

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.

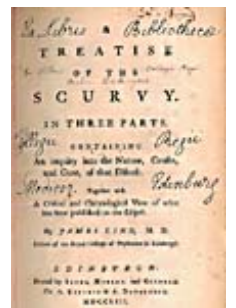
En el siglo XVIII, el cirujano William Cheselden ya era consciente del problema que representaban los "grupos disímiles" cuando los cirujanos comparaban sus respectivas tasas de mortalidad después de las operaciones para extraer cálculos de la vejiga. Cheselden señaló que era importante tener en cuenta 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](http://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](http://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](#)