

**teorema**

Vol. XLI/3, 2022, pp. 47-63

ISSN 0210-1602

[BIBLID 0210-1602 (2022) 41:3; pp. 47-63

## La validez interna de los RCT: el caso de los errores aleatorios en nutrición

Roberto López-Mas

### ABSTRACT

RCTs are considered the gold standard in various evidence hierarchies for assessing the quality of scientific research, as well as for guiding regulations and public policy. In this article, the challenges of this methodology in nutrition science are identified by means of analyzing the difficulties that the reduction of random error presents. The minimization of random error could be impossible in a context in which large-scale RCTs are not feasible and meta-analyses are not an option. The conclusion is that guaranteeing the reliability of the evidence in nutrition RCTs can be a highly complex task.

KEYWORDS: *RCT, Nutrition, Internal Validity, Causality, Random Error, Uncertainty.*

### RESUMEN

Los RCT se consideran el estándar de oro en diversas jerarquías evidenciales propuestas, tanto para valorar la calidad de la investigación científica como para orientar regulaciones y políticas públicas. En este artículo se identifican los retos de esta metodología en las ciencias de la nutrición, a partir del análisis de las dificultades que presenta la reducción de los errores aleatorios. En contextos en los que los RCT de grandes dimensiones no son realizables y los meta-análisis no son una opción, la minimización de los errores aleatorios puede ser imposible. Se concluye que garantizar la fiabilidad de la evidencia de los RCT en nutrición puede ser una tarea considerablemente compleja.

PALABRAS CLAVE: *RCT, nutrición, validez interna, causalidad, error aleatorio, incertidumbre.*

## I. Introducción

Los ensayos controlados aleatorizados (*randomized controlled trials*, RCT) son estudios en los que una muestra de la población es dividida en, al menos, dos grupos: el grupo experimental, cuyos miembros reciben la sustancia o el tratamiento a prueba, y el grupo de control, cuyos integrantes no reciben la sustancia o el tratamiento. La asignación de los individuos a cada uno de los grupos se lleva a cabo mediante un proceso aleatorio.

En las ciencias de la salud, los participantes del grupo de control pueden recibir un placebo, que es una sustancia que no tiene ningún efecto bioquímico relevante para la enfermedad investigada. Los RCT suelen ser a doble ciego, lo que significa que ni los participantes ni los investigadores saben en qué grupo se encuentra cada uno de los individuos de la muestra.

Los RCT se consideran la mejor metodología para la inferencia causal en diversos ámbitos científicos. Este tipo de estudio puede tener una validez interna superior con respecto al resto de métodos: puede generar datos más precisos y fiables para el establecimiento de conclusiones causales. No obstante, diversos filósofos de la ciencia [Cartwright (2007), Cartwright y Stegenga (2011), Clarke *et al.* (2013), Jukola (2019), Reiss (2015), Stegenga (2014), Vandenbroucke, Broadbent y Pearce (2016), Worrall (2007)] han puesto en cuestión que los RCT constituyan una metodología universalmente superior. En el contexto de este debate, el objetivo de este artículo es analizar la validez interna de los RCT en las ciencias de la nutrición sobre la base de la incertidumbre en los resultados introducida por los errores aleatorios.

El RCT internamente válido se define como el que cumple, al menos, tres requisitos: a) hay un grupo experimental y un grupo de control correctamente diseñados; b) los errores sistemáticos son eliminados; y c) los errores aleatorios son minimizados. Se argumenta que, en el ámbito de la nutrición, alcanzar cada uno de los tres requisitos plantea retos difíciles de superar. Los límites éticos y técnicos de esta metodología en nutrición han sido ampliamente reconocidos [Biesalski *et al.* (2011), Blumberg *et al.* (2010), Heaney (2006), (2008), Richardson (2012), Visioli (2012)]. Sin embargo, existe un aspecto relevante no suficientemente estudiado hasta el momento: si los meta-análisis o los RCT con muestras grandes no son realizables, la reducción de los errores aleatorios puede ser imposible. En tal caso, el grado de incertidumbre en los datos de los RCT en nutrición puede ser significativo, llegando a poner en cuestión la fiabilidad de los resultados.

En la segunda sección, procedo a presentar las jerarquías evidenciales, en las que los RCT ocupan una de las posiciones más destacadas. Expongo algunas de las propuestas más relevantes de filósofos de la ciencia que mantienen una postura crítica con tales jerarquías [Cartwright (2007), Cartwright y Stegenga (2011), Clarke *et al.* (2013), Osimani (2020), Reiss (2015), Russo y Williamson (2007), Stegenga (2014), Vandenbroucke, Broadbent y Pearce (2016)]. En la tercera sección, planteo una caracterización de los RCT internamente válidos. Además, identifico algunas de las limitaciones de esta metodología reconocidas en el ámbito

de la nutrición [Biesalski *et al.* (2011), Blumberg *et al.* (2010), Heaney (2006), (2008), Richardson (2012), Visioli (2012)]. En la cuarta sección, desarrollo mi tesis principal: la minimización de los errores aleatorios en los RCT puede ser imposible en determinados casos. Argumento que, en las ciencias de la nutrición, factores prácticos y factores metodológicos pueden llevar a una ausencia de vías epistémicas para reducir los errores aleatorios en los RCT.

Como conclusión se establece que, aunque los RCT pueden generar resultados con una incertidumbre mínima, en la práctica científica los estudios pueden estar lejos de tal escenario. En las ciencias de la nutrición, las explicaciones alternativas a la hipótesis causal pueden tener un nivel de plausibilidad importante. Los errores aleatorios podrían no haber sido minimizados debido a los retos de los RCT en diferentes contextos científicos. Cuando la hipótesis de la acción del azar no puede ser rechazada, la incertidumbre en los resultados del RCT se puede considerar demasiado elevada para aceptar la conclusión causal.

## II. ¿RCT COMO ESTÁNDAR DE ORO?

Los RCT se consideran el estándar de oro en diversas disciplinas científicas. Esta metodología se halla en las posiciones más destacadas de las jerarquías evidenciales, que clasifican los distintos tipos de estudios en función de su capacidad para generar datos relevantes para el establecimiento de relaciones causales. Las jerarquías constituyen un criterio en la evaluación de la fiabilidad de la evidencia científica para la toma de decisiones en la ciencia reguladora, la medicina y las ciencias de la nutrición [Jukola (2019), La Caze (2009), Luján y Todt (2021)]. En este contexto, los estudios de cohorte sin aleatorización se suelen situar por debajo de los RCT, mientras que los estudios de laboratorio y los estudios de casos ocupan las posiciones inferiores [Stegenga (2014)]. En las ciencias de la salud, las revisiones sistemáticas y los metaanálisis de RCT se hallan en lo más alto de la clasificación metodológica [La Caze (2009), Osimani (2020)]. El concepto fundamental sobre el que se construyen las jerarquías evidenciales es el de *validez interna* [Osimani y Mignini (2015)]. Se considera que las metodologías en las posiciones más destacadas presentan una validez interna superior: tienen un diseño que posibilita la neutralización de los factores de confusión, generando así evidencia más precisa [Andreoletti y Teira (2019), Jukola (2019), La Caze (2009), Luján y Todt (2021), Osimani (2020)].

Una tesis ampliamente aceptada en diversos ámbitos científicos es que las metodologías situadas en las posiciones inferiores de las jerarquías evidenciales tienen una menor capacidad para establecer causalidad. Desde este enfoque, los estudios observacionales, por ejemplo, podrían mostrar una correlación entre dos factores, pero no podrían concluir que tal relación es causal. Las metodologías con un nivel bajo de validez interna no pueden neutralizar todos los factores de confusión, ni pueden eliminar otros sesgos que podrían explicar los resultados del estudio [EFSA (2010), Heaney (2008), Martínez y Siani (2017), Neale y Tapsell (2019), Richardson (2012), Spieth *et al.* (2016), Stephenson y Babiker (2000), Todt y Luján (2017a), (2017b)]. Los RCT sí podrían llegar a minimizar los sesgos y los intereses en la toma de decisiones. Esta metodología se considera imparcial ante elementos sociales o económicos. Según Mattia Andreoletti y David Teira [(2019)], en el marco de la regulación de fármacos, una estrategia de reglas centrada en la exigencia de evidencia procedente de RCT presenta la ventaja de ser resistente a la presión y los intereses de los agentes externos, en comparación a estrategias alternativas que se basarían en el juicio experto.

Algunos filósofos de la ciencia han puesto en cuestión los RCT como estándar de oro, así como las jerarquías evidenciales en general [Cartwright (2007), Clarke *et al.* (2013), Jukola (2019), Reiss (2015), Stegenga (2014), Vandenbroucke, Broadbent y Pearce (2016), Worrall (2007)]. Uno de los argumentos principales que se han planteado desde esta postura es el del problema de la *validez externa*. Un estudio carece de validez externa cuando sus resultados no son extrapolables a la población objetivo. Aunque un RCT genere datos precisos y fiables sobre la eficacia de un determinado tratamiento, no proporciona la evidencia suficiente para extender la conclusión causal a poblaciones diferentes a la del estudio [Cartwright (2007), Clarke *et al.* (2013), Jukola (2019), Osimani (2020)]. La aplicación de los resultados de un RCT internamente válido en otros contextos distintos al experimental no estaría garantizada. Esta metodología no explica qué interacciones de factores se dan cuando dos individuos muestran efectos diferentes ante un mismo tratamiento [Osimani *et al.* (2019)]. Obtener información sobre las distintas combinaciones relevantes de cofactores podría requerir el uso de métodos situados en las posiciones inferiores de las jerarquías evidenciales, tales como los estudios de casos o los estudios *in vitro* [Cartwright (2007), Cartwright y Stegenga (2011), Osimani (2020), Stegenga (2014)]. Si bien los datos de estos estudios pueden considerarse menos precisos, podrían generar un conocimiento sobre mecanismos que sería fundamental para la evalua-

ción de la validez externa. De este modo, se podrían investigar los diferentes contextos en los que puede actuar un determinado tratamiento y hasta qué punto es eficaz en cada uno de los casos [Cartwright (2007), Osimani (2020), Stegenga (2014)].

Como alternativa a las jerarquías evidenciales que priman los RCT, se ha propuesto un pluralismo que defiende que la inferencia causal implica diversos tipos de evidencia científica y, por lo tanto, el uso de metodologías distintas [Cartwright (2007), Cartwright y Stegenga (2011), Clarke *et al.* (2013), Osimani (2020), Reiss (2015), Russo y Williamson (2007), Stegenga (2014), Vandenbroucke, Broadbent y Pearce (2016)]. Este enfoque sostiene que los estudios con un grado relativamente bajo de validez interna pueden llevar a conclusiones causales dentro de una estructura compleja que incorpore los datos procedentes de diferentes métodos, o incluso diversas disciplinas. Aunque no se pueda probar una relación causal a partir de un estudio *in vitro* particular o un único estudio de cohorte, se podría llegar a tal conclusión tomando en consideración distintas metodologías científicas. Algunos pluralistas sostienen que el método adecuado para establecer causalidad depende del tipo de relación estudiada [Vandenbroucke, Broadbent y Pearce (2016)]. Otros filósofos críticos con las jerarquías evidenciales no ponen el foco sobre las metodologías científicas, sino sobre el tipo de evidencia. Federica Russo y Jon Williamson (2007), proponen un pluralismo evidencial según el cual tanto la evidencia probabilística como la evidencia mecanística son necesarias para la inferencia causal. En esta misma línea, Brendan Clarke *et al.* (2013), defienden que la evidencia de los estudios mecanísticos no es inferior a la evidencia estadística de RCT, sino que es complementaria. Los pluralistas no suelen negar que los RCT sean útiles bajo determinadas circunstancias. Sin embargo, esta metodología no se consideraría un estándar de oro, pues ningún tipo de estudio es universalmente superior [Cartwright (2007), Reiss (2015), Stegenga (2014)].

### III. VALIDEZ INTERNA DE LOS RCT

Mientras que los pluralistas apelan al problema de la validez externa en el análisis metodológico, los defensores de las jerarquías evidenciales ponen el foco sobre la cuestión de la validez interna. El RCT internamente válido elimina la posibilidad de que los resultados del experimento se puedan explicar mediante determinadas hipótesis alternativas al efecto causal del tratamiento, tales como los *errores sistemáticos*. Este tipo de error

produce una desviación sistemática de los resultados con respecto a los valores verdaderos. Los errores sistemáticos – tales como los sesgos de selección, sesgos de confirmación o sesgos de desempeño – suelen ser consecuencia de características de los participantes o los investigadores del estudio. Por ejemplo, se pueden producir sesgos de confirmación cuando los pacientes de un ensayo clínico son conscientes de que están recibiendo un tratamiento particular. En este caso, aunque el tratamiento no tenga realmente ningún efecto clínico, los pacientes pueden afirmar sentirse mejor al final del ensayo. Los resultados del estudio podrían estar influidos por las expectativas sobre los beneficios del tratamiento. Este error se considera sistemático porque se repite necesariamente en RCT con un diseño y desarrollo idénticos.

La ocultación de la asignación aleatoria y el doble ciego pueden ayudar a eliminar diferentes errores sistemáticos que podrían generarse a lo largo del estudio [Akobeng (2008), Althubaiti (2016), Karanicolas, Farrukhyar y Bhandari (2010), Renjith (2017), Senn (2007), Shaver (1993), Spieth *et al.* (2016), Stephenson y Babiker (2000), Viera y Bangdiwala (2007)]. En los RCT sin ciego, los efectos de los tratamientos se suelen sobreestimar considerablemente, en comparación a ensayos a doble ciego. En consecuencia, el desarrollo de un placebo que sea idéntico a la sustancia a estudio en cuanto al aspecto, sabor y olor puede ser importante para garantizar la fiabilidad de los resultados del experimento. El doble ciego es crucial –aunque puede no ser suficiente– para erradicar la posibilidad de distintos sesgos humanos y, en términos generales, para asegurar la validez interna de los RCT.

Si bien los errores sistemáticos pueden ser eliminados, el caso de los *errores aleatorios* es distinto. Este tipo de error es el que ocurre por acción del azar, generando una desviación de los valores verdaderos de manera no consistente y que no se puede predecir [Akobeng (2008), Althubaiti (2016), Osimani (2020), Stephenson y Babiker (2000)]. Los errores aleatorios no se repetirían necesariamente en estudios idénticos. Los resultados de un ensayo podrían deberse no a que exista una relación causal entre el tratamiento a prueba y la recuperación, sino a un error aleatorio por la variabilidad de los datos medidos, que generaría una medición imprecisa del efecto del tratamiento. Los errores aleatorios están relacionados con el tamaño de la muestra: en la medida en que la muestra crece, los errores aleatorios disminuyen. En estudios con una muestra suficientemente grande, los errores aleatorios pueden ser reducidos. A diferencia de los errores sistemáticos, los errores aleatorios nunca se pueden evitar completamente a través del diseño y el control adecuado del RCT.

Los errores aleatorios no se pueden eliminar, pero sí se pueden minimizar. Existen, en principio, tres vías para conseguir la reducción de este tipo de error: a) mediante la réplica idéntica del estudio original; b) con ensayos cuyas muestras sean tan grandes como sea posible; y c) a través de los meta-análisis, que combinan los resultados de varios estudios suficientemente homogéneos [Akobeng (2008), Osimani (2020), Osimani y Mignini (2015), Spieth *et al.* (2016), Stanley y Spence (2014), Stephenson y Babiker (2000)]. Réplicas idénticas del estudio original que utilicen la misma muestra permitirían, a la larga, determinar la diferencia media entre los dos grupos del RCT, minimizando los errores aleatorios. La replicación podría considerarse necesaria para estimar el verdadero tamaño del efecto investigado. Aunque los errores aleatorios son inevitables –poniendo en peligro las inferencias causales si no son reducidos–, se pueden llegar a minimizar.

El RCT internamente válido tiene que cumplir, por lo menos, tres requisitos: a) hay un grupo experimental y un grupo de control correctamente diseñados; b) los errores sistemáticos son eliminados; y c) los errores aleatorios son minimizados. Si se cumple cada uno de estos puntos y se obtienen unos resultados positivos, se puede inferir una conclusión causal a partir de evidencia precisa y fiable, con un nivel bajo de incertidumbre. En los RCT con una validez interna elevada, los resultados solo se pueden explicar por el efecto causal del tratamiento o por unos errores aleatorios reducidos. En este caso,  $p(H/E) \approx 1$ , donde la hipótesis  $H$  es *el tratamiento causa la recuperación* y la evidencia  $E$  es el resultado del ensayo, que muestra una mayor recuperación en el grupo experimental. No obstante, alcanzar los requisitos del RCT internamente válido puede presentar considerables dificultades en varias disciplinas científicas.

Los RCT surgieron en el campo de la medicina como método estándar para la evaluación de fármacos, y, posteriormente, fueron aplicados a las ciencias de la nutrición [Blumberg *et al.* (2010), Todt y Luján (2017b)]. El uso de estos estudios en nutrición se basa en la idea de que pueden proporcionar una fundamentación científica sólida a las recomendaciones relacionadas con la nutrición y, concretamente, las declaraciones de salud. En la mayoría de las regulaciones de las declaraciones de salud, los RCT se consideran la metodología superior por su capacidad para establecer relaciones causales sólidas entre el consumo de un alimento y beneficios para la salud [Blumberg *et al.* (2010), EFSA (2010), Lichtenstein *et al.* (2021), Martínez y Siani (2017), Richardson (2012), Todt y Luján (2017a), (2017b)]. Sin embargo, alcanzar un nivel alto de validez interna en los RCT puede ser un reto complejo por diversos factores.

En diversas disciplinas científicas, los RCT pueden presentar dificultades en el correcto diseño de un grupo experimental y un grupo de control. En las ciencias de la salud, por ejemplo, proporcionar un placebo a pacientes enfermos puede no ser ético en determinados contextos. En el ámbito de la nutrición, destacan los límites de los estudios utilizados para investigar los efectos del consumo de nutrientes esenciales. Crear un grupo de control al que se le proporcione un placebo en lugar de, por ejemplo, vitamina C no es aceptable. Administrar una dieta nutritivamente insuficiente no es ético porque no se pueden provocar deficiencias artificiales en los participantes del estudio [Biesalski *et al.* (2011), Blumberg *et al.* (2010), Heaney (2006), (2008), Richardson (2012), Visioli (2012)]. En consecuencia, disponer de un control apropiado puede resultar imposible en algunos contextos. De este modo, alcanzar el primer requisito de los RCT internamente válidos puede ser imposible en algunos casos, debido a los límites éticos de la experimentación en humanos.

La eliminación de los distintos errores sistemáticos también puede ser un reto complejo en los RCT en nutrición. Esto se debería a las limitaciones técnicas que pueden existir en relación a la creación de un placebo adecuado. Cuando se investigan los efectos de un complemento vitamínico, la producción de un placebo en forma de cápsula puede ser relativamente sencillo, como en ciertos casos de ensayos con fármacos. No obstante, si el objetivo es estudiar los efectos de alimentos o dietas completas, la técnica necesaria para el desarrollo de un placebo adecuado puede no estar disponible. Por ejemplo, en un RCT que estudia la relación entre una dieta baja pero suficiente en sodio y una mejora de la salud, los resultados pueden estar sesgados si los participantes descubren por el sabor que su dieta tiene un bajo contenido en sal. Estos participantes podrían buscar información sobre los efectos para la salud de la dieta a la que se exponen, lo que podría generar errores sistemáticos. La validez interna de este ensayo puede estar en peligro sin la técnica necesaria para producir un placebo suficientemente similar a la sal común en todos los aspectos relevantes, tales como el sabor, la textura o incluso el aspecto. La eliminación de los sesgos de confirmación o los sesgos de deserción puede ser complicada cuando el doble ciego no está garantizado.

#### IV. EL RETO DE LOS ERRORES ALEATORIOS

Diseñar un control adecuado o eliminar los diferentes errores sistemáticos en los RCT puede llegar a ser una tarea compleja. Diversos au-



tores han señalado que, en las ciencias de la nutrición, cumplir los dos primeros requisitos de los RCT internamente válidos puede ser un reto difícil en determinados contextos [Biesalski *et al.* (2011), Blumberg *et al.* (2010), Heaney (2006), (2008), Richardson (2012), Visioli (2012)]. Sin embargo, estas no son las únicas limitaciones que puede presentar el diseño y desarrollo de RCT internamente válidos en nutrición. En este apartado, argumento que también el tercer requisito de estos estudios –la minimización de los errores aleatorios– puede ser difícil o incluso imposible de alcanzar en algunos casos. La reducción de los errores aleatorios puede constituir una tarea importante en nutrición y, al mismo tiempo, considerablemente complicada debido tanto a factores metodológicos como a factores prácticos.

Cuando los errores aleatorios en un RCT son significativos, no se puede considerar implausible que los resultados sean producto del azar. Esta es una cuestión relevante para la investigación en nutrición –aunque no es exclusiva a este campo–, ya que los efectos de los nutrientes suelen ser sutiles o incluso difíciles de medir. Si bien el consumo de varios alimentos puede generar una mejora importante para la salud humana, el efecto de los ingredientes individuales suele ser pequeño [de Boer, Vos y Bast (2014), Heaney (2008), Todt y Luján (2017b)]. En un RCT en el que se somete a prueba la hipótesis de que un ingrediente particular causa una mejora concreta en la salud, los errores aleatorios pueden ser considerables. Esto se debería a que existe una relación entre la magnitud del efecto medido y los errores aleatorios. Cuando el efecto verdadero del estudio es pequeño, se suele dar una mayor variabilidad en los valores medidos [Stanley y Spence (2014)]. Por lo tanto, la minimización de los errores aleatorios es esencial para incrementar la fiabilidad de los datos generados mediante RCT.

Existen, en principio, tres vías para reducir los errores aleatorios en los RCT: la réplica idéntica del estudio original, los ensayos con muestras suficientemente grandes, y los metaanálisis. Si bien la réplica exacta de los RCT no es posible en sentido estricto [Osimani y Mignini (2015), Worrall, (2007)], las otras dos vías podrían ser adecuadas para la minimización de los errores aleatorios [Akobeng (2008), Blair (2004), Viera y Bangdiwala (2007)]. No obstante, al menos en algunos contextos, ninguna de estas dos vías constituye una opción en las ciencias de la nutrición. En esta línea, defiendo que: a) los estudios con muestras grandes pueden ser complicados debido a limitaciones prácticas; y b) el uso de metaanálisis puede estar desincentivado o incluso descartado en la jerarquía evidencial estándar.

Los RCT con muestras grandes se consideran una vía adecuada para la minimización de los errores aleatorios. No obstante, llevar a cabo estudios con un número elevado de participantes constituye un reto difícil de superar en el ámbito de la nutrición. Varios autores han especificado las limitaciones prácticas de los RCT con muestras grandes: a) son necesarios unos recursos económicos elevados; b) la duración de los ensayos debería ser en algunos casos de décadas para poder medir efectos con largos periodos de latencia; y c) contar con muchos participantes que finalicen el ensayo es complicado a causa de la falta de adherencia a las dietas y la alta tasa de abandono [Aggett (2012), Aggett *et al.* (2005), Biesalski *et al.* (2011), Blumberg *et al.* (2010), Jacobs, Tapsell y Temple (2011), Jukola (2019), Richardson (2012)]. Kathryn Foti *et al.* [(2020)] evalúan el diseño y desarrollo que debería tener un RCT a gran escala para investigar los efectos cardiovasculares de una reducción de sodio en la dieta. Las estimaciones del tamaño y la duración del estudio son extrapolaciones a partir del diseño de ensayos anteriores sobre la prevención de enfermedades cardiovasculares graves. Foti *et al.* concluyen que, si se persigue un RCT internamente válido con un error aleatorio reducido, la muestra debería incluir 16.996 participantes procedentes de una población en riesgo elevado o 37.124 participantes de una población general, con un coste mínimo de 407.904.000 dólares y una duración de 5 años. Sostienen que, aunque los costes económicos fueran reducidos a la mitad, seguirían siendo difícilmente asumibles.

Cuando los RCT a gran escala plantean dificultades prácticas considerables, los meta-análisis pueden llegar a constituir un recurso útil para minimizar los errores aleatorios. Los meta-análisis combinan los resultados de diferentes estudios individuales. Aunque los distintos RCT tengan muestras muy pequeñas, es posible aumentar el número total de participantes a partir de meta-análisis. De este modo, se pueden llegar a reducir los errores aleatorios al incrementar suficientemente el tamaño de la muestra. Los estudios incluidos en un meta-análisis deben ser tan homogéneos como sea posible —en los criterios de inclusión y exclusión, el control utilizado, la población del estudio y el contexto— para que los resultados sean significativos [Barnard, Willett y Ding (2017), La Caze (2009), Osimani (2020), Osimani y Mignini (2015)]. Esta homogeneidad necesaria puede ser difícil de alcanzar en las ciencias de la nutrición, ya que los RCT pueden presentar diferencias metodológicas importantes. Sin embargo, cuando son posibles y adecuados, los meta-análisis pueden proporcionar resultados con un grado alto de fiabilidad.

En los casos en que la realización de meta-análisis es una opción – a partir de un conjunto de estudios suficientemente homogéneos–, es necesario analizar hasta qué punto sus resultados finales se consideran relevantes. En las ciencias de la salud, los métodos sistemáticos de síntesis de datos se hallan en lo más alto de la jerarquía evidencial, por encima de los RCT individuales. En el ámbito de la nutrición, en cambio, estos métodos ocupan las posiciones inferiores de determinadas jerarquías. Según la EFSA [(2021)], los meta-análisis no proporcionan una evidencia científica sólida para la confirmación de hipótesis relativas a los efectos causales de los alimentos funcionales. La jerarquía evidencial definida por la agencia reguladora otorga un valor mínimo a los trabajos de revisión de estudios:

- *Estudios de intervención en humanos*
    - Ensayos controlados aleatorizados
    - Ensayos aleatorizados sin control
    - Ensayos controlados sin aleatorización
    - Otros estudios de intervención
  - *Estudios observacionales en humanos*
    - Estudios de cohorte
    - Estudios de casos y controles
    - Estudios transversales
    - Otros estudios observacionales
  - *Trabajos de revisión de estudios en humanos*
    - Revisiones sistemáticas
    - Análisis combinados
    - Meta-análisis
    - Otros trabajos de revisión
- [EFSA (2021)]

En este marco, los meta-análisis no se consideran decisivos para probar una conclusión causal. Aunque la combinación de los resultados de RCT individuales suficientemente homogéneos sea posible, esta tarea podría no estar justificada. El bajo valor de los meta-análisis en la jerarquía de la EFSA contrasta con la postura de autores del ámbito de la nutrición: “las revisiones sistemáticas y los meta-análisis representan el recurso más valioso, fiable y objetivo para resumir la evidencia relativa a una cuestión de estudio específica en la investigación en nutrición” [Schwingshackl *et al.* (2016) p. 994]. Cuando los meta-análisis son posibles y adecuados, po-

drían constituir una vía epistémica para minimizar los errores aleatorios en la investigación en nutrición, al incrementar el tamaño de la muestra. Sin embargo, los métodos sistemáticos de síntesis de datos no son una opción relevante en el estudio de los efectos de los nutrientes porque, a priori, no proporcionarían resultados suficientemente fiables, según la jerarquía evidencial de la EFSA [(2021)].

En una jerarquía evidencial que clasifica las distintas metodologías por su capacidad para generar datos precisos, los estudios científicos que ocupan las posiciones inferiores se consideran relativamente poco fiables. La situación de los meta-análisis en la jerarquía estándar lleva a la idea de que este método presenta un nivel bajo de fiabilidad, en comparación al de cualquier otra metodología científica. Esto es especialmente relevante porque las jerarquías no constituyen simples propuestas teóricas. Tienen consecuencias directas en las elecciones metodológicas dentro de la investigación científica. Jacob Stegenga [(2014)] sostiene que la aplicación de las jerarquías evidenciales conduce a ignorar los métodos situados en las posiciones inferiores. Esto se aplicaría tanto a ámbitos disciplinares como a ámbitos reguladores. En la regulación europea de los alimentos funcionales, el establecimiento de una jerarquía evidencial implica un monismo metodológico que prima los datos de RCT individuales para estudiar relaciones causales relativas a la nutrición [López-Mas (2021), Luján y Todt (2021), Sanz-Merino y Luján (2021)]. En la práctica científica, la realización de meta-análisis puede quedar desincentivada o incluso descartada por el bajo valor otorgado en contextos como el de la regulación europea.

En definitiva, en lo que respecta a los errores aleatorios, los RCT en nutrición muestran tres aspectos principales: a) presentan una gran variabilidad de los valores medidos debido a que los efectos de los ingredientes son pequeños; b) desarrollar RCT con muestras suficientemente grandes, en algunos casos, no es posible a no ser que se disponga de unos recursos económicos extraordinarios y décadas para su finalización; y c) el recurso de los meta-análisis puede no ser una vía para la minimización de los errores porque la jerarquía evidencial estándar sitúa los métodos de síntesis de datos en las posiciones inferiores. La minimización de los errores aleatorios en los RCT en nutrición puede ser imposible en determinados casos.

## V. CONCLUSIÓN

Los RCT con el mayor grado de validez interna presentan un grupo experimental y un grupo de control correctamente diseñados, un error sistemático eliminado y un error aleatorio minimizado. Sobre la base de estos estudios,  $p(H/E) \approx 1$ , donde  $H$  es (*dosis más altas de*) el ingrediente o alimento causa una mejora para la salud y  $E$  es el resultado del ensayo, en el ámbito de la nutrición. Los resultados del RCT solo se pueden explicar por la hipótesis sobre la eficacia (y seguridad) del alimento o por la acción del azar. El nivel de incertidumbre en los datos viene determinado por unos errores aleatorios reducidos. Si bien la conclusión causal siempre se puede poner en cuestión, la probabilidad de que los resultados del ensayo se deban a la acción del azar es mínima.

En la práctica científica, la evidencia de los RCT puede encontrarse muy lejos de ese nivel de certidumbre. En un RCT en nutrición, los resultados podrían explicarse por la eficacia del alimento o ingrediente para causar una mejora para la salud. Sin embargo, esta es solo una explicación entre otras con un nivel de plausibilidad importante: los resultados podrían deberse a errores sistemáticos si no se garantiza el doble ciego o podrían ser producto de la acción del azar. Los errores aleatorios pueden ser significativos en el ámbito de la nutrición debido a los efectos pequeños, las muestras reducidas o a la falta de meta-análisis.

La minimización de los errores aleatorios en los RCT puede ser difícil o incluso imposible en contextos en los que: a) los ensayos con muestras suficientemente grandes presentan complicaciones considerables, y b) la jerarquía evidencial estándar desincentiva el uso de meta-análisis, al situarlos en la posición inferior de la clasificación metodológica. En el ámbito de la nutrición, factores prácticos y factores metodológicos pueden llevar a que no se pueda rechazar la hipótesis de la acción del azar en los resultados de los RCT. En tal caso,  $p(H/E)$  puede estar muy lejos de 1, siendo el nivel de incertidumbre en los datos demasiado elevado.

*Departamento de Filosofía  
Universidad de las Islas Baleares  
Ed. Ramon Llull  
Cra. de Valldemossa, km 7.5  
07122 Palma, España  
E-mail: robertolopezmas@gmail.com*

## AGRADECIMIENTOS

Esta publicación es parte del proyecto de I+D+i PID2020-113449GB-I00, financiado por MCIN/AEI/10.13039/501100011033/. Agradezco el apoyo financiero de la ayuda pre-doctoral PRE2018-085271 y de la Universidad de las Islas Baleares. Aprecio las valiosas observaciones de Barbara Osimani y José Luis Luján, así como los detallados comentarios de dos revisores anónimos.

## REFERENCIAS BIBLIOGRÁFICAS

- AGGETT, P. J. (2012), 'Dose-Response Relationships in Multifunctional Food Design: Assembling the Evidence'; *International Journal of Food Sciences and Nutrition*, vol. 63(S1), pp. 37-42. doi:10.3109/09637486.2011.636344
- AGGETT, P. J., ANTOINE, J. M., ASP, N. G., BELLISLE, F., CONTOR, L., CUMMINGS, J. H., HOWLETT, J., MÜLLER, D. J. G., PERSIN, C., PIJLS, L. T. J., RECHKEMMER, G., TUIJELAARS, S., y VERHAGEN, H. (2005), 'Passclaim: Consensus on Criteria'; *European Journal of Nutrition*, vol. 44(Suppl. 1), pp. I/5-I/30. doi:10.1007/s00394-005-1104-3
- AKOBENG, A. K. (2008), 'Assessing the Validity of Clinical Trials'; *Journal of Pediatric Gastroenterology and Nutrition*, vol. 47(3), pp. 277-282. doi: 10.1097/MPG. 0b013e31816c749f
- ALTHUBAITI, A. (2016), 'Information Bias in Health Research: Definition, Pitfalls, and Adjustment Methods'; *Journal of Multidisciplinary Healthcare*, vol. 9(1), pp. 211-217. doi:10.2147/JMDH.S104807
- ANDROLETTI, A., y TEIRA, D. (2019), 'Rules Versus Standards: What Are the Costs of Epistemic Norms in Drug Regulation?'; *Science, Technology, & Human Values*, vol. 44(6), pp. 1093-1115. doi:10.1177/0162243919828070
- BARNARD, N. D., WILLETT, W. C., y DING, E. L. (2017), 'The Misuse of Meta-analysis in Nutrition Research'; *JAMA*, vol. 318(15), pp. 1435-1436. doi:10.1001/jama.2017.12083
- BIESALSKI, H. K., AGGETT, P. J., ANTON, R., BERNSTEIN, P. S., BLUMBERG, J., HEANEY, R. P., HENRY, J., NOLAN, J. M., RICHARDSON, D. P., VAN OMMEN, B., WITKAMP, R. F., RIJKERS, G. T., y ZÖLLNER, I. (2011), '26th Hohenheim Consensus Conference, September 11, 2010 Scientific Substantiation of Health Claims: Evidence-Based Nutrition'; *Nutrition*, vol. 27(10 Suppl.), pp. S1-S20. doi:10.1016/j.nut.2011.04.002
- BLAIR, E. (2004), 'Gold is Not Always Good Enough: The Shortcomings of Randomization When Evaluating Interventions in Small Heterogeneous Samples'; *Journal of Clinical Epidemiology*, vol. 57(12), pp. 1219-1222. doi:10.1016/j.jclinepi.2004.06.003
- BLUMBERG, J., HEANEY, R. P., HUNCHAREK, M., SCHOLL, T., STAMPFER, M., VIETH, R., WEAVER, C. M., y ZEISEL, S. H. (2010), 'Evidence-Based Criteria in the Nutritional Context'; *Nutrition Reviews*, vol. 68(8), pp. 478-484. doi:10.1111/j.1753-4887.2010.00307.x

- CARTWRIGHT, N. (2007), 'Are RCTs the Gold Standard?'; *BioSocieties*, vol. 2(1), pp. 11-20. doi:10.1017/S1745855207005029
- CARTWRIGHT, N., y STEGENGA, J. (2011), 'A Theory of Evidence for Evidence-Based Policy'; en Dawid, P., Twining, W. y Vasilaki, M. (eds.), *Evidence, inference and enquiry*, Oxford, Oxford University Press, pp. 291-322.
- CLARKE, B., GILLIES, D., ILLARI, P., RUSSO, F., y WILLIAMSON, J. (2013). 'The Evidence that Evidence-Based Medicine Omits'; *Preventive Medicine*, vol. 57(6), pp. 745-747. doi:10.1016/j.ypmed.2012.10.020
- DE BOER, A., VOS, E., y BAST, A. (2014), 'Implementation of the Nutrition and Health Claim Regulation – The Case of Antioxidants'; *Regulatory Toxicology and Pharmacology*, vol. 68(3), pp. 475-487. doi:10.1016/j.yrtph.2014.01.014
- EFSA (2010), 'Guidance on Human Health Risk-Benefit Assessment of Food'; *EFSA Journal*, 8(7):1673, pp. 1-41. doi:10.2093/j.efsa.2010.1673
- (2021), 'Scientific and Technical Guidance for the Preparation and Presentation of a Health Claim Application (Revision 3)'; *EFSA Journal*, 19(3):6554, pp. 1-35. doi:10.2903/j.efsa.2021.6554
- FOTI, K., FORAKER, R. E., MARTYN-NEMETH, P., ANDERSON, C. A. M., COOK, N. R., LICHTENSTEIN, A. H., DE FERRANTI, S. D., YOUNG, D. R., HIVERT, M-F., ROSS, R., DEEDWANIA, P., WHITSEL, L. P., y APPEL, L. J. (2020), 'Evidence-Based Policy Making for Public Health Interventions in Cardiovascular Diseases: Formally Assessing the Feasibility of Clinical Trials'; *Circulation: Cardiovascular Quality and Outcomes*, vol. 13(11):e0063781-10. doi:10.1161/CIRCOUTCOMES.119.006378
- HEANEY, R. P. (2006), 'Nutrition, Chronic Disease, and the Problem of Proof'; *The American Journal of Clinical Nutrition*, vol. 84(3), pp. 471-472. doi:10.1093/ajcn/84.3.471
- (2008), 'Nutrients, Endpoints, and the Problem of Proof', *The Journal of Nutrition*, vol. 138(9), pp. 1591-1595. doi:10.1093/jn/138.9.1591
- JACOBS, D. R., TAPSELL, L. C., y TEMPLE, N. J. (2011), 'Food Synergy: The Key to Balancing the Nutrition Research Effort'; *Public Health Reviews*, vol. 33(2), pp. 507-529. doi:10.1007/BF03391648
- JUKOLA, S. (2019), 'On the Evidentiary Standards for Nutrition Advice', *Stud Hist Philos Biol Biomed Sci*, vol. 73, pp. 1-9. doi:10.1016/j.shpsc.2018.05.007
- KARANICOLAS, P. J., FARROKYAR, F., y BHANDARI, M. (2010), 'Blinding: Who, What, When, Why, How?'; *Canadian Journal of Surgery*, vol. 53(5), pp. 345-348.
- LA CAZE, A. (2009), 'Evidence-Based Medicine Must Be...'; *Journal of Medicine and Philosophy*, vol. 34(5), pp. 509-527. doi:10.1093/jmp/jbp034
- LICHTENSTEIN, A. H., PETERSEN, K., BARGER, K., HANSEN, K. E., ANDERSON, C. A. M., BAER, D. J., LAMPE, J. W., RASMUSSEN, H., y MATTHAN, N. R. (2021), 'Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials'; *Advances in Nutrition*, vol. 12(1), pp. 4-20. doi:10.1093/advances/nmaa109

- LÓPEZ-MAS, R. (2021), 'Evidencia científica y alimentos funcionales: la regulación de las declaraciones de salud en la Unión Europea'; *Política y Sociedad*, vol. 58(3), pp. 1-11. doi:10.5209/poso.72995
- LUJÁN, J. L., y TODT, O. (2021), 'Evidence Based Methodology: A Naturalistic Analysis of Epistemic Policies in Regulatory Science'; *European Journal for Philosophy of Science*, vol. 11(26), pp. 1-19. doi:10.1007/s13194-020-00340-7
- MARTÍNEZ, S. V., y SIANI, A. (2017), 'Health Claims Made on Food in The EU: The Edge between Scientific Knowledge and Regulatory Requirements'; *Trends in Food Science & Technology*, vol. 69(B), pp. 315-323. doi:10.1016/j.tifs.2017.01.005
- NEALE, E. P., y TAPSELL, L. C. (2019), 'Perspective: The Evidence-Based Framework in Nutrition and Dietetics: Implementation, Challenges, and Future Directions'; *Advances in Nutrition*, vol. 10(1), pp. 1-8. doi:10.1093/advances/nmy113
- OSIMANI, B. (2020), 'Epistemic Gains and Epistemic Games: Reliability and Higher Order Evidence in Medicine and Pharmacology'; en La Caze, A. y Osimani, B. (eds.), *Uncertainty in pharmacology: epistemology, methods, and decisions*, Cham, Springer, pp. 345-372.
- OSIMANI, B., BERTOLASO, M., POELLINGER, R., y FRONTONI, E. (2019), 'Real and Virtual Clinical Trials: A Formal Analysis'; *Topoi*, vol. 38(4), pp. 411-422. doi:10.1007/s11245-018-9563-3
- OSIMANI, B., y MIGNINI, F. (2015), 'Causal Assessment of Pharmaceutical Treatments: Why Standards of Evidence Should Not Be the Same for Benefits and Harms?'; *Drug Safety*, vol. 38, pp. 1-11. doi:10.1007/s40264-014-0249-5
- REISS, J. (2015), 'A Pragmatist Theory of Evidence'; *Philosophy of Science*, vol. 82(3), pp. 341-362. doi:10.1086/681643
- RENJITH, V. (2017), 'Blinding in Randomized Controlled Trials: What Researchers Need to Know?'; *Manipal Journal of Nursing and Health Sciences*, vol. 3(1), pp. 45-50.
- RICHARDSON, D. P. (2012), 'Preparing Dossiers: Strength of the Evidence and Problems of Proof'; *Proceedings of the Nutrition Society*, vol. 71(1), pp. 127-140. doi:10.1017/S002966511100317X
- RUSSO, F., y WILLIAMSON, J. (2007), 'Interpreting Causality in the Health Sciences'; *International Studies in the Philosophy of Science*, vol. 21(2), pp. 157-170. doi:10.1080/02698590701498084
- SANZ-MERINO, N., y LUJÁN, J. L. (2021), 'Políticas Epistémicas y Gobernanza. Un Análisis Comparativo de las Regulaciones Norteamericana y Europea sobre Declaraciones de Propiedades Saludables'; *Revista Española de Ciencia Política*, vol. 55, pp. 93-117. doi:10.21308/recp.55.04
- SCHWINGSHACKL, L., KNÜPPEL, S., SCHWEDHELM, C., HOFFMANN, G., MISSBACH, B., STELMACH-MARDAS, M., DIETRICH, S., EICHELMANN, F., KONTOPANTEILS, E., IQBAL, K., ALEKSANDROVA, K., LORKOWSKI, S., LEITZMANN, M. F., KROKE, A., y BOEING, H. (2016), 'Perspective: NutriGra-



- de: a Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research?; *Advances in Nutrition*, vol. 7(6), pp. 994-1004. doi:10.3945/an.116.013052
- SENN, S. (2007), *Statistical Issues in Drug Development*, Chichester, John Wiley & Sons. ISBN:978-0-470-01877-4
- SPIETH, P. M., KUBASCH, A. S., PENZLIN, A. I., ILLIGENS, B. M., BARLINN, K., y SIEPMANN, T. (2016), 'Randomized Controlled Trials – A Matter of Design?'; *Neuropsychiatric Disease and Treatment*, vol. 12(1), pp. 1341-1349. doi:10.2147/NDT.S101938
- STANLEY, D. J., y SPENCE, J. P. (2014), 'Expectations for Replications: Are Yours Realistic?'; *Perspectives on Psychological Science*, vol. 9(3), pp. 305-318. doi:10.1177/1745691614528518
- STEGENGA, J. (2014), 'Down with the Hierarchies'; *Topoi*, vol. 33(2), pp. 313-322. doi:10.1007/s11245-013-9189-4
- STEPHENSON, J. M., y BABIKER, A. (2000), 'Overview of Study Design in Clinical Epidemiology'; *Sexually Transmitted Infections*, vol. 76(4), pp. 244-247. doi:10.1136/sti.76.4.244
- TODT, O., y LUJÁN, J. L. (2017a), 'The Role of Epistemic Policies in Regulatory Science: Scientific Substantiation of Health Claims in The European Union'; *Journal of Risk Research*, vol. 20(4), pp. 551-565. doi:10.1080/13669877.2015.1100661
- (2017b), 'Health Claims and Methodological Controversy in Nutrition Science'; *Risk Analysis*, vol. 37(5), pp. 958-968. doi:10.1111/risa.12665
- VANDENBROUCKE, J. P., BROADBENT, A., y PEARCE, N. (2016), 'Causality and Causal Inference in Epidemiology: The Need for a Pluralistic Approach'; *International Journal of Epidemiology*, vol. 45(6), pp. 1776-1786. doi:10.1093/ije/dyw341
- VIERA, A. J., y BANGDIWALA, S. I. (2007), 'Eliminating Bias in Randomized Controlled Trials: Importance of Allocation Concealment and Masking'; *Family Medicine*, vol. 39(2), pp. 132-137.
- VISIOLI, F. (2012), 'Can Experimental Pharmacology Be Always Applied to Human Nutrition?'; *International Journal of Food Sciences and Nutrition*, vol. 63(S1), pp. 10-13. doi:10.3109/09637486.2012.665439
- WORRALL, J. (2007), 'Why There's No Cause to Randomize'; *The British Journal for the Philosophy of Science*, vol. 58(3), pp. 451-488. doi:10.1093/bjps/axm024

# Frank Ramsey

Cheryl Misak



A SHEER  
EXCESS *of*  
POWERS