas creen que ya “conocen” el efecto de un tratamiento o cuando tienen determinadas razones para preferir uno de los tratamientos que se están comparando. Cuando no se toman medidas para reducir los sesgos en las evaluaciones de los resultados al comparar tratamientos, se tiende a sobreestimar los efectos del tratamiento (Schulz et al. 1995). Cuanto mayor es el elemento subjetividad al evaluar los resultados, mayor será la necesidad de reducir estos sesgos del observador para garantizar que las pruebas de los tratamientos sean auténticas.

En estas circunstancias para nada extraordinarias, “cegar” a los pacientes y a los médicos es un elemento importante para que las pruebas sean auténticas. La primera evaluación ciega (encubierta) de un tratamiento parece haberla llevado a cabo una comisión de inquisición designada por Luis XVI en 1784 para que investigara las afirmaciones de Anton Mesmer respecto de los efectos del “magnetismo animal” ([Comisión Real de 1784](#)). La Comisión evaluó si los supuestos efectos de este nuevo método de curación se debían a alguna fuerza “real” o a las “ilusiones de la mente”. Se les vendaron los ojos a las personas y se les informó que estaban recibiendo o no estaban recibiendo magnetismo cuando, en realidad, en ocasiones sucedía lo contrario. Las personas estudiadas sintieron los efectos del “magnetismo animal” sólo cuando se les decía que estaban recibiendo el tratamiento pero no en la situación opuesta (Kaptchuk 1998; Schulz et al. 2002).



La utilización de placebos para lograr el método ciego

Pocos años después de las pruebas de los efectos del magnetismo animal, John Haygarth llevó a cabo un experimento en el que utilizó un dispositivo de simulación (placebo) para lograr el ciego ([Haygarth 1800](#)). La caricatura que acompaña a este párrafo muestra a un médico tratando a un cliente adinerado con un mecanismo patentado y comercializado por Elisha Perkins. Perkins afirmaba que sus “tractores” —pequeñas varillas de metal— curaban una variedad de dolencias a través de una “fuerza electrofísica”. En un folleto titulado *Of the imagination as a cause and as a cure of disorders of the body: exemplified by fictitious tractors* (En una serie de pacientes que desconocían los detalles de su evaluación, utilizó un estudio cruzado para comparar los “tractores” metálicos patentados (que, según se suponía, trabajaban mediante “fuerza electrofísica”) con “tractores” de madera, iguales en apariencia (“tractores placebo”). No pudo detectar ningún beneficio a favor de los tractores metálicos ([Haygarth 1800](#)).



La prueba auténtica a la que John Haygarth sometió a los tractores de Perkins es uno de los primeros ejemplos del uso de placebos para lograr el ciego y así reducir los sesgos en las evaluaciones de los resultados de los tratamientos. Los placebos se convirtieron en una herramienta de investigación en los debates sobre homeopatía, la otra forma no convencional de cura importante en el siglo XIX. Los homeópatas utilizaron con frecuencia la evaluación con el método ciego y controles con placebo para sus “demostraciones”, que probaron los efectos de sus remedios en voluntarios sanos ([Löhner 1835](#); Kaptchuk 1998). Una de las pruebas más sofisticadas controladas por placebo fue implementada por la Milwaukee Academy of Medicine en 1879-1880. Este estudio fue “doble ciego”: tanto los pacientes como quienes realizaban el experimento desconocían si el tratamiento era un remedio homeopático genuino o una píldora de azúcar ([Storke et al. 1880](#)).

No fue sino hasta mucho tiempo después que una actitud más escéptica en la medicina convencional permitió reconocer que existía la necesidad de adoptar las evaluaciones ciegas y los placebos para verificar la validez de sus propias afirmaciones. Con la



inspiración que les aportaron principalmente los farmacólogos, los investigadores alemanes adoptaron gradualmente la investigación encubierta. En 1918, por ejemplo, Adolf Bingel informó que había tratado de ser “tan objetivo como fuera posible” al comparar dos tratamientos distintos para la difteria ([Bingel 1918](#)). Evaluó si él mismo o sus colegas podían adivinar qué pacientes habían recibido cuál tratamiento: “No confié sólo en mi propio juicio sino que pedí las opiniones de los médicos ayudantes de la sala de difteria, sin informarles cuál era la naturaleza del suero que se estaba probando. De esta manera, su opinión estaría totalmente desprovista de sesgos. Tengo enormes deseos de que mis observaciones se verifiquen en forma independiente, y para ello recomiendo fervorosamente este método ‘ciego’.” ([Bingel 1918](#)). De hecho, no se detectó ninguna diferencia entre los dos tratamientos. Se desarrolló en Alemania una sólida tradición de evaluaciones ciegas, que el farmacólogo clínico Paul Martini codificó ([Martini 1932](#)).



La evaluación ciega en el mundo moderno angloparlante se introdujo por primera vez cuando los farmacólogos recibieron la influencia tanto de la tradición alemana como de un movimiento indígena de “curanderismo”, que utilizaban la evaluación encubierta (Kaptchuk 1998). Así, para la década de 1930, se habían convertido en líderes en el uso de controles con placebo en los experimentos clínicos. Dos de las primeras pruebas auténticas del Consejo Médico de Investigación del Reino Unido, por ejemplo, se hicieron para tratamientos del resfriado común. Habría sido muy difícil interpretar sus resultados si no se hubiera utilizado el “doble ciego” para evitar que los enfermos y los médicos supieran qué pacientes habían recibido los nuevos medicamentos y cuáles habían recibido placebo ([MRC 1944](#); [MRC 1950](#)). La enérgica defensa que hace Harry Gold de la importancia de las evaluaciones ciegas parece haber influido de manera particular en los Estados Unidos ([Congreso sobre terapias 1954](#)).

Cegamiento de los observadores ante la imposibilidad de cegar a los pacientes y a los médicos

En determinadas circunstancias es sencillamente imposible ocultar a los pacientes y a los médicos la identidad de los tratamientos que se comparan. Tal es el caso de tratamientos quirúrgicos que se comparan con tratamientos con fármacos o con ningún tratamiento. Aún en estos casos, sin embargo, pueden tomarse medidas para reducir los sesgos en la evaluación de los resultados del tratamiento: es posible ocultar qué pacientes han recibido qué tratamientos a observadores independientes. A principios de la década de 1940, por ejemplo, una prueba comparó pacientes con tuberculosis pulmonar que recibieron el tratamiento estándar —reposo en cama— con otros pacientes que recibieron, además, inyecciones de estreptomycin. Los investigadores sintieron que no sería ético inyectar placebos inactivos a los pacientes asignados al grupo que haría sólo reposo en cama para “cegar” a los pacientes y a los médicos que los trataban ([MRC 1948](#)), pero tomaron otras precauciones para reducir los sesgos en la evaluación de los criterios de valoración. Si bien el riesgo de una evaluación prejuiciosa era muy bajo para el criterio de valoración principal (la supervivencia), la subjetividad podría haber influido sobre la evaluación de las radiografías de tórax. En consecuencia, las radiografías fueron evaluadas por médicos que desconocían si estaban evaluando los resultados en un paciente que había recibido estreptomycin o en un paciente que sólo había hecho reposo en cama.

Junto con la aleatorización, la evaluación encubierta —utilizando placebos siempre que sea posible— se ha convertido en uno de los componentes metodológicos esenciales en las pruebas auténticas de los tratamientos.

Cite as: Editorial commentary (2007). Diferencias en la manera en que se evalúan los resultados de los tratamientos. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [control of observer bias](#)

Next essay: [Interpretación de comparaciones no prejuiciosas](#)

Select essay:

Referencias

Bingel A (1918). Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum. Deutsches Archiv für Klinische Medizin 125:284-332.

Commission Royale (1784). Rapport des commissaires chargés par le roi du magnetisme animal. Paris: Imprimerie royale.

Conference on Therapy (1954). How to evaluate a new drug. American Journal of Medicine 17:722-727.

Faure (1759). Recueil des pieces qui ont concouru pour le prix de L'Académie Royale de Chirurgie. Vol 8. Paris, P.Al Le Prieur.

Haygarth J (1800). Of the imagination, as a cause and as a cure of disorders of the body: exemplified by fictitious tractors, and epidemical convulsions. Bath: R. Crutwell.

Kaptchuk TJ (1998). Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine* 72: 389-433.

Löhner G (1835), on behalf of a Society of truth-loving men. Die Homoöopathischen Kochsalzversuche zu Nürnberg [The homeopathic salt trials in Nuremberg].

Martini P (1932). *Methodenlehre der Therapeutischen Untersuchung*. Berlin:Springer.

Medical Research Council (1944). Clinical trial of patulin in the common cold. *Lancet* 2: 373-375.

Medical Research Council (1948). Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 2: 769-782.

Medical Research Council (1950). Clinical trials of antihistaminic drugs in the prevention and treatment of the common cold. *BMJ* 2: 425-431.

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273: 408-412.

Schulz KF, Chalmers I, Altman D (2002). The landscape and lexicon of blinding. *Annals of Internal Medicine* 136: 254-259.

Storke EF, Martin R, Rosenkrans EM, Ford J, Schloemilch A, McDermott GC, Carlson OW (1880). Final report of the Milwaukee test of the thirtieth dilution. *Homeopathic Times: A Monthly Journal of Medicine, Surgery and the Collateral Sciences* 7: 280-281.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)[jameslindlibrary.org](#)

Interpretación de comparaciones no prejuiciosas

Una comparación auténtica de tratamientos es aquella que [evita los sesgos](#). Esto requiere tomar medidas para minimizar los sesgos provenientes de las [diferencias entre los pacientes que se comparan](#) y los provenientes de las [diferencias en la forma en que se evalúan los resultados del tratamiento](#).

No obstante, aún cuando se eviten estos sesgos, a menudo no es sencillo interpretar las comparaciones no prejuiciosas. Por ejemplo, ¿se ha tenido en cuenta alguna [diferencia entre los tratamientos que se pretende administrar y los que realmente se recibieron?](#) y, ¿se ha tenido en cuenta [la obra de la casualidad?](#)

A veces, un estudio nuevo proporciona evidencia muy contundente de los efectos de un tratamiento. Por ejemplo, decenas de miles de personas participaron en un notable estudio que demostró que un comprimido de aspirina podía reducir substancialmente el riesgo de muerte en las personas con infarto (ISIS-2 1988). Es, sin embargo, extraordinario que un único estudio genere una evidencia tan contundente. Por lo tanto, al leer los informes de la mayoría de los estudios, es importante preguntar si la nueva evidencia ha sido integrada en [revisiones sistemáticas de todo el resto de la evidencia pertinente](#). Si lo fue, ¿se tomaron medidas durante ese proceso de síntesis para minimizar el impacto de [los sesgos en los informes sobre la evidencia disponible](#) y de [los sesgos en la selección de la misma?](#) ¿Se consideró utilizar el potencial del [metanálisis para reducir la influencia de la casualidad?](#)

Cite as: Editorial commentary (2007). Interpretación de comparaciones no prejuiciosas. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Diferencias entre el tratamiento que se pretende administrar y el que realmente se recibe](#)

Select essay:

Referencias

ISIS-2 Second International Study of Infarct Survival Collaborative Group (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 2: 349 60.

[Home](#)[Contenido](#)

[Comments welcome](#)

Home

Contenido

jameslindlibrary.org

Interpretación de comparaciones no prejuiciosas:

Diferencias entre el tratamiento que se pretende administrar y el que realmente se recibe

Las pruebas auténticas de los tratamientos médicos deben planificarse con cuidado. Nos referimos a los documentos que exponen esta planificación como protocolos y, entre otras cosas, en ellos se especifican los detalles sobre los tratamientos que van a compararse. Sin embargo, aún los planes más organizados no siempre funcionan como se espera. Los tratamientos que los pacientes realmente reciben en las pruebas a veces difieren de los que se tenía la intención de administrarles. Al interpretar los resultados de las comparaciones entre tratamientos deben tenerse en cuenta estas desviaciones de la intención.

Una de las razones por las que se introdujeron los placebos en la evolución de las pruebas auténticas de los tratamientos médicos fue reducir estas desviaciones del tratamiento que se intentaba administrar (Kaptchuk 1998). Pero todo puede ir por mal camino aún en los estudios controlados con placebo. Durante la segunda Guerra Mundial, personas que estaban resfriadas recibieron una solución de un fármaco denominado patulina y se las comparó con otras personas que sólo recibieron el líquido en el que se había disuelto la medicación ([MRC 1944](#)). El análisis de los resultados no reveló ningún efecto beneficioso de la medicación pero entonces surgió una duda: el líquido utilizado para disolver el medicamento podría haberlo inactivado. En otras palabras, ¡más de 1000 pacientes podrían haber participado en la comparación de dos tratamientos inactivos! Por fortuna, las pruebas confirmaron que la patulina utilizada en el estudio sí había estado activa, aunque no había generado sobre los resfriados efectos que pudieran detectarse (Chalmers y Clarke 2004).

Los tratamientos que se reciben pueden diferir de los tratamientos que se intenta administrar debido a una cantidad de razones. Los médicos, por ejemplo, pueden decidir que el tratamiento al que se han asignado algunos de sus pacientes en una comparación formal de tratamientos no debería administrárseles; los pacientes pueden rechazar el tratamiento que se les ha asignado o no seguirlo como deberían; puede que se administren dosis diferentes de las previstas; o es posible que se agote la provisión de uno de los tratamientos.

Cuando surgieron diferencias en los resultados de tratamientos aparentemente idénticos para la leucemia administrados a niños británicos y estadounidenses, por ejemplo, la investigación reveló que los resultados peores obtenidos en Gran Bretaña reflejaban la resistencia de los médicos británicos a continuar con la quimioterapia cuando se desarrollaron efectos tóxicos desagradables (Grupo de Trabajo del Consejo Médico de Investigación sobre leucemia en niños 1986).

Por estas razones, las interpretaciones de pruebas auténticas deben tener en cuenta la posibilidad de que los tratamientos recibidos no fueran aquellos que se había intentado administrar. Si han surgido discrepancias entre la intención y la práctica, es importante considerar las implicaciones para interpretar la evidencia.

Cite as: Editorial commentary (2007). Diferencias entre el tratamiento que se pretende administrar y el que realmente se recibe. The James Lind Library (www.jameslindlibrary.org).

Next essay: [Cómo considerar la obra de la casualidad](#)

Select essay:

Referencias

Chalmers I, Clarke M (2004). The 1944 Patulin Trial: the first properly controlled multicentre trial conducted under the aegis of the British Medical Research Council. *International Journal of Epidemiology* 32: 253-260.

Kaptchuk TJ (1998). Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine* 72: 389-433.

Medical Research Council (1944). Clinical trial of patulin in the common cold. *Lancet* 2: 373-375.

Medical Research Council Working Party on Leukaemia in Children (1986). Improvement in treatment for children with acute lymphoblastic leukaemia. *Lancet* 1: 408-11.

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

Interpretación de comparaciones no prejuiciosas:

Cómo considerar la obra de la casualidad

Cuando se comparan dos tratamientos, cualquier diferencia en los resultados puede deberse, simplemente, a la casualidad. Tomemos, por ejemplo, la comparación de un nuevo tratamiento con un tratamiento estándar en el que 4 personas mejoraron con el primero y 6 mejoraron con el segundo. Deducir con certeza que el nuevo tratamiento fue peor que el estándar sería, claramente, un error, dado que estos resultados podrían sólo reflejar la obra de la casualidad. Si se repitiera la comparación, la cantidad de pacientes que mejoraran podría revertirse (6 contra 4) o podría ser la misma (5 contra 5) o la relación podría ser otra.

Sin embargo, si 40 personas mejoraran con el tratamiento nuevo y 60 con el estándar, sería menos probable que la casualidad pudiera explicar la diferencia. Y si 400 personas mejoraran con el nuevo tratamiento y 600 con el estándar, quedaría claro que el nuevo tratamiento tendría muchas probabilidades de ser peor que el tratamiento estándar. La manera de reducir las probabilidades de que la casualidad nos impulse al error al comparar tratamientos es, por lo tanto, asegurarse de que las pruebas auténticas incluyan cantidades suficientemente grandes de personas que muestren los criterios de valoración que uno espera que influyan, como la mejoría o el deterioro.

En determinadas circunstancias, para obtener cálculos confiables de los efectos del tratamiento, es necesario que participe un número muy grande de personas: miles y, a veces, decenas de miles. Se necesitan grandes cantidades de participantes, por ejemplo, si los criterios de valoración que nos interesan son raros, como podría ser el caso de infartos y accidentes cerebrovasculares en mujeres de edad mediana, aparentemente sanas, que utilizan la terapia de reemplazo hormonal (TRH). También se necesitan grandes números de personas si se desea detectar con certeza efectos moderados pero importantes de los tratamientos, por ejemplo, la reducción en un 20 por ciento del riesgo de muerte temprana en personas con infarto.

Para evaluar el papel que pudo haber desempeñado la casualidad en los resultados de pruebas auténticas, los investigadores utilizan "pruebas de significación estadística". Cuando los estadísticos y otras personas se refieren a "diferencias significativas" entre los tratamientos, normalmente se refieren a la significación estadística. Las diferencias estadísticamente significativas entre tratamientos no necesariamente tienen importancia práctica. Pero, de todos modos, las pruebas de significación estadística son importantes porque nos ayudan a evitar que erróneamente deduzcamos que existen diferencias reales entre los tratamientos cuando, de hecho, no las hay. Esta deducción equivocada es lo que a veces llamamos error tipo I.

También es importante tener en cuenta una cantidad suficientemente amplia de resultados del tratamiento para evitar un peligro mucho más común: deducir que no existen diferencias entre los tratamientos cuando, de hecho, sí las hay. Esta deducción equivocada es lo que a veces llamamos error tipo II. Thomas Graham Balfour era consciente de este tipo de riesgo cuando interpretó los resultados de su prueba del potencial de la belladonna para evitar que los huérfanos a su cuidado desarrollaran fiebre escarlata ([Balfour 1854](#)). En comparación con 2 de 75 niños que no recibieron *belladonna*, 2 de 76 niños asignados al grupo que sí la recibía desarrollaron fiebre escarlata. Balfour observó que "las cantidades son demasiado pequeñas para justificar las deducciones respecto del potencial profiláctico de la *belladonna*". Si hubieran sido más los niños que desarrollaron fiebre escarlata, quizá Balfour hubiera podido arribar a una conclusión más certera sobre los posibles efectos de la *belladonna*. En cambio, sólo observó que 4 casos de fiebre escarlata entre 151 niños era un número demasiado pequeño para llegar a una conclusión confiable.

Un método que reduce la probabilidad de que los efectos de la casualidad nos engañen es estimar un rango de diferencias entre los tratamientos dentro del cual deberían caer las diferencias reales ([Gavarret 1840](#); Huth 2006). Estos cálculos de los rangos se conocen como intervalos de confianza. Como se ilustra en el párrafo inicial de este ensayo, es probable que la repetición de una comparación entre tratamientos arroje cálculos variados de los efectos diferenciales de los tratamientos sobre los criterios de valoración, en particular si los cálculos se hacen en función de cantidades pequeñas de dichos criterios. Los intervalos de confianza tienen en cuenta esta variación, son más informativos que las meras pruebas de significación estadística y, por lo tanto, más útiles para reducir la probabilidad de que la intervención de la casualidad nos conduzca a error.

Las pruebas estadísticas y los intervalos de confianza —ya sea para el análisis de estudios individuales o en el [metanálisis](#) de una cantidad de estudios diferentes pero similares— nos ayudan a determinar la dimensión de la influencia de la casualidad y a no deducir que los efectos y las diferencias entre los tratamientos existen cuando no los hay o no existen cuando sí los hay.

Cite as: Editorial commentary (2007). Cómo considerar la obra de la casualidad. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [taking account of the play of chance](#)

Next essay: [Identificación de los efectos no previstos de los tratamientos](#)

Select essay:

Referencias

Balfour TG (1854). Quoted in West C. Lectures on the Diseases of Infancy and Childhood. London, Longman, Brown, Green and Longmans, p 600.

Gavarret LDJ (1840). Principes généraux de statistique médicale: ou développement des règles qui doivent présider à son emploi. Paris: Bechet jeune & Labé.

Huth EJ (2006). Jules Gavarret's Principes Généraux de Statistique Médicale: a pioneering text on the statistical analysis of the results of treatments.

[Home](#)

[Contenido](#)

[Comments welcome](#)

Identificación de los efectos no previstos de los tratamientos

Resulta prácticamente inevitable que surjan efectos imprevistos cuando se introducen nuevos tratamientos de una manera más amplia. Las pruebas iniciales –por ejemplo, las exigidas para otorgar la licencia a fármacos nuevos– incluyen como máximo unos pocos cientos o miles de personas tratadas durante algunos meses. En esta etapa, es probable que sólo se detecten los efectos no previstos relativamente frecuentes y de corto plazo.

Los efectos menos comunes del tratamiento o los que necesiten un cierto tiempo para manifestarse no se descubrirán hasta que las pruebas del tratamiento hayan durado lo suficiente o hasta que la utilización de los tratamientos se haya extendido más. Además, los nuevos tratamientos suelen administrarse a personas que pueden tener diferencias importantes con quienes participaron en las pruebas originales: pueden tener más o menos edad, ser de otro sexo, estar más o menos enfermos, vivir en circunstancias diferentes o padecer otros problemas de salud además de la enfermedad para la que se administra el tratamiento. Es posible que estas divergencias modifiquen los efectos del tratamiento y que surjan efectos nuevos e imprevistos (vea el número especial de [BMJ del 3 de julio de 2004](#)).



La detección y la verificación de los efectos no previstos, ya sean [adversos](#) o [beneficiosos](#), suele producirse de manera diferente de la de los métodos utilizados para evaluar los efectos esperados de los nuevos tratamientos. A veces, los profesionales de la salud o los pacientes sospechan algunos efectos imprevistos de los tratamientos desde el comienzo. La identificación de cuáles de estos presentimientos iniciales reflejan los efectos verdaderos de los tratamientos representa un desafío ya conocido para los lectores de otros ensayos de esta serie: evitar ser inducidos a error por los [sesgos](#) y [la obra de la casualidad](#).

Si el efecto imprevisto de una terapia es muy sorprendente y se produce con marcada frecuencia después de que se ha empleado el tratamiento, es posible que los profesionales de la salud o los pacientes lo noten espontáneamente. Por ejemplo, casi no se registran casos de bebés que nazcan sin extremidades, por lo tanto, cuando se produjo un aumento repentino en el número de estos nacimientos en la década de 1960, naturalmente surgieron interrogantes. Todas las madres de estos bebés habían tomado [talidomida](#), un medicamento de reciente comercialización en ese entonces, recetado a principios del embarazo para tratar las náuseas; por lo tanto, era probable que ésta fuera la causa y que se necesitaran más evaluaciones. Los efectos imprevistos beneficiosos de los fármacos suelen detectarse en forma similar, como sucedió, por ejemplo, cuando se descubrió que un medicamento para tratar la psicosis también reducía el colesterol (Goodwin 1991).

Cuando se observan relaciones tan asombrosas, a menudo se las confirma como efectos imprevistos reales del tratamiento ([Venning 1982](#)). No obstante, muchos presentimientos sobre los efectos no previstos de los tratamientos se basan en evidencias mucho menos convincentes. En consecuencia, al igual que sucede con las pruebas diseñadas para detectar los efectos esperados de los tratamientos, la planificación de pruebas para confirmar o descartar efectos imprevistos sospechados pero menos asombrosos requiere [evitar los sesgos en las comparaciones](#).

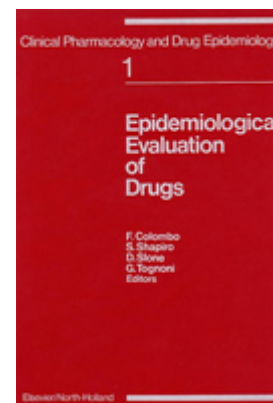
Los estudios para determinar si los efectos imprevistos sospechados de un tratamiento son reales deben ajustarse al principio de comparación entre iguales. La asignación aleatoria a los tratamientos es la forma ideal de lograr que se cumpla este principio. Sin embargo, son pocas las veces en que los efectos sospechados de los tratamientos se pueden investigar mediante análisis adicionales o mediante el seguimiento de personas aleatoriamente asignadas a tratamientos antes de que se los administre (Hemminki y McPherson 1997). Por lo tanto, el desafío consiste en reunir grupos de comparación no prejuiciosa de otras formas, a menudo recurriendo a información recabada sistemáticamente durante la atención médica.

De hecho, en estos estudios ayuda que los efectos sospechados no se previeran en el momento en que se tomaron las decisiones respecto del tratamiento. Esto significa que el riesgo de la afección sospechada no se había podido tener en cuenta en el momento en que se seleccionaron las personas para el tratamiento: el efecto imprevisto suele ser una afección o una enfermedad distinta de la afección o la enfermedad para las cuales se indicó el tratamiento (Vandenbroucke 2004a).

Por ejemplo, cuando se introdujo la terapia de reemplazo hormonal (TRH) para tratar los síntomas de la menopausia, era improbable que se tuviera en cuenta el riesgo de que las mujeres sufrieran trombosis venosa ya que la mayoría de los médicos y de las mujeres no la consideraban pertinente. En consecuencia, no había motivos para prever que el riesgo de desarrollar trombosis venosa fuera diferente para las mujeres a quienes se les prescribiría la TRH que para aquéllas que no recibirían el tratamiento. De este modo se estableció el fundamento para las pruebas auténticas, y éstas demostraron que la TRH aumenta el riesgo de desarrollar trombosis venosa.

Cuando un efecto imprevisto sospechado se relaciona con un tratamiento para un problema común de salud (como un infarto) pero no se produce con mucha frecuencia con el nuevo tratamiento (o éste no lo alivia totalmente), es necesario realizar una vigilancia en gran escala de las personas que reciben el tratamiento para detectar el efecto no previsto. Por ejemplo, aunque algunas personas creían que la [aspirina](#) reduciría el riesgo de ataque cardíaco y comenzaron pruebas auténticas de esta teoría en pacientes en los últimos años de la década de 1960 ([Elwood et al. 1974](#)), la mayoría de las personas hubiera pensado que la teoría era bastante inverosímil. El cambio se produjo cuando se realizó un estudio extenso para detectar los efectos adversos imprevistos de los fármacos: los investigadores observaron que, en comparación con otros pacientes aparentemente similares, era menos probable que las personas que ingresaban al hospital con infartos hubieran tomado aspirina recientemente ([Boston Collaborative Drug Surveillance Group 1974](#)). Estos resultados coincidieron con los de una prueba auténtica, en la que las personas habían sido asignadas aleatoriamente para recibir o no recibir aspirina después de un infarto. Los dos informes se publicaron juntos en el mismo número del British Medical Journal (BMJ 1974).

A fines de la década de 1970 se establecieron las reglas básicas para detectar e investigar los efectos imprevistos de los tratamientos ([Jick 1977](#); [Colombo et al. 1977](#)). Estas reglas se basaron en la experiencia colectiva producto de la investigación de los efectos imprevistos que se habían acumulado después del desastre de la [talidomida](#). Los requisitos para una clase importante de investigaciones —los estudios de control de casos de los posibles efectos adversos de los tratamientos— se volcaron en un trabajo escrito basado en las experiencias de investigadores en Boston y Oxford ([Jick and Vessey 1978](#)). Con un gran número de tratamientos potentes introducidos a partir de esa fecha, este aspecto de las pruebas auténticas sigue suponiendo retos tan importantes hoy como lo hacía entonces (Vandenbroucke 2004b; Vandenbroucke 2006; Papanikolaou et al. 2006).



Como se destacó en ensayos previos de esta serie, es importante reconocer que los informes individuales que sugieren o descartan sospechas sobre efectos imprevistos de tratamientos pueden inducir a error. Al igual que sucede con otras pruebas auténticas de los tratamientos, se deben investigar los posibles efectos imprevistos del tratamiento utilizando [revisiones sistemáticas](#) de toda la evidencia pertinente, como las que confirmaron la relación entre la TRH y las cardiopatías, y el accidente cerebrovascular y el cáncer de mama (Hemminki y McPherson 1997; Collaborative Group on Hormonal Factors in Breast Cancer 1997).

Cita: Editorial commentary (2007). Identificación de los efectos no previstos de los tratamientos. The James Lind Library (www.jameslindlibrary.org).

Mostrar JLL registros: ilustrando [la identificación de los efectos no previstos](#)

Siguiente ensayo: [Revisiones sistemáticas de toda la evidencia pertinente](#)

Seleccione un ensayo:

Referencias

Boston Collaborative Drug Surveillance Group (1974). Regular aspirin intake and acute myocardial infarction. BMJ 1:440-443.

Collaborative Group on Hormonal Factors in Breast Cancer (1997). Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet 350:1047-1059 .

Colombo F, Shapiro S, Slone D, Tognoni G, eds (1977). Epidemiological Evaluation of Drugs. Amsterdam: Elsevier/ North Holland Biomedical Press, 1977.

Elwood PC, Cochrane AL, Burr ML, Sweetnam PM, Williams G, Welsby E, Hughes SJ, Renton R (1974). A randomised controlled trial of acetyl salicylic acid in the secondary prevention of mortality from myocardial infarction. BMJ 1:436-440.

Goodwin JS (1991). The empirical basis for the discovery of new therapies. Perspectives in Biology and Medicine 35:20-36.

Hemminki E, McPherson K (1997). Impact of postmenopausal hormone therapy on cardiovascular events and cancer:

pooled data from clinical trials. *BMJ*; 315:149-153.

Jick H (1977). The discovery of drug-induced illness. *New England Journal of Medicine* 296:481-485.

Jick H, Vessey M (1978). Case-control studies in the evaluation of drug-induced illness. *American Journal of Epidemiology* 107:1-7.

Papanikolaou PN, Christidi GD, Ioannidis JPA (2006). Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 174:635-641.

Vandenbroucke JP (2004a). When are observational studies as credible as randomised trials? *Lancet* 363:1728-1731.

Vandenbroucke JP (2004b). Benefits and harms of drug treatments. *BMJ* 329:2-3.

Vandenbroucke JP (2006). What is the best evidence for determining harms of medical treatment? *CMAJ* 174:645-646.

Venning GR (1982). Validity of anecdotal reports of suspected adverse drug reactions: the problem of false alarms. *BMJ* 284:249-254.

[Página inicial](#) [Contenidos](#)



Traducido del original en inglés por la Organización Panamericana de la Salud.
Derechos reservados sobre los materiales traducidos.

Translated from the original English by the Pan American Health Organization.

All rights reserved. www.paho.org

[Comentarios](#)

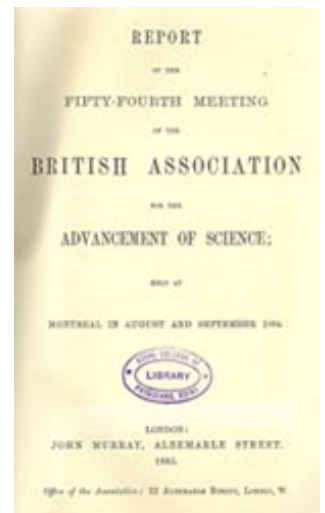
[Home](#)[Contenido](#)jameslindlibrary.org

Revisiones sistemáticas de toda la evidencia pertinente

[Austin Bradford Hill](#), uno de los pioneros del siglo XX en las pruebas auténticas para los tratamientos, observó que quienes leían los informes de las investigaciones querían respuestas a las cuatro preguntas siguientes: '¿Por qué empezó?', '¿Qué hizo?', '¿Que encontró?' y '¿Y esto qué significa?' (Hill 1965). Es especialmente importante la calidad de la respuesta a la última pregunta de Hill porque es el elemento del informe de investigación que tiene mayor probabilidad de influir sobre las opciones y las decisiones reales acerca de los tratamientos.

Son muy escasas las ocasiones en que una única prueba auténtica de un tratamiento ofrecerá evidencia suficientemente sólida como para proporcionar una respuesta segura a la pregunta '¿Y esto qué significa?'. Una prueba auténtica de un tratamiento suele ser una de una serie de estudios que abordan el mismo interrogante. Por lo tanto, para obtener una respuesta confiable para la pregunta '¿Y esto qué significa?', es importante interpretar la evidencia de una prueba auténtica concreta en el contexto de una evaluación exhaustiva de toda la evidencia obtenida de las pruebas auténticas que han abordado el interrogante en cuestión.

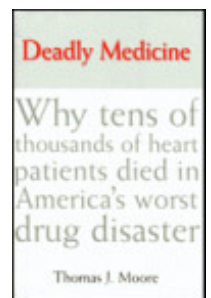
El presidente de la British Association for the Advancement of Science manifestó la necesidad de observar este principio hace más de un siglo:



"Si, como se supone a veces, la ciencia consistiera solamente en la mera y laboriosa acumulación de hechos, muy pronto quedaría paralizada, aplastada, por así decirlo, bajo su propio peso... Por consiguiente, dos procesos operan simultáneamente: la recepción de material nuevo y la digestión y la asimilación del antiguo...El trabajo que merece, pero me temo que no siempre recibe, el mayor crédito es aquél en el que el descubrimiento y la explicación avanzan juntos, en el que no solamente se presentan los nuevos hechos, sino que también se señala su relación con los precedentes". ([Rayleigh 1885](#))

Quizá se deba a que la aplicación de este principio en la práctica aún obtiene poco crédito en los círculos académicos, que muy pocos informes de pruebas auténticas de tratamientos analizan sus resultados en el contexto de una evaluación sistemática del resto de la evidencia pertinente (Clarke et al. 2002). En consecuencia, suele ser difícil que quienes leen esos informes obtengan una respuesta confiable a la pregunta '¿Y esto qué significa?' de los informes de nuevas investigaciones.

Como se observó en un ensayo explicativo previo, es peligroso, poco ético y un desperdicio embarcarse en nuevas pruebas de tratamientos médicos sin haber revisado sistemáticamente lo que se puede aprender de investigaciones previas (vea [Por qué las comparaciones deben abordar una incertidumbre genuina](#)). Comunicar los resultados de nuevas pruebas sin haber interpretado la nueva evidencia a la luz de la evaluación sistemática del resto de la evidencia pertinente también es peligroso, ya que genera demoras en la identificación de tratamientos tanto útiles como perjudiciales (Antman et al. 1992). Por ejemplo, entre la década de 1960 y los primeros años de la de 1990, se realizaron más de 50 pruebas auténticas de fármacos para reducir las anomalías de la frecuencia cardíaca en las personas con infarto antes de que se dieran cuenta de que estos fármacos estaban matando a los pacientes. Si cada informe hubiera evaluado los resultados de las nuevas pruebas en el contexto de toda la evidencia pertinente, los efectos mortales de los fármacos se podrían haber identificado diez años antes y se podrían haber evitado muchas muertes innecesariamente prematuras.



En la era de las publicaciones electrónicas debería ser posible superar las limitaciones presentes en la mayoría de los informes de las nuevas investigaciones (Chalmers y Altman 1999; Smith y Chalmers 2001). Sin embargo, más

que fundamentar las conclusiones acerca de los tratamientos en uno o unos pocos estudios individuales, los usuarios de la evidencia de las investigaciones se inclinan cada vez más a recurrir a información confiable de [revisiones actualizadas y sistemáticas de toda la evidencia pertinente y confiable](#), porque se reconoce, cada vez más, que dichas reseñas proporcionan la base más sólida para extraer conclusiones acerca de los efectos de los tratamientos médicos.

De la misma forma en que es importante tomar medidas para evitar ser inducido a error por los [sesgos](#) y [la obra de la casualidad](#) en la planificación, la realización, el análisis y la interpretación de cada prueba auténtica concreta de los tratamientos, es esencial que se tomen medidas para planificar, realizar, analizar e interpretar las revisiones sistemáticas. Esto implica:

- especificar el interrogante que abordará la revisión sistemática;
- definir los criterios de elegibilidad de los estudios que se incluirán;
- identificar (todos) los estudios que posiblemente reúnan los requisitos;
- aplicar los criterios de elegibilidad de manera que se limiten los sesgos;
- recabar un porcentaje lo más elevado posible de la información pertinente proveniente de los estudios;
- utilizar el metanálisis y una variedad de análisis para examinar esta información, si corresponde y es factible;
- y
- preparar un informe estructurado.

La rápida evolución de los métodos para mejorar la confiabilidad de las propias revisiones es una manifestación del reconocimiento cada vez mayor de la importancia fundamental que tienen las revisiones sistemáticas en la evaluación de los efectos de los tratamientos. La primera edición de un libro titulado *Systematic Reviews* (Revisiones sistemáticas) contaba con menos de 100 páginas ([Chalmers and Altman 1995](#)): tan sólo seis años después, la segunda edición terminó en unas casi 500 páginas e incluyó estrategias que evolucionaban rápidamente para aumentar la cantidad de información obtenida de las investigaciones (Egger et al. 2001).

En la actualidad, existen importantes desarrollos en los métodos utilizados para preparar revisiones sistemáticas, incluidas aquéllas necesarios para identificar los efectos imprevistos de los tratamientos (Glasziou et al 2004) y para incorporar los resultados de la investigación que describe y analiza la experiencia de personas que administraron y recibieron los tratamientos (Thomas 2004). A medida que vaya surgiendo, el material pertinente se incorporará a *The James Lind Library*.

Cite as: Editorial commentary (2007). Revisiones sistemáticas de toda la evidencia pertinente. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [systematic review](#)

Next essay: [Cómo manejar los sesgos en los informes de la evidencia disponible](#)

Select essay:

Referencias

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* 268: 240-48.

Chalmers I, Altman DG (1995). *Systematic Reviews*. London: BMJ Publications.

Chalmers I, Altman DG (1999). How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 353: 490-493.

Egger M, Davey Smith G, Altman D (2001). *Systematic Reviews in Health Care: meta-analysis in context*. 2nd Edition of *Systematic Reviews*. London: BMJ Books.

Glasziou P, Vandenbroucke J, Chalmers I (2004). Assessing the quality of research *BMJ* 328: 39-41.

Hill AB (1965). Cited in 'The reasons for writing'. *BMJ* 4: 870.

Rayleigh (1885). Address by the Rt. Hon. Lord Rayleigh. In: Report of the fifty-fourth meeting of the British Association for the Advancement of Science; held at Montreal in August and September 1884, London: John Murray.

Smith R, Chalmers I (2001). Britain's gift: a 'Medline' of synthesized evidence. BMJ 323:1437-1438.

Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J (2004). Integrating qualitative research with trials in systematic reviews BMJ 328:1010-1012.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

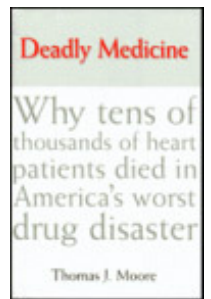
Cómo manejar los sesgos en los informes de la evidencia disponible

[Evitar el sesgo en las comparaciones](#) implica identificar y tener en cuenta toda la evidencia confiable pertinente de las revisiones sistemáticas. Esta tarea presenta muchos desafíos, en especial debido a que parte de la evidencia pertinente no se publica ya que se toman decisiones prejuiciosas acerca de qué resultados de las investigaciones se presentan y se aceptan para su publicación. Los estudios que hayan arrojado resultados “decepcionantes” o “negativos” tienen menos probabilidades de ser comunicados que otros trabajos. A menudo, esta selección se llama “sesgo de publicación” o “sesgo de comunicación”.

Hace siglos que se reconocen estos sesgos o prejuicios de comunicación (Dickersin 2004a). Por ejemplo, en 1792, James Ferriar subrayó la importancia de comunicar los fracasos en los tratamientos así como se comunican los éxitos (Ferriar 1792). Este principio se reiteró en un artículo editorial publicado en el Boston Medical and Surgical Journal casi un siglo después ([Editorial 1909](#)).

En la actualidad, existe un amplio cuerpo de evidencia que confirma que el sesgo de comunicación representa un problema esencial. Asimismo, la evidencia sugiere que el sesgo de comunicación deriva principalmente de la falta de elaboración o de presentación de informes de investigaciones para su publicación por parte de los investigadores y no de rechazos prejuiciosos de los informes presentados por parte de los editores de publicaciones (Dickersin 2004b). Investigaciones recientes han revelado un problema adicional: si los cálculos aproximados de los efectos de los tratamientos sobre algunos de los criterios de valoración estudiados no apoyan las conclusiones de los investigadores, a veces estos datos tampoco se comunican (Chan et al. 2004).

Por ejemplo, si se hubieran comunicado todos los estudios sobre los efectos de administrar medicamentos para reducir las anomalías de la frecuencia cardíaca en los pacientes con infarto, se podrían haber evitado cientos de miles de muertes provocadas por estos fármacos. En 1993, Dr. Cowley y sus colaboradores destacaron cómo un estudio no publicado que había sido realizado hacía 13 años podría haber “constituido un aviso precoz del problema que se avecinaba”. Nueve de los 49 pacientes asignados al fármaco contra la arritmia (lorcainida) habían muerto en comparación con un único paciente de un número similar del grupo que recibió placebo. “Cuando realizamos nuestro estudio en 1980”, informaron, “creíamos que el aumento en la tasa de mortalidad era un efecto de la casualidad...El desarrollo de la lorcainida se interrumpió por razones comerciales y, por lo tanto, este estudio jamás se publicó; ahora constituye un ejemplo sólido de ‘sesgo de publicación’” (Cowley et al. 1993).



Los sesgos de comunicación suelen conducir a conclusiones de que los tratamientos médicos son más útiles de lo que en realidad son. En consecuencia, pueden generar sufrimientos innecesarios y muerte, así como también el desperdicio de recursos gastados en tratamientos ineficaces o peligrosos (Chalmers 2004). Las personas que aceptan participar en pruebas de tratamientos a pedido de los investigadores, suponen que su participación dará por resultado una expansión de los conocimientos. Los investigadores que no dan a conocer los resultados de sus investigaciones infringen este contrato implícito entre investigadores y participantes.

El sesgo que induce a la comunicación parcial o nula de las investigaciones es una falta de ética y mala conducta científica (Chalmers 1990). La comunicación selectiva de los estudios patrocinados por la industria farmacéutica constituye un problema especial (Melander et al. 2003), aunque el problema no se limita a aquéllos con intereses comerciales creados. Los comités de ética en la investigación, los médicos defensores de la ética y quienes financian las investigaciones todavía no han hecho lo suficiente para proteger a los pacientes y al público de los efectos adversos del sesgo de comunicación (Savulescu et al. 1996). Las pruebas auténticas de los tratamientos – en especial de los tratamientos en los que hay un interés comercial– seguirán estando comprometidas en tanto los gobiernos y quienes deben proteger los intereses del público toleren esta clase de falta de conducta ética en las investigaciones.

La Organización Mundial de la Salud (OMS) ha propuesto soluciones para abordar el problema del sesgo no identificable en las investigaciones y las publicaciones (o en su difusión): En primer término, establecer normas para el registro y el intercambio de datos para la inscripción de los estudios clínicos. En segundo término, proponer la inscripción de los protocolos de investigación en bases de datos que cumplan con las normas mencionadas antes de que se inicie el reclutamiento de los pacientes. En último término, proponer la implementación de un portal de acceso abierto (www.who.int/ictpr) que reúna los datos de todos los registros y que permita que las personas conozcan los protocolos de investigación futuros, en curso y terminados.

Todos debemos apoyar la iniciativa de la Organización Mundial de la Salud para reducir los sesgos de comunicación

exigiendo la inscripción de todas las pruebas auténticas de tratamiento en sus inicios e insistiendo en la publicación de sus resultados.

Cite as: Editorial commentary (2007). Cómo manejar los sesgos en los informes de la evidencia disponible. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [reporting bias](#)

Other material: view [Monty Python's take](#) on reporting bias - YouTube video

Next essay: [Cómo evitar los sesgos en la selección de la evidencia disponible](#)

Select essay:

Referencias

Chalmers I (1990). Under-reporting research is scientific misconduct. *JAMA* 263:1405-1408.

Chalmers I (2004). In the dark: drug companies should be forced to publish all the results of clinical trials. *New Scientist* 181:19.

Chan A-W, Hróbjartsson A, Haahr M, Gøtzsche PC, Altman DG (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to publications. *JAMA* 291:2457-2465.

Cowley AJ, Skene A, Stainer, Hampton JR (1993). The effect of lorcaïnide on arrhythmias and survival in patients with acute myocardial infarction. *International Journal of Cardiology* 40:161-166.

Dickersin K (2004a). Publication bias: recognising the problem, understanding its origins and scope, and preventing harm. In: Rothstein H, Sutton A, Borenstein M, eds. *Handbook of publication bias*. New York: Wiley.

Dickersin K (2004b). How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9 (1 Suppl): 15-21.

Editorial (1909). The reporting of unsuccessful cases. *Boston Medical and Surgical Journal* 161:263-264.

Ferriar J (1792). *Medical histories and reflexions*. Vol 1. London: Cadell and Davies, 1792.

Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003). Evidence b(i)ased medicine - selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 326:1171-3.

Savulescu J, Chalmers I, Blunt J (1996). Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 313:1390-1393.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)[jameslindlibrary.org](http://www.jameslindlibrary.org)

Revisiones sistemáticas de toda la evidencia pertinente:

Cómo evitar los sesgos en la selección de la evidencia disponible

Los sesgos pueden distorsionar las pruebas de los tratamientos médicos y derivar en conclusiones erróneas. También pueden distorsionar las revisiones de la evidencia. Deberán establecerse en los protocolos planes para la revisión sistemática, como los publicados por [The Cochrane Collaboration](#), que estipulen claramente las medidas que se tomarán para reducir el sesgo.

Esto incluye la especificación clara de:

- las preguntas acerca de los tratamientos que se abordarán en la revisión;
- los criterios que harán que un estudio reúna los requisitos para su inclusión;
- las estrategias que se utilizarán para buscar los estudios que posiblemente reúnan los requisitos; y
- las medidas que se tomarán para minimizar el sesgo en la selección de los estudios y los datos para la inclusión en la revisión (Berlin 1997).

Las diferentes revisiones sistemáticas que abordan lo que parece ser la misma pregunta acerca de los efectos de los tratamientos médicos suelen llegar a conclusiones diferentes. A veces, la divergencia deriva de las sutiles diferencias en las preguntas que se abordan. Otras, refleja variaciones en los materiales y métodos utilizados por los revisores y, en estas circunstancias, es importante discernir cuáles de las revisiones tienen mayores probabilidades de haber reducido satisfactoriamente los sesgos en la asignación.

Asimismo, se debería considerar si los revisores tienen otros intereses que pudieran afectar la realización o la interpretación de su revisión. Por ejemplo, personas asociadas con los fabricantes de aceite de primula revisaron los efectos del fármaco sobre el eczema (Morse et al. 1989). Llegaron a una conclusión mucho más entusiasta acerca del valor del fármaco que una revisión a cargo de investigadores sin interés comercial, quienes incluyeron los resultados de estudios no publicados en su evaluación (Williams 2003).

No sólo los intereses comerciales pueden derivar en una selección prejuiciosa de la evidencia disponible para su inclusión en las revisiones. Todos tenemos prejuicios que pueden derivar en una selección poco imparcial de la evidencia, y los investigadores, los profesionales de la salud, los pacientes y otras personas que evalúen los efectos del tratamiento no son la excepción. Se ha reconocido que los conflictos de interés son importantes y se están tomando algunas medidas para tenerlos en cuenta.

Cite as: Editorial commentary (2007). Cómo evitar los sesgos en la selección de la evidencia disponible. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [reporting bias](#)

Other material: view [Monty Python's take](#) on reporting bias - YouTube video

Next essay: [Cómo reducir la obra de la casualidad a través del metanálisis](#)

Select essay:

Referencias

Berlin JA (1997). Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *Lancet* 350: 185-186.

Morse PF, Horrobin DF, Manku MS, Stewart JC, Allen R, Littlewood S, Wright S, Burton J, Gould DJ, Holt PJ, et al (1989). Meta-analysis of placebo-controlled studies of the efficacy of Epogam in the treatment of atopic eczema. Relationship between plasma essential fatty acid changes and clinical response. *British Journal of Dermatology* 121: 75-90.

Sackett DL, Oxman AD (2003). HARLOT plc: an amalgamation of the world's two oldest professions *BMJ* 327:1442-1445.

Williams HC (2003). Evening primrose oil for atopic dermatitis. BMJ 327: 1358-1359.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

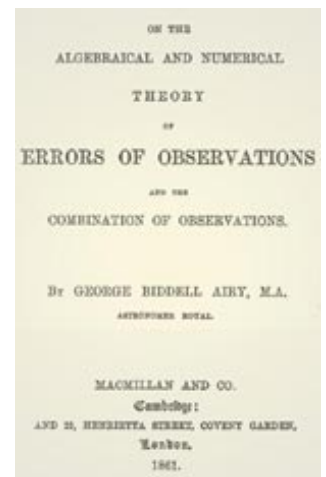
Revisiones sistemáticas de toda la evidencia pertinente:

Cómo reducir la obra de la casualidad a través del metanálisis

Para realizar pruebas auténticas de los tratamientos de atención médica se necesitan [revisiones sistemáticas](#) de toda la evidencia pertinente y confiable. A fin de no llegar a conclusiones erróneas acerca de los efectos de los tratamientos, quienes preparan dichas revisiones deben tomar medidas para evitar diversas clases de [sesgo](#), por ejemplo, deben tener en cuenta [toda la evidencia pertinente](#) y [evitar la selección prejuiciosa de la evidencia disponible](#).

Aunque se puede tener cuidado de minimizar el sesgo en las revisiones, es posible que se llegue a conclusiones engañosas acerca de los efectos de los tratamientos como resultado de [la obra de la casualidad](#). El análisis de estudios independientes pero similares, de a uno por vez, en revisiones sistemáticas también puede dejar una impresión confusa debido a la obra de la casualidad. Si es posible y corresponde, este problema puede reducirse mediante la combinación de datos de todos los estudios pertinentes y la utilización de un procedimiento estadístico que en la actualidad se conoce como "metanálisis".

La mayoría de las técnicas estadísticas utilizadas en la actualidad en el metanálisis derivan del trabajo que realizaron el matemático alemán Karl Gauss y el matemático francés Pierre-Simon Laplace durante la primera mitad del siglo XIX. La astronomía fue uno de los campos en los que sus métodos encontraron aplicaciones prácticas: la medición de la posición de las estrellas en una serie de ocasiones a menudo dio por resultado cálculos ligeramente diferentes, por lo tanto, se necesitaban técnicas para combinar los cálculos y producir un promedio derivado de los resultados recabados. En 1861, el Astrónomo Real Británico, George Airy, publicó un 'manual' para astrónomos ([Airy 1861](#)) en el que describió los métodos utilizados para este proceso de síntesis cuantitativa. Poco más de un siglo después, un científico social estadounidense, Gene Glass, denominó el proceso "metanálisis" (Glass 1976).



Karl Pearson publicó uno de los primeros ejemplos médicos de metanálisis en el British Medical Journal en 1904 ([Pearson 1904](#); O'Rourke 2006). El gobierno le había pedido que revisara la evidencia sobre los efectos de una vacuna contra la fiebre tifoidea. Si bien durante los siguientes 70 años los estadísticos desarrollaron métodos para el metanálisis, fue recién en la década del 70 que dichos métodos comenzaron a aplicarse con mayor frecuencia, inicialmente los aplicaron los científicos sociales (Glass 1976) y luego los investigadores médicos (Stjernsward J 1974; Stjernsward et al. 1976; Cochran et al. 1977; Chalmers et al. 1977; Chalmers 1979; Editorial 1980).

El metanálisis se puede ilustrar mediante el logotipo de [The Cochrane Collaboration](#). Este logotipo ilustra un metanálisis de datos de siete pruebas auténticas: cada línea horizontal representa los resultados de una prueba (cuanto más corta la línea, mayor la certeza del resultado) y cada diamante representa sus resultados combinados. La línea vertical indica la posición alrededor de la cual se agruparán las líneas horizontales si los dos tratamientos comparados en los estudios han tenido efectos similares; si una línea horizontal cruza la línea vertical, significa que esa prueba específica no encontró una diferencia clara ("estadísticamente significativa") entre los tratamientos. Cuando las líneas horizontales individuales cruzan la línea vertical de "no diferencia", se infiere que el tratamiento puede aumentar o disminuir las muertes de bebés. Sin embargo, en conjunto, las líneas horizontales suelen ubicarse en el lado beneficioso (izquierdo) de la línea de "no diferencia". El diamante representa los resultados combinados de estas pruebas, generados mediante el proceso estadístico de metanálisis. La posición del diamante claramente hacia la izquierda de la línea de "no diferencia" indica que el tratamiento es beneficioso.



Este diagrama muestra los resultados de una revisión sistemática de pruebas auténticas de un tratamiento corto y económico con un esteroide administrado a mujeres que se esperaba que tuvieran partos prematuros. La primera de estas pruebas se informó en 1972. El diagrama sintetiza la evidencia que habría sido revelada si se hubieran revisado sistemáticamente las pruebas disponibles una década después, en 1981: indica claramente que los esteroides reducen el riesgo de que los bebés mueran debido a las complicaciones generadas por la falta de

madurez. Para 1991, se habían comunicado siete estudios más y la imagen del logotipo se había vuelto más fuerte.

Hasta 1989 no se publicó ninguna revisión sistemática de estos estudios (Crowley 1989), en consecuencia, la mayoría de los obstetras, las parteras y las embarazadas no se dio cuenta de que el tratamiento era tan eficaz. En última instancia, algunas de las pruebas no habían arrojado beneficios "estadísticamente significativos", y quizá estas fueran las únicas pruebas que se habían tenido en cuenta. Dado que no se realizaron revisiones sistemáticas, decenas de miles de bebés prematuros sufrieron y murieron innecesariamente, y se desperdiciaron recursos en investigaciones innecesarias. Este es sólo uno de los muchos ejemplos de los costos humanos que pueden derivarse de no evaluar los efectos de los tratamientos en [revisiones sistemáticas y actualizadas de pruebas auténticas](#), utilizando metanálisis para reducir la probabilidad de que [la obra de la casualidad](#) nos conduzca al error.

A fines del siglo XX, se había aceptado en forma generalizada que el metanálisis era un elemento importante de las pruebas auténticas de los tratamientos y que ayudaba a evitar las conclusiones incorrectas de que los tratamientos no tenían efectos cuando, de hecho, eran beneficiosos o dañinos.

Cite as: Editorial commentary (2007). Cómo reducir la obra de la casualidad a través del metanálisis. The James Lind Library (www.jameslindlibrary.org).

Show JLL records: illustrating [meta-analysis](#)

Next essay: [Revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable](#)

Select essay:

Referencias

Airy GB (1861). On the algebraical and numerical theory of errors of observations and the combination of observations. London: Macmillan.

Chalmers I (1979). Randomized controlled trials of fetal monitoring 1973-1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. Perinatal Medicine. Stuttgart: Georg Thieme, 260-265.

Chalmers TC, Matta RJ, Smith H, Kunzler A-M. (1977). Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. New England Journal of Medicine 297:1091-1096.

Crowley P (1989). Promoting pulmonary maturity. In: Chalmers I, Enkin M, Keirse MJNC, eds. Effective care in pregnancy and childbirth. Oxford: Oxford University Press, pp 746-762.

Editorial (1980). Aspirin after myocardial infarction. Lancet 1:1172-3.

Glass GV (1976). Primary, secondary and meta-analysis of research. Educational Researcher 10, 3-8.

O'Rourke K (2006). An historical perspective on meta-analysis: dealing quantitatively with varying study results. The James Lind Library.

Pearson K (1904). Report on certain enteric fever inoculation statistics. BMJ 3:1243-1246.

Stjernswärd J (1974). Decreased survival related to irradiation postoperatively in early operable breast cancer. Lancet 2:1285-1286.

Stjernswärd J, Muenz LR, von Essen CF (1976). Postoperative radiotherapy and breast cancer. Lancet 1:749.

[Home](#)

[Contenido](#)

[Comments welcome](#)

[Home](#)[Contenido](#)jameslindlibrary.org

Revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable

Pruebas auténticas de los tratamientos en la atención médica

Los resultados de las pruebas auténticas individuales de los tratamientos médicos rara vez se insertan sistemáticamente en el contexto de otros estudios similares, utilizando métodos para reducir los [sesgos](#) y [la obra de la casualidad](#). Esta falta de revisiones sistemáticas de las investigaciones sobre los efectos de los tratamientos ha provocado mucho sufrimiento que podría haberse evitado. Las [pruebas auténticas de los tratamientos en la atención médica](#) también implican la preparación de revisiones sistemáticas de todos los estudios pertinentes y confiables de investigación de los tratamientos que se evalúan.

Hay algunos ejemplos de este proceso que se remontan a más de 200 años atrás. En 1753, por ejemplo, en su revisión del gran número de informes acerca de la prevención y el tratamiento del escorbuto, James Lind observó:



“No es sencillo desarraigar los prejuicios,... resultó esencial mostrar una visión completa e imparcial de lo que hasta ese momento se había publicado sobre el escorbuto... De hecho, antes de que el tema se pudiera explicitar clara y correctamente, fue necesario eliminar una gran cantidad de basura”. ([Lind 1753](#))

Las revisiones sistemáticas de todas las investigaciones pertinentes que abordan las preguntas acerca de los efectos de los tratamientos se consideran cada vez más como los fundamentos más confiables para las conclusiones acerca de dicho efectos. A veces, las revisiones sistemáticas mostrarán que no existe evidencia confiable, de hecho, ésta es una de sus funciones más importantes. Asimismo, las revisiones sistemáticas a veces podrán confirmar que la evidencia confiable se limita a un estudio único y aquí, también, es importante que esta situación sea explícita.

La comprensión de que es necesario realizar revisiones sistemáticas para obtener pruebas auténticas de tratamientos se ha reflejado en el rápido aumento de los números de informes de revisiones sistemáticas que se publican en papel y en Internet ([DARE](#); [The Cochrane Collaboration](#)). Se utilizan para (i) informar a la práctica clínica, por ejemplo, a través de la publicación [Clinical Evidence](#) del British Medical Journal y de la red [Scottish Intercollegiate Guidelines Network](#); (ii) evaluar qué tratamientos médicos son eficaces en función de los costos, por ejemplo, las evaluaciones del [National Institute for Health and Clinical Excellence](#); y (iii) satisfacer las necesidades que tienen los pacientes de contar con información confiable acerca de los efectos de los tratamientos, por ejemplo, a través de [Informed Health Online](#) y de la [National Library for Health](#).

Asuntos pendientes

Estos desarrollos y otros similares muestran que quienes intentan mejorar el acceso a la evidencia necesaria para informar sobre las elecciones en la atención médica han aceptado la importancia de las revisiones sistemáticas. Sin embargo, aún resta mucho por hacer: se calcula que la producción actual de la Cochrane Collaboration de varios miles de revisiones sistemáticas deberá aumentarse a mucho más de 10.000 para abarcar la evidencia existente (Mallet y Clarke 2002) y luego mantenerse actualizada a medida que vaya surgiendo nueva evidencia. De hecho, un editor de publicaciones ha sugerido que debería haber una moratoria sobre toda nueva investigación hasta que nos hayamos puesto al día con la información que nos puede ofrecer la evidencia existente (Bausell 1993).

Los responsables de desembolsar fondos para la investigación deben asegurarse de que se proporcionen recursos para superar este atraso, de que sólo se apoyen nuevos estudios si las revisiones sistemáticas de la evidencia existente han demostrado que se necesitan estudios adicionales y de que dichos estudios se hayan diseñado para aprovechar las lecciones de las investigaciones previas. Si los editores de publicaciones desean ofrecer un mejor servicio a sus lectores, deberán seguir el ejemplo de The Lancet y asegurarse de que las comunicaciones de los nuevos estudios aclaren qué contribución ha hecho la nueva evidencia a la revisión sistemática actualizada de toda la evidencia pertinente (Young y Horton 2005).

El aumento de la disponibilidad de revisiones actualizadas y sistemáticas está mejorando la calidad de la

información acerca de los efectos de los tratamientos, pero no se deben aceptar las conclusiones de las revisiones sistemáticas sin crítica. Revisiones diferentes que aparentemente abordan el mismo interrogante acerca de los tratamientos a veces llegan a conclusiones divergentes. Sus autores son humanos y debemos ser conscientes de que pueden seleccionar, analizar y presentar evidencia en muchas formas que respalden sus prejuicios e intereses. La evolución continua de métodos confiables para preparar y mantener las revisiones sistemáticas contribuirá a reducir este problema, pero no se puede esperar que lo elimine.

Aunque el crecimiento en el número de revisiones sistemáticas ha aumentado la disponibilidad de las pruebas auténticas primarias de los tratamientos en la atención de la salud, estas revisiones suelen revelar la escasa calidad y la falta de pertinencia de gran parte de la investigación de los efectos de los tratamientos. Como lo expresó un editorialista en un comentario acerca del "escándalo de la falta de calidad en la investigación médica": necesitamos menos investigaciones, mejores investigaciones e investigaciones que se emprendan por las razones correctas (Altman 1994). Parece poco probable que esto se logre sin una mayor comprensión pública de cuáles son las razones para realizar pruebas auténticas de los tratamientos y de sus características, y sin una mayor influencia y compromiso públicos en todas las fases de las pruebas auténticas de los tratamientos. La imposición y difusión de estas prioridades dependen de que nuevas alianzas entre pacientes y médicos hagan frente a la incertidumbre acerca de los efectos de los tratamientos (Chalmers 2004; www.duets.nhs.uk; [James Lind Alliance](http://JamesLindAlliance.org)).

El público y los profesionales de la salud recibirán lo que merecen cuando puedan acceder más fácilmente a revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable que aborde las dudas importantes sobre los efectos de los tratamientos y a información acerca de las investigaciones en curso que aborden dichas incertidumbres (Smith and Chalmers 2001).

Cite as: Editorial commentary (2007). Revisiones sistemáticas actualizadas de toda la evidencia pertinente y confiable. The James Lind Library (www.jameslindlibrary.org).

Select essay:

Referencias

Altman (1994). The scandal of poor medical research. *BMJ* 308:283-284.

Bausell BB (1993). After the meta-analytic revolution. *Evaluation and the Health Professions* 16:3-12.

Bunker JP, Frazier HS, Mosteller F (1994). Improving health: measuring effects of medical care. *Milbank Quarterly* 72:225-258.

Chalmers I (2004). Well informed uncertainties about the effects of treatments: how should clinicians and patients respond? *BMJ* 328:475-476.

Lind J (1753). A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson.

Mallett S, Clarke M (2002). The typical Cochrane Review. *International Journal of Technology Assessment in Health Care* 18:820-823.

Smith R, Chalmers I (2001). Britain's gift: a 'Medline' of synthesized evidence. *BMJ* 323:1437-1438.

Young C, Horton R (2005). Putting clinical trials into context. *Lancet* 366:107-8.

[Home](#)

[Contenido](#)

[Comments welcome](#)