

NORMAS ÉTICAS Y ESTADÍSTICAS EN LA JUSTIFICACIÓN DE LOS ENSAYOS CLÍNICOS ALEATORIZADOS

MARTA GARCÍA ALONSO

Departamento de Filosofía y Filosofía Moral y Política
Universidad Nacional de Educación a Distancia (UNED)
mgalonso@fsf.uned.es

DAVID TEIRA SERRANO

Departamento de Lógica, Historia y Filosofía de la Ciencia
Universidad Nacional de Educación a Distancia (UNED)
Department of Science and Technology Studies
UCL
dteira@fsf.uned.es

RESUMEN: En este trabajo analizamos cómo se articula la justificación estadística de las conclusiones obtenidas en los ensayos clínicos aleatorizados con sus condiciones de aceptabilidad ética y política. Pretendemos mostrar cómo varía tal justificación dependiendo de nuestra concepción de la probabilidad y, correlativamente, qué argumentos éticos se pueden ofrecer sobre cada una de ellas. Veremos cómo los enfoques frecuentista y bayesiano resultan defendibles en el ámbito de los ensayos clínicos y defenderemos que, para su aceptación pública, se les debe exigir garantías de imparcialidad cuyo mejor ejemplo es la propia técnica de aleatorización.

PALABRAS CLAVE: bioética, aleatorización, imparcialidad, experimentación médica, probabilidad

SUMMARY: This paper analyses how to combine the statistical justification of the results obtained in a randomised clinical trial and the ethical and social arguments that support its implementation. It aims at showing how this justification varies according to the view on probability we adopt and the different ethical cases that can be made to sustain it. Both the frequentist and the Bayesian approaches to the design of clinical trials will be seen as equally defensible on these grounds. To be socially acceptable both approaches must meet an impartiality requirement, which is best served by the technique of randomisation.

KEY WORDS: bioethics, randomisation, impartiality, medical experimentation, probability

1. *Introducción*

Un ensayo clínico, en sentido amplio, es toda evaluación experimental de una sustancia o medicamento mediante su administración a un sujeto, para comprobar sus efectos y establecer su eficacia. Una vez superada la etapa de investigación en animales, se suele proceder en varias fases experimentando sobre seres humanos; de estas fases suele considerarse decisiva la tercera: aquella en la que se compara

el efecto del tratamiento experimental con respecto a uno de control, asignando uno u otro aleatoriamente a los pacientes de acuerdo con un diseño experimental justificado estadísticamente.¹ Los ensayos clínicos aleatorizados (de ahora en adelante, ECA) constituyen actualmente la *regla de oro* tanto de los investigadores como de las autoridades sanitarias para decidir sobre la eficacia de un fármaco.²

En este trabajo pretendemos analizar cómo se articula la justificación inferencial de los ECA con su aceptabilidad ética y social. Los ECA, como muchas otras técnicas de contraste de hipótesis, se apoyan en una concepción frecuentista de la probabilidad, según la cual nuestra inferencia sobre los resultados del ensayo se justificaría por inducirnos a error sólo en un porcentaje prefijado de repeticiones de nuestro experimento. Hoy, sin embargo, se están proponiendo ensayos clínicos apoyados sobre una concepción subjetiva de la probabilidad, en la que, con independencia del diseño experimental, basamos nuestras conclusiones en la revisión bayesiana de nuestras probabilidades *a priori* (grados de creencias) a la vista de los datos. Como veremos, la justificación implícita en ambos criterios de prueba es muy distinta, y el dilema que se abre ante nosotros es decidir si son igualmente admisibles.

Vale la pena abordar este dilema epistemológico desde el punto de vista de sus implicaciones éticas. Existen ya buenos análisis en los que se pone de manifiesto que un ECA será aceptable sólo si en su desarrollo se observan ciertos principios normativos (no maleficencia y equidad), por los que deben velar los comités encargados de su supervisión. Como observaremos, la justificación inferencial se subordina aquí a la justificación ética, aun cuando después le proporcione unas coordenadas argumentales notablemente exigentes (*e.g.*, en qué consiste el consentimiento informado en un ECA). La aparición de una alternativa bayesiana en el diseño y la ejecución de un ensayo clínico desafía el consenso ético que hoy existe sobre los ECA, pues apela a principios alternativos en su justificación. Como se hará patente, la autonomía del paciente (al optar por un tratamiento) puede prevalecer en su diseño sobre el propio juicio del médico.

En suma, al diseñar un ensayo clínico tenemos criterios de justificación inferencial alternativos, cuyas condiciones de aceptabilidad ética son potencialmente divergentes. ¿Debemos aceptar ambos? ¿En qué condiciones?

¹ Esta definición se apoya en Abajo 1993.

² En 1977, la Food and Drug Administration de Estados Unidos, incorporó los ECA a la normativa de buena práctica clínica, como ocurrió después con la Comunidad Europea en 1990.

Por así decirlo, en este trabajo queremos plantear una respuesta *política* a este dilema. Para ello recordaremos el papel de garante de la imparcialidad médica que se les asignó a los ECA desde su mismo origen y que, todavía hoy, sostiene su aceptación. Que los intereses particulares del médico en el resultado de un ensayo no afecten sus conclusiones es crucial para que el Estado, en representación de los intereses generales de los pacientes, pueda autorizar el uso de un medicamento. Defenderemos aquí que la imparcialidad de nuestros procedimientos estadísticos constituye una condición políticamente anterior a su justificación inferencial y a su aceptabilidad ética. Examinaremos en qué condiciones se garantiza que nuestro criterio estadístico de prueba (frecuentista o bayesiano) sobre la eficacia de un medicamento será imparcial y, en esa medida, socialmente aceptable por equitativa —en el sentido que esto tiene entre los teóricos de la justicia que evalúan nuestras reglas de decisión social bajo un *velo de la ignorancia*—.

Para elaborar nuestra tesis, procederemos según los siguientes pasos. Analizaremos, en primer lugar, el origen de los ECA y la articulación original de argumentos estadísticos y políticos que se dio en ellos. Explicitaremos después su fundamentación ética para explorar, a continuación, la alternativa bayesiana en sus vertientes inferencial y bioética. Veremos, finalmente, el dilema que plantea la elección entre estos dos enfoques y en qué medida la consideración de su imparcialidad nos ayuda a resolverlo.

2. *Dos argumentos en el origen de la aleatorización*

La introducción de los ensayos clínicos aleatorizados se produjo inicialmente a mediados del siglo XX en los sistemas sanitarios británico y estadounidense. Según el análisis de este último propuesto Harry Marks (1997), cabría interpretar este proceso como un tránsito de la *autoridad del experto* a la *autoridad del método*. A principios del siglo XX, la aprobación de un producto para usos terapéuticos era emitida por el Consejo de Farmacia y Química de la Asociación Médica Estadounidense, integrado por investigadores cuyo juicio personal sobre su eficacia era generalmente aceptado por la Administración estadounidense con independencia de los argumentos que lo justificasen. En la segunda mitad del siglo comenzó a exigirse la aplicación de criterios metodológicos impersonales para la evaluación clínica de un medicamento, y los ECA se generalizaron gradualmente como estándar. Los evaluadores debían así asegurar su imparcialidad contra la sospecha de connivencia con los intereses comerciales de las empresas

farmacéuticas, ya evidentes en el último cuarto del siglo XIX (Marks 1997, p. 19).

La autoridad de la ciencia experimental debía regular el mercado de los medicamentos (Marks 2000). No obstante, durante toda la primera mitad del siglo XX se careció de un procedimiento universalmente consensuado para efectuar su evaluación, y la experiencia clínica acumulada personalmente por cada médico solía entrar en contradicción con las conclusiones de la investigación experimental. En 1938, el Congreso estadounidense asignó a la Food and Drug Administration (FDA) la misión de analizar y aprobar el uso terapéutico de nuevos compuestos farmacéuticos, tarea que llevó a cabo de acuerdo con el criterio del Consejo antes citado. No resultó difícil alcanzar un consenso para restringir productos cuya toxicidad era evidente, pero la FDA se encontró con una enconada oposición al negar su autorización a ciertos productos por carecer de eficacia terapéutica, en oposición al criterio de sus fabricantes avalado por el juicio particular de múltiples médicos.

El desarrollo de la investigación militar durante la Segunda Guerra Mundial propició la introducción ulterior de protocolos estadísticos en la experimentación médica y, en particular, la generalización de los ECA. El impulso provino del Reino Unido, donde un matemático contratado por la estación experimental agrícola de Rothamsted venía gestando desde los años 1920 una revolución conceptual en la inferencia estadística. Ronald Aylmer Fisher (1890–1962) estableció algunas de las distinciones y técnicas todavía hoy vigentes en la disciplina (*e.g.*, la distinción entre población y muestra o los estimadores de máxima verosimilitud) y elaboró sobre estos principios una nueva teoría del diseño de experimentos, en las que se fundamentarían los ECA (Marks 2003).

De modo paradigmático, pensemos en el análisis de los efectos de un fertilizante sobre cierto cultivo: ¿cómo interpretar las diferencias de rendimiento entre un terreno abonado y otro sin abonar? El principio aplicado en Rothamsted, la estación experimental para la que Fisher trabajaba, consistía en controlar cuantos factores pudieran afectar al cultivo de modo que tuvieran un valor semejante en ambas parcelas; no obstante, era imposible asegurar que se disponía de una relación completa de tales factores, de modo que la justificación de nuestro análisis sería siempre objetable. Con el fin de superar este dilema inductivo, Fisher propuso aplicar su técnica para contrastar nuestra hipótesis sobre el origen de las diferencias de rendimiento entre ambos cultivos.

Nuestros datos sobre el efecto del fertilizante en las cosechas podrían interpretarse como una muestra extraída del conjunto de resultados que cabría obtener si se repitiese indefinidamente el experimento (plantar en terreno abonado y sin abonar). Supongamos que juzgamos, a la vista de los datos efectivamente obtenidos, si el fertilizante aumenta o no el rendimiento del terreno y pongamos que así lo indican: si repetimos el experimento n veces más, podremos preguntarnos cuántas veces nos habríamos equivocado al juzgar el efecto del abono sobre cada nueva serie de datos y compararlo con nuestra evaluación inicial. Pues bien, la técnica de Fisher nos permite afirmar que, si juzgamos la muestra inicial conforme a sus principios, podemos justificar nuestra evaluación del abono afirmando que sólo nos equivocaremos al razonar así en un mínimo porcentaje de repeticiones del experimento (*e.g.*, 5 de cada 100 veces). Es decir, equivocarse se convierte en un suceso raro (pero no imposible).

Semejante justificación requiere, ante todo, que los datos muestrales tengan un origen perfectamente aleatorio (Fisher 1971, § 26); *i.e.*, al distribuir dos semillas entre las dos parcelas, ambas deberán tener la misma probabilidad de ser asignadas al terreno abonado que al no abonado. De lo contrario, existiría un sesgo que afectaría la probabilidad de equivocarnos al juzgar en sucesivas repeticiones del experimento. Para garantizar el cumplimiento de este requisito, es necesario aleatorizar la distribución de semillas, es decir, asegurarnos de que la asignación a cada terreno sea equiprobable e independiente en todos los casos. De este modo, argumentaba Fisher, nos aseguramos también de que cualesquier otros factores que puedan influir en el desarrollo del cultivo actuarán también al azar sobre las semillas, de manera que sus efectos no afectarán la vigencia de nuestras conclusiones.

Austin Bradford Hill (1897–1991), pionero de la estadística médica en el Reino Unido, supo advertir que tales principios se podían aplicar también al análisis de la eficacia clínica de un medicamento. Además de divulgarla en numerosos artículos y manuales, propició su aplicación en un ensayo clínico precursor sobre el uso de la estreptomina en el tratamiento de la tuberculosis pulmonar (1948) y su generalización ulterior en la investigación médica. El mayor mérito de Hill, al decir de muchos intérpretes, consistió en haber transformado los principios estadísticos de Fisher en argumentos convincentes para el investigador clínico (*e.g.*, Armitage 2003). Para ello evitó apelar directamente a su justificación metodológica como técnica de inferencia inductiva y se concentró en la aleatorización como garante de la imparcialidad médica con respecto al medicamento evaluado.

Al asignar aleatoriamente un tratamiento a cada paciente, el médico evitaba que sus propios sesgos afectasen al resultado del ensayo. En palabras del propio Hill:

[La aleatorización] asegura que ni nuestra idiosincrasia personal (nuestros gustos o disgustos aplicados consciente o inadvertidamente) ni nuestra carencia de un juicio ponderado afecte la construcción de los diferentes grupos de tratamiento —la asignación está más allá de nuestro control y los grupos resultan, por lo tanto, insesgados—. (Hill 1952, p.)³

Tenemos, por consiguiente, un doble sentido de sesgo. En primer lugar, una propiedad deseable de un estimador estadístico (*e.g.*, la media muestral evaluada en un ensayo clínico) es que sea insesgado, es decir, el error de estimación en que podemos incurrir será, en promedio, nulo. La aleatorización no es condición suficiente, pero sí necesaria para que se obtenga esta propiedad en algunos estimadores, pues, según supo apreciar Hill, con ella evitamos que se dé una desviación en los datos a consecuencia de la intervención del experimentador. Esto constituye, en segundo lugar, una garantía de su imparcialidad: sus preferencias personales no afectarán la asignación de tratamientos y, por lo tanto, nos aseguramos de que no existirán conflictos de intereses que vicien el resultado del experimento. Es decir, se sustituye el juicio personal por la autoridad del método y queda así descartada cualquier sospecha de connivencia entre el experimentador y la empresa farmacéutica cuyo producto se somete a ensayo. Como decíamos, este segundo argumento resultó, al parecer, decisivo en la aceptación generalizada de los ECA entre los médicos británicos y estadounidenses en la segunda mitad del siglo XX (Marks 2000).

De hecho, si atendemos al análisis de Marks, la justificación inferencial de la aleatorización propuesta por Fisher apenas influyó en la opinión médica. Es más, a menudo resulta dudoso que se haya interpretado correctamente, y esta duda todavía persiste en nuestros días, a juicio de algunos críticos. No obstante, como veremos a continuación, la conexión entre ambos argumentos (la justificación inferencial y la garantía de imparcialidad) resulta crucial para establecer en qué condiciones resulta éticamente aceptable la aleatorización.

³ Citado y comentado por Armitage 2003, p. 927.

3. *Bioética de los ensayos clínicos aleatorizados*

Partimos, por lo tanto, de un doble sentido de imparcialidad: estimación insesgada del valor verdadero de un parámetro e indiferencia del resultado del ensayo respecto a los intereses particulares del investigador. Hemos de añadir ahora una tercera acepción: también desde un punto de vista clínico ambas terapias deben resultarle indiferentes al investigador:

Si el médico sabe, o tiene buenas razones para creer, que una nueva terapia (*A*) es mejor que otra (*B*), no puede participar en un ensayo comparativo de ambas terapias. Éticamente, el médico está obligado a proporcionar la terapia *A* a cada nuevo paciente con necesidad de una de ellas.

Si el médico (o sus pares) tienen dudas genuinas acerca de cuál sea la mejor terapia, debe proporcionar a cada paciente igual oportunidad de recibir cada una de ellas. (Levine 1988, p. 187)

En otras palabras, un ensayo clínico sólo resultará aceptable desde un punto de vista ético si sometemos a contraste estadístico la hipótesis (“nula”) de que ambas terapias son equivalentes (frente a la hipótesis de que no lo son). En ese caso, la aleatorización constituirá la mejor expresión de la igualdad de oportunidades del paciente. Tenemos aquí, por lo tanto, un tercer argumento (esta vez bioético) para justificarla que se desarrolla en dos pasos. En primer lugar, aplicando al caso el principio de *no maleficencia*, podremos defender la aleatorización sobre la base de la indeterminación clínica [*clinical equipoise*] de ambos tratamientos. Como establecen De Abajo y Gracia (1997), “una intervención en la que no exista una indeterminación clínica real es maleficiente porque entra un exceso de riesgo evitable en las personas a las que se ofrece el tratamiento inferior”. Sólo en estas condiciones resulta justa una asignación equiprobable de cada par de tratamientos, suponiendo que al iniciar el ensayo podamos creer que *el beneficio esperado sea igual en ambos*; aplicamos, además, un segundo principio, el de *equidad*, que nos exige una distribución paritaria no sólo de costos, sino de ganancias.

La articulación de este tercer argumento con los dos anteriores no es inmediata. La aleatorización es necesaria para justificar inferencialmente las conclusiones del ensayo, pero puede ocurrir que un médico ponga reparos a la inclusión de un determinado paciente en un ensayo, por dudar de que *en su caso* la equivalencia de ambos

tratamientos sea cierta. El juicio médico sobre el caso particular socava aquí el juicio estadístico con respecto a la validez general del experimento. El origen de la dificultad se encuentra en el tipo de perspectiva implícita en ambas posturas. El estadístico se propone efectuar una inferencia sobre la eficacia de un tratamiento en una población de enfermos a partir de una muestra que extrae de ella aleatoriamente; es decir, cada enfermo tiene iguales probabilidades de ser escogido y, a estos efectos, resultan todos ellos equivalentes. Desde un punto de vista clínico, a menudo se puede dudar de tal equivalencia sobre la base de criterios derivados con frecuencia de la propia experiencia personal del médico. Los dos principios éticos anteriores nos exigen, en este caso, restar justificación estadística a nuestro ensayo, asignando un tratamiento no aleatorio a un paciente. En este sentido, el médico deja de ser imparcial o, al menos, no puede justificar su neutralidad sobre la base de un procedimiento estadístico; sin embargo, ya que no sirve a sus intereses particulares, sino a los del paciente, su parcialidad resulta éticamente aceptable.

Según sugiere Robert J. Levine, dificultades como éstas sugieren la conveniencia de pronunciarse caso por caso sobre la aceptabilidad ética de los ECA. No obstante, existe un principio general asentado sobre el consenso médico (articulado, en una proporción u otra, sobre los tres argumentos expuestos): un ECA constituye la *regla de oro* para evaluar la eficacia de una terapia y, por lo tanto, sólo cabe dar por plenamente justificado nuestro juicio sobre ella una vez concluido éste. En otras palabras, al dar su consentimiento a participar en un ECA, el paciente reconoce su valor como criterio (a menudo, exclusivo) de prueba, de acuerdo con lo establecido por la comunidad científica (Levine 1988, p. 201).

Dos son las implicaciones que aquí nos interesa señalar a este respecto. En primer lugar, quien acepta tomar parte en un ECA obviamente basa su consentimiento en un juicio médico sobre la indeterminación clínica de nivel epistemológico inferior al que se obtendrá con el propio ensayo. El conocimiento adquirido en los ensayos efectuados en las fases previas de evaluación de un medicamento carece de la justificación inferencial que la aleatorización proporciona en la fase III. Por consiguiente, el paciente debe confiar en el juicio personal del médico, que el ECA pretende superar. Por la misma razón, si durante el desarrollo del ensayo aparecieran datos preliminares sugiriendo la superioridad de uno de los dos tratamientos sobre el otro, el paciente deberá seguir con el que le

fuese originalmente asignado, pues tales datos no serán considerados concluyentes mientras no se complete el ECA.⁴

En segundo lugar, una vez finalizado el ECA, idealmente estaríamos en situación de establecer (*e.g.*) que la eficacia de un fármaco *A* es superior a la de *B* con un nivel de significación $p = 0.05$. Es decir, como ya adelantábamos en la sección 2, si se repitiera el ensayo, supongamos, 100 veces, sólo en 5 concluiríamos erróneamente que la eficacia de *A* es superior a la de *B*.⁵ Obviamente, cabe preguntarse si los datos que acabamos de obtener en nuestro ECA pertenecerán a ese 5 por ciento de los casos. El argumento de Fisher es que si así fuera, ése sería un suceso raro, de modo que si aceptamos por convención ese umbral de certidumbre, estaremos probabilísticamente justificados al afirmar nuestras conclusiones. Éste es el criterio de prueba que se le ofrece al paciente al que se le receta un medicamento aprobado en la fase III.

En resumidas cuentas, la aleatorización es objeto de un argumento ético (sólo es aceptable si se cumplen los principios de no maleficencia y equidad), pero la justificación inferencial de los ECA tiene también implicaciones éticas. Determina los niveles de aceptabilidad de la información estadística a la que debe dar su consentimiento expreso el paciente que participa en un ECA —y su asentimiento tácito, el enfermo que se beneficia después del medicamento aprobado—. Así, a continuación veremos cómo la articulación de los ensayos clínicos sobre un patrón estadístico alternativo viene a romper este consenso y qué consecuencias éticas se seguirán de ello.

4. *La crítica bayesiana*

Como pudo apreciarse en los dos epígrafes anteriores, la justificación inferencial que podemos dar a nuestras conclusiones en un ECA

⁴ *Cfr.* Freedman 1987 para una discusión de este último punto.

⁵ Un evaluador nos señala el problema de la justificación ética de semejante umbral, que constituye de por sí una simple convención estadística. De hecho, este umbral puede aumentar o disminuir según el criterio de los autores del estudio —por ejemplo, dependiendo de los riesgos para la salud que puedan derivarse de su error—. No obstante, conviene advertir que las agencias reguladoras supervisan, en la denominada fase IV del ensayo, los efectos de cada medicamento tras su comercialización, lo cual constituye una admisión implícita de la posibilidad de que el ensayo clínico yerre en sus conclusiones (y del deber de evitar efectos perversos de ese error). Por lo tanto, más que buscar una justificación ética de un nivel de confianza en particular, creemos que debiera ser objeto de discusión política cuál ha de ser el procedimiento de las agencias reguladoras para evitar los errores que puedan derivarse de cualquier prueba estadística (frecuentista o bayesiana). *Cfr.*, a este respecto, Cranor 2004.

apela a las *repeticiones* del experimento: estimamos la probabilidad de equivocarnos a partir de las secuencias de errores y aciertos que generaría nuestra regla para interpretar sus resultados. Como es sabido, esta concepción de la probabilidad se denomina *frecuentismo*, y su representante más eximio fue seguramente Richard von Mises.⁶ Su intuición central es que la teoría de la probabilidad constituía una ciencia matemática de los fenómenos de masas (*e.g.*, el crecimiento de una población biológica) o los sucesos repetitivos (*e.g.*, el lanzamiento de un dado). Cada uno de estos colectivos está compuesto por elementos que poseen un atributo cuyo valor varía aleatoriamente —en los ejemplos anteriores, el número de hijos por familia o el valor de la cara del dado—. En colectivos empíricos, estos valores dan lugar a secuencia finitas, que la teoría de la probabilidad analizaría matemáticamente abstrayendo sus leyes —a partir de las cuales se obtendría, a su vez, la axiomática de Kolmogorov—. Fisher compartía esta concepción y supo implementarla en sus técnicas estadísticas.

Pero, como también es sabido, existe una concepción alternativa de la probabilidad, a partir de la cual también se puede derivar la definición axiomática de Kolmogorov. Interpretamos ahora la probabilidad como una representación numérica de nuestros grados de creencia, que alcanzamos analizando nuestras apuestas. Al apostar, en efecto, ponderamos la probabilidad de que se dé un acontecimiento o de que una proposición sea verdadera según el principio de que más estaremos dispuestos a apostar cuanta más certeza tengamos sobre lo que vaya a ocurrir. Numéricamente, podemos expresar este grado de creencia mediante nuestro cociente de apuestas (CA): la proporción entre la cantidad que se juega cada cual con el total de la apuesta. Establecido esto, cabe evaluar si nuestras apuestas son coherentes —esto es, si nuestros CA no dan lugar a una *apuesta holandesa*—. Sólo podemos garantizar la coherencia de nuestros CA si y sólo si cumplen con los tres principios básicos de la probabilidad —normalidad, certeza y aditividad—, generalizados por Kolmogorov en su axiomática.

Ahora bien, es posible aplicar también esta segunda concepción de la probabilidad al análisis de los ensayos clínicos, y una consecuencia inmediata al hacerlo es que la asignación aleatoria de tratamientos ya no será necesaria para la justificación inferencial de nuestras conclusiones. No necesitamos suponer que el experimento se repetirá para estimar nuestra probabilidad de equivocarnos, pues ésta sólo expresa-

⁶ Sobre las distintas concepciones de la probabilidad, nos servimos como referencia de Gillies 2000.

rá la intensidad de nuestra creencia en nuestra conclusión a partir de la evidencia disponible (con independencia de su origen). De hecho, a juicio de los partidarios de la probabilidad personal, resulta dudoso que la aleatorización nos proporcione unas bases adecuadas para los contrastes de significación. Veamos dos objeciones al respecto.

Idealmente, en un ECA quedarían distribuidas al azar cualesquier influencias que el investigador no controlara *ex ante*, bien a consecuencia de la propia aleatorización (*e.g.*, la asignación de tratamientos), bien de un modo espontáneo (*e.g.*, la ubicación del paciente en el hospital). En realidad, el investigador siempre debe decidir qué factores son relevantes y deben ser aleatorizados (las *variables pronóstico*) y cuáles no. Pero puede ocurrir, argumentan los críticos, que alguno de estos últimos afecte de distinto modo a nuestros dos grupos de pacientes y cause un sesgo en el resultado del experimento.⁷ La objeción no es puramente escéptica: no se trata de poner en duda el alcance de nuestra ciencia médica, sino de mostrar que la aleatorización no nos protege contra *todo* factor incontrolado. Por consiguiente, el resultado de un ECA no será tan concluyente como muchos de sus partidarios pretenden.

Un segundo argumento se dirige contra la propia idea de población de la que se extraen los pacientes sobre los que se efectúa la aleatorización (Urbach 1993, p. 1424). En un ECA podemos afirmar que cada paciente tiene igual probabilidad de recibir un tratamiento; pero, ¿es posible generalizar sobre esta base las conclusiones que obtengamos al conjunto de la población? Como vimos en la sección 2, éste es precisamente el tipo de justificación inferencial que pretenden alcanzar los contrastes de significación. A diferencia de lo que ocurre con las semillas en la experimentación agrícola, resulta difícil establecer una definición estadística de población respecto de la cual los pacientes incluidos en un ECA constituyan una muestra aleatoria. De nuevo, la dificultad radica en establecer el conjunto de variables relevantes para el análisis de una enfermedad que definan en qué consiste ser un enfermo, de modo tal que se incluyan todos aquellos casos no diagnosticados en el presente y todos aquellos que se verán aquejados en un futuro. En consecuencia, nuevamente, la justificación que la aleatorización proporciona a nuestras inferencias estadísticas en los ECA no es tan concluyente como podría ser en otros contrastes de significación.

Lo que objeciones como éstas pretenden poner de manifiesto es que el diseño de los ECA es internamente problemático, aun para

⁷ *Cfr.* Urbach 1985, pp. 261–262; Worrall 2002, pp. 323–324.

aquellos que explícitamente adoptan un enfoque frecuentista. Por lo tanto, deberíamos considerar qué tipo de justificación nos ofrecen alternativamente los ensayos clínicos articulados en una perspectiva bayesiana.⁸

Supongamos que asignamos una probabilidad *a priori* $P(H)$ a una hipótesis H sobre la eficacia de un fármaco de acuerdo con nuestro grado de creencia (a mayor intensidad, su valor será más cercano a 1). Nos interesará saber cómo varía esa probabilidad a la vista de los datos que obtengamos de un ensayo clínico, es decir, la probabilidad condicional $P(H|\text{datos})$. El teorema de Bayes establece que $P(H|\text{datos})$ es proporcional al producto $P(\text{datos}|H) \cdot P(H)$. Cualquier investigador tendrá su asignación subjetiva $P(H)$, como también cualquier paciente interesado en el fármaco. Aplicando este teorema, podrá estimar su probabilidad personal sobre la superioridad de un fármaco frente a la terapia de control para un futuro enfermo.

Al diseñar el ensayo, la asignación de los pacientes a los grupos de control dependerá de la probabilidad que el investigador asigne al efecto que cada fármaco tendrá sobre cada uno de ellos. Un par de pacientes será intercambiable *sii* tal probabilidad es idéntica. La aleatorización puede ser útil, por lo tanto, para evitar sesgos personales en la asignación, pero ya no por consideraciones probabilísticas. Por eso mismo, se vuelve posible utilizar, con las cautelas adecuadas, los registros médicos disponibles sobre los efectos de un fármaco.

Las ventajas de este enfoque con respecto al frecuentista al justificar nuestra inferencia no son menores. Una muy significativa se encuentra en la simplicidad de interpretación al poder asignarse directamente las probabilidades. En una perspectiva bayesiana, podemos concluir “la probabilidad de que no haya diferencia entre las terapias A y B es de 0.05”, como expresión de nuestro grado de creencia particular. En una perspectiva frecuentista, como se recordará, la conclusión será más bien “la probabilidad de que se dé un suceso tan extremo o más que el observado es de 0.05, dado el diseño del estudio y suponiendo que no haya otra diferencia en las terapias”. Correlativamente, cada médico podrá asignar una probabilidad personal sobre el efecto del tratamiento a un paciente en particular, lo cual resulta impracticable para el frecuentista.

Los inconvenientes tampoco son menores: el diseño del experimento se complica para considerar cuantos interrogantes resten generalidad a sus conclusiones y éstas no podrán pretender la objetividad

⁸ La exposición siguiente está basada en Berry 1993.

pretendidamente alcanzada con los contrastes de significación. No obstante, aun siendo una técnica minoritaria, el debate está abierto: la FDA estadounidense está considerando ensayos clínicos diseñados en una perspectiva bayesiana en lo que quizá constituya el preámbulo de una discusión internacional sobre su implantación (Couzin 2004).

5. *Argumentos éticos sobre el bayesianismo*

La situación es entonces la siguiente. Tenemos dos técnicas estadísticas para justificar nuestras inferencias sobre la eficacia de un fármaco a partir del resultado de un ensayo clínico. Cada una de ellas está articulada sobre una concepción diferente de la probabilidad (aun cuando asuman una misma definición axiomática). Esto pone de manifiesto que ambas técnicas, a pesar de resultar ocasionalmente coincidentes en sus estimaciones, nos proponen dos tipos de justificación, en principio, inconmensurables. El frecuentista apelará a la objetividad de las series de datos que se darían al repetirse un experimento, mientras que el bayesiano argumentará sobre el carácter irreductiblemente personal de nuestros grados de creencia. Dada esta divergencia de principio, aun cuando el debate estadístico profundice seguramente en los defectos y virtudes que en cada caso se puedan apreciar en ambas técnicas, es presumible que no baste para zanjar la cuestión de qué valor concedemos a cada una de ellas para justificar nuestras decisiones clínicas. ¿Qué ayuda podemos esperar de la argumentación ética y política?

Los partidarios del enfoque bayesiano defienden su causa también sobre bases éticas. Como veíamos a propósito de la justificación inferencial, comienzan por socavar la pretendida superioridad de los ECA, cuestionando el sentido que pueda darse a la *indeterminación clínica*. Ya se dijo que, según el diseño actualmente vigente de los ERC, tal indeterminación depende del juicio personal de los investigadores, pues su validación estadística es justamente el objeto del ensayo. En un enfoque bayesiano, en cambio, un juicio semejante puede expresarse como una probabilidad personal *a priori*: si existe indeterminación genuina, ambos tratamientos tendrán una probabilidad personal de éxito para un investigador de 0.5. Para sostener esta creencia tal investigador no necesitará esperar a la conclusión del ensayo, puesto que se puede apoyar, como buen bayesiano, en datos previamente existentes e incluso en los datos que va generando el ensayo a medida que se desarrolla.

En otras palabras, el frecuentista no puede apelar a la probabilidad para afirmar la indeterminación de ambos tratamientos al iniciar el

ensayo y, desde un punto de vista ético, esto va en detrimento de la asignación aleatoria de tratamientos. Pero, recordémoslo, en un enfoque bayesiano, ésta ya no es necesaria para la validación estadística del ensayo. En realidad, la aceptación del diseño experimental depende ahora del *consentimiento informado* del paciente, en un sentido muy preciso: debe decidir cuánta información necesita para formarse su propia probabilidad personal sobre sus efectos.

Donald Berry propuso a este respecto su propia cláusula de consentimiento informado para tomar parte en un ensayo con un diseño convencional:

“Me gustaría que participase en un ensayo clínico aleatorizado. Lanzaremos una moneda al aire y le asignaremos la terapia *A* si sale cara y la *B* si sale cruz. Ni usted ni yo sabremos qué terapia recibe a menos que aparezcan problemas ulteriores.

[Presentación de las terapias y de sus posibles efectos secundarios]

“Nadie sabe realmente qué terapia es mejor, y por eso efectuamos este ensayo. No obstante, tenemos ya alguna experiencia con ambas terapias, y datos procedentes de este ensayo ya iniciado. La evidencia disponible sugiere que usted viviría en promedio cinco meses más con el tratamiento *A* que con el tratamiento *B*. Pero los datos todavía no son concluyentes, y sucede que hay gente que ha recibido el tratamiento *B* y ha vivido más que algunos pacientes que han seguido el *A*. Si yo fuera usted, quizá preferiría el tratamiento *A*, pues mi probabilidad personal de que usted viva más con *A* es del 25 por ciento.

“Su participación en este ensayo nos ayudará a tratar a otros pacientes con su enfermedad, cosa que también le ruego que tome en consideración en su nombre. Si usted elige no participar en este ensayo, recibirá la terapia *A* o *B* que usted decida.” (Berry 1993, p. 1397)

Es decir, de modo implícito, se pone en manos de los pacientes la elección del tipo de justificación inferencial que podamos dar a un ensayo clínico. Ya no se trata de que acepten sin más el criterio de prueba consensuado por la comunidad científica, sino que, ante la existencia de alternativas, sus intereses particulares deciden. Estos intereses, como pone de manifiesto el último párrafo de la cláusula, pueden ser egoístas, pero ello no obsta para que prevalezca el principio de autonomía.

Por una parte, se le ofrece una elección real entre ambos tratamientos, cuando en el diseño convencional su única oportunidad de recibir el nuevo fármaco depende a menudo de que acepte participar en el ensayo y el azar se lo asigne. Por otro lado, el juicio particular del médico es parte explícita del consentimiento, lo cual corresponde

a la demanda de consejo personal por parte de los expertos al tomar tales decisiones. Obviamente, cláusulas como éstas probablemente supongan un descenso en el número de pacientes dispuestos a participar en un ECA. De ahí la insistencia bayesiana en reemplazarlos con análisis elaborados sobre bases de datos.⁹

En suma, cabe articular cada una de nuestras estrategias inferenciales con sendos argumentos éticos igualmente poderosos, pues hoy tenemos por paritarios los tres principios sobre los que éstos se apoyan: autonomía, equidad y no maleficencia. Para romper semejante indeterminación, creemos que conviene regresar a las bases sociales sobre las que se asienta la aceptabilidad de tales técnicas. Se trata de ponernos de acuerdo con respecto a qué reglas utilizaremos para establecer la eficacia de un medicamento, aun a sabiendas de que su aplicación resultará a menudo complicada, según las circunstancias éticas y estadísticas de cada caso. Pero para poder optar entre ambas técnicas debemos asegurar a sus detractores de que se observará algún principio general que todas las partes (médicos y pacientes) convengan en reconocer como indispensable para analizar los resultados de un ensayo clínico. Estamos en una situación, digámoslo así, *constituyente*.

Creemos que ese principio es el de *imparcialidad*, en la segunda de las acepciones que exponíamos al principio: puesto que debemos elegir si autorizamos un tratamiento, el procedimiento de decisión debe ser tal que *vele* al experimentador con respecto a sus intereses particulares; *i.e.*, el resultado tiene que ser independiente de sus intereses particulares, cualesquiera que éstos sean. El diseño de los ECA sirvió como garantía de imparcialidad y ello, como veíamos, resultó determinante en su aceptación por parte de los médicos y, en realidad, por parte de los propios pacientes, pues el Estado lo adoptó como criterio legal para establecer la eficacia de un medicamento. Creemos que este principio de imparcialidad sigue siendo hoy igualmente necesario.

De hecho, cabe afirmar que la imparcialidad médica ante un potencial conflicto de intereses es el reto más urgente de la investigación

⁹ Un evaluador apunta que la auténtica dimensión de la autonomía radica, en este contexto, en el consentimiento informado que se otorga al diseño estadístico del experimento. Ahora bien, cabe preguntarse qué tipo de información estadística es necesaria para dar un consentimiento autónomo, y si a este respecto son más eficaces las frecuencias o las asignaciones de probabilidad personal. Pero, sin perjuicio de las consideraciones expuestas en el texto, creemos que se trata de una cuestión que debiera dilucidarse experimentalmente, como se muestra en Gigerenzer 2003.

clínica de nuestro tiempo. La explosión comercial de las biotecnologías se está sintiendo también en el mundo académico, al multiplicarse las oportunidades de recibir patrocinio privado, o privatizar sin más la propia investigación patentando sus resultados. En el otoño de 2000, el diario *USA Today* publicó un informe sobre los expertos que tomaron parte en 18 comisiones consultivas del Centro para la Investigación y Evaluación del Medicamento de la FDA entre el 1 de enero de 1998 y el 30 de junio de 2000. Para poner sólo un ejemplo, el 54 por ciento de dichos expertos tenían intereses económicos directos en el fármaco sobre el que se requería su evaluación, bien por tener acciones en la compañía productora, cobrar en ella como consultores o recibir fondos para su propia investigación (Krimsky 2003, p. 96). Como sugiere Sheldon Krimsky, el dilema para la FDA radica probablemente en la dificultad de encontrar grandes especialistas exentos de compromisos empresariales.

6. ¿Son ambas técnicas igualmente imparciales?

Hoy son numerosos los ensayos clínicos subvencionados por empresas privadas en busca de evidencia a favor de sus productos. Kjaergard y Als-Nielsen (2002) analizaron 159 ECA de diversas especialidades médicas publicados en el *British Medical Journal* entre 1997 y 2001.¹⁰ De acuerdo con la declaración de potenciales conflictos de interés exigida para la publicación de cada artículo, en 65 casos se dejó constancia de su existencia: 27 por recibir patrocinio exclusivamente privado, 19 por recibir patrocinio mixto y 19 por otros motivos (personales, académicos, etc.). El análisis de Kjaergard y Als-Nielsen descubrió que “las conclusiones de los autores resultaban significativamente más positivas para la intervención experimental en los ensayos financiados exclusivamente por organizaciones con ánimo de lucro que en los ensayos sin intereses en conflicto” (2002, p. 251). Otros estudios obtuvieron conclusiones parecidas antes, aun cuando ninguno de ellos se pronuncie tajantemente en cuanto al origen del sesgo, si es que hay tal.¹¹

¿Dónde queda pues la imparcialidad de los ECA? ¿Constituye esto un argumento en defensa de la perspectiva bayesiana? Como vere-

¹⁰ En el *BMJ* es obligatorio que los autores declaren sus potenciales conflictos de interés: *cfr.* “Beyond Conflict of Interest”, *British Medical Journal*, 1998, vol. 325, p. 3.

¹¹ Puede ocurrir, en efecto, que entre los estudios que reciben subvención privada tiendan a publicarse exclusivamente los que obtienen conclusiones positivas para el tratamiento experimental.

mos, es posible sostener la siguiente respuesta. En cuanto a la generación de los datos, la aleatorización constituye la mejor garantía de su imparcialidad, con independencia de sus inconvenientes inferenciales. Con respecto a su interpretación, el enfoque bayesiano sólo será aceptable si alcanza una imparcialidad semejante a la que ahora atribuimos al contraste de hipótesis; esto es, si nos procuran una regla de interpretación de los datos que no dependa de una probabilidad *a priori* en particular.

Desde un punto de vista normativo, la aleatorización sigue siendo garantía de imparcialidad, aun cuando sea susceptible de las objeciones estadísticas que veíamos anteriormente. Para ello podemos apelar al siguiente argumento de Berry y Kadane (1997).

Supongamos que tenemos dos estadísticos, Daniel y Pedro, el primero de los cuales tendrá que diseñar un experimento y el segundo analizar los datos; tenemos también a un médico, Felipe, que asignará los pacientes a cada tratamiento de acuerdo con el diseño del experimento. Daniel quiere proporcionarle a Pedro los mejores datos para su análisis y debe establecer una regla de asignación para que Felipe (o el paciente, si le corresponde a él mismo elegir) los genere a partir de su diseño. Supongamos que Felipe tiene interés en que el resultado favorezca al tratamiento experimental, para lo cual, si tiene la ocasión, propenderá a asignárselo a los pacientes más sanos. Lo que Pedro debe descubrir es la probabilidad de que cada tratamiento tenga éxito con los grupos de población sano y enfermo. Por lo tanto, Felipe sesgará este valor cargando la probabilidad de que si un paciente está sano, reciba el tratamiento experimental, siempre y cuando se le conceda la oportunidad de asignarlo. Puede demostrarse que Daniel sólo podrá estar seguro de que el sesgo de Felipe no afectará el resultado si la probabilidad de que un paciente, sano o enfermo, reciba ambos tratamientos es igual. En otras palabras, si se aleatoriza la asignación. En esas condiciones, el estimador corresponderá insesgadamente al parámetro poblacional.

Lo interesante de este argumento es que es posible probarlo con independencia de nuestra concepción de la probabilidad; es decir, demuestra que si Daniel es un estadístico bayesiano, la aleatorización constituirá para él un procedimiento de asignación óptimo.¹² En otras palabras, con independencia del tipo de justificación inferencial que elijamos, la aleatorización constituye un garante de la imparcialidad

¹² Cosa que, en general, no sucedería si nuestro estadístico efectuase su estimación aislado. Sobre la pertinencia de este punto de vista *estratégico* en el análisis de la ciencia, *cfr.* Zamora Bonilla 2003.

de un ensayo clínico con respecto a los intereses del médico. Si la aleatorización de los ensayos clínicos examinados por Kjaergard y Als-Nielsen se efectuó correctamente, no tenemos mejor garantía, por ese lado, de su imparcialidad. Correlativamente, si se nos propone una interpretación bayesiana de los datos generados por un ensayo clínico, debemos exigir que la asignación de tratamientos se haya efectuado aleatoriamente.

Para comprobar el alcance normativo de esta exigencia, podemos pensar en la polémica que generaron los activistas antisida sobre los ECA. Tal como se evidencia en estudios como el de Steven Epstein (1996), los ECA promovidos por la administración estadounidense resultaron objeto de amplias críticas por parte de organizaciones como Act Up, deseosas de obtener un acceso rápido y generalizado a los tratamientos experimentales. Para ello, estos activistas socavaron tanto su justificación inferencial como ética. Por una parte, se cuestionaba en qué medida los ECA representaban a la población realmente afectada, sobre la base de análisis que mostraban cómo sus criterios de admisión resultaban demasiado restrictivos en su intento por controlar “demasiados” factores pronósticos y trabajar sobre una muestra homogénea que se pudiese aleatorizar. A cambio, los grupos de activistas promovieron diferentes tipos de ensayos no controlados, de interpretación controvertida según el paradigma vigente y beneficiarios potenciales de un análisis bayesiano. Por otro lado, los principios de equidad y no maleficencia que sostienen éticamente la asignación aleatoria de fármacos se vieron también impugnados. Ante la expectativa de una muerte segura, es inevitable restar importancia a la no maleficencia de un fármaco si ofrece alguna esperanza. Del mismo modo, resulta menos defendible la equidad de asignar al azar un placebo, por escasas que sean las probabilidades de éxito que se atribuyan al fármaco. Los activistas antisida reivindicaban precisamente su autonomía para asumir el riesgo y tener acceso a nuevos medicamentos desde las primeras fases de su ensayo —o a productos no autorizados en Estados Unidos—. Cabría decir, su derecho a obrar de acuerdo con su probabilidad personal.

¿Hemos de prescindir de la aleatorización en nuestros criterios de prueba ante demandas como éstas? Como defiende Epstein, citando a Bryan Winne, “extender el mito de que la producción de conocimiento es una empresa sometida a reglas no es la solución” (1996, p. 346), y deberíamos más bien tomar como criterio de prueba aquello que sea objeto de consenso entre todas las partes implicadas. Nosotros creemos, más bien, que puesto que las partes implicadas a menudo tienen intereses contrapuestos, la única garantía de que los

datos en que se apoye nuestro criterio de prueba (frecuentista o bayesiano) sean imparciales radica en la aleatorización. Una cosa es que, por razones prudenciales, las autoridades sanitarias accedan a las demandas de acceso a un tratamiento experimental por parte de un determinado colectivo (indiferente a la prueba de su efectividad) y otra que lo autoricen como seguro para el conjunto de la población.

Más allá del procedimiento de generación de los datos, y conviniendo en que la aleatorización es normativamente necesaria, cabe plantearse si es posible equiparar las reglas de interpretación clásicas actualmente vigentes a las que se nos proponen desde el bayesianismo. ¿Puede, por ejemplo, una empresa cuyos fármacos resulten descartados por los ECA buscar la aprobación mediante un ensayo no controlado interpretado desde un punto de vista bayesiano? Los partidarios de esta segunda alternativa lo niegan, pero hay quien piensa lo contrario. Whitehead (1993) nos propone, por ejemplo, el siguiente argumento a partir del análisis de los objetivos que se persiguen en la fase III de un ensayo clínico.

Para los organismos estatales se pretende, ante todo, negar la autorización a productos inefectivos o nocivos. Las farmacéuticas quieren probar la seguridad y la efectividad de sus medicamentos. El investigador intenta obtener información comparativa sobre los dos tratamientos sometidos a prueba. En una perspectiva frecuentista, estos tres objetivos suponen intereses estadísticos distintos: un nivel de significación pequeño, para los primeros; gran potencia del contraste, para los segundos; y estimación de intervalos de confianza, para los terceros. Pero la regla de interpretación que rige el contraste de hipótesis es idéntica para todas las partes y generará, por ello, un consenso. Desde el punto de vista de la probabilidad personal, dependemos del teorema de Bayes; el dilema es si las probabilidades *a priori* de cada una de las partes podrían converger para facilitar el acuerdo, suponiendo que, en general, serán distintas, según los intereses de cada parte.

La respuesta bayesiana se basa aquí en establecer un procedimiento para derivar una probabilidad *a priori* que describa, de modo imparcial, nuestra carencia de información *a priori* acerca de la magnitud que deseamos analizar, de modo que podamos evaluar su idoneidad maximizando la cantidad de información que vayamos obtener. Técnicas como las sugeridas por J.M. Bernardo (2003) pueden ser extraordinariamente útiles al respecto.¹³

¹³ Cfr. Spiegelhalter, Freedman y Parmar 1994, pp. 363–367, para un catálogo de alternativas en el contexto de los ensayos clínicos.

7. Conclusión

Hemos mostrado la existencia de diferentes tipos de argumentos sobre la metodología que debe adoptarse en los ensayos clínicos, destacando tres de ellos: su justificación inferencial, su aceptabilidad ética y su institución pública. Hemos visto cómo en los dos primeros planos es posible construir argumentos de contundencia parecida a favor de los dos enfoques que aquí presentamos, ensayos clínicos controlados mediante contrastes de significación y ensayos clínicos analizados en una perspectiva bayesiana. La cuestión es en qué medida resultan aceptables ambas perspectivas desde un punto de vista más general: con independencia del tipo de justificación por el que optemos, al investigador le interesa evaluar la eficacia de un medicamento; al médico y al paciente, en qué condiciones resulta éticamente aceptable el experimento. Pero al resto de los ciudadanos les interesará, en cambio, decidir qué tipo de pruebas pueden aceptar por válidas (para que el investigador experimente, y médicos y pacientes actúen según sus resultados). Dado que existen hoy distintos criterios de prueba: ¿qué se les puede exigir a tales ensayos para que el Estado acepte su uso? Nuestra respuesta es que se les debe exigir que garanticen su imparcialidad en relación con los intereses privados del investigador.

Contra quienes sostienen que la imparcialidad debe darse por descontada en el investigador, en virtud de sus compromisos con la veracidad científica y el bien público, creemos más bien que los intereses particulares son a menudo tan poderosos que resulta deseable una garantía metodológica en el propio procedimiento, la cual sirva, en alguna medida, como *criba*. Aun si en casos puntuales la aleatorización es discutible desde un punto de vista ético, constituye un criterio público de imparcialidad y debe exigirse, con independencia de su justificación probabilística inferencial, a cualquier técnica que se nos proponga para desarrollar ensayos clínicos. En cuanto a las reglas de interpretación de los datos, cabe esperar (y defender) la aceptación de criterios bayesianos en la medida en la que acierten a incorporar una criba semejante, introduciendo una probabilidad *a priori* que sea objeto de consenso en virtud de los mismos principios impersonales que rigen los contrastes de significación.¹⁴

¹⁴ Este artículo ha recibido el apoyo de los proyectos de investigación “Autorregulación y principio de justicia” (UNED) y “Aspectos del contrato social epistémico en las instituciones científicas” (BF2003-04372). Se benefició también de los comentarios planteados a versiones previas en el VII Seminario Internacional de Bioética (Universidad de Salamanca), y en distintas sesiones en la Universidad Complutense de Madrid, University College London y la Universidad de Valencia. Agradecemos

BIBLIOGRAFÍA

- Abajo, F. de, “Fundamentos de los ensayos clínicos”, en Carvajal 1993, pp. 83–105.
- Abajo, F. de, y D. Gracia, 1997, “Ética del uso del placebo en la investigación clínica”, *Investigación y Ciencia*, vol. 254, pp. 90–99.
- Armitage, P., 2003, “Fisher, Bradford Hill, and Randomization”, *International Journal of Epidemiology*, vol. 32, pp. 925–928.
- Bernardo, J.M., 2003, “Bayesian Statistics”, en R. Viertl (comp.), *Encyclopedia of Life Support Systems. Probability and Statistics*, UNESCO, Oxford.
- Berry, D.A., 1993, “A Case for Bayesianism in Clinical Trials [with discussion]”, *Statistics in Medicine*, vol. 12, nos. 15–16, pp. 1377–1404.
- Berry, S.M. y J.B. Kadane, 1997, “Optimal Bayesian Randomization”, *Journal of the Royal Statistical Society B*, vol. 59, no. 4, pp. 813–819.
- Carvajal, A. (comp.), 1993, *Farmacoepidemiología*, Universidad de Valladolid, Valladolid.
- Couzin, J., 2004, “The New Math of Clinical Trials”, *Science*, vol. 303, pp. 783–786.
- Cranor, C., 1994, “Conocimiento experto y políticas públicas en las sociedades tecnológicas. En busca del apoyo científico apropiado para la protección de la salud pública”, en J.L. Luján y J. Echeverría (comps.), *Gobernar los riesgos. Ciencia y valores en la sociedad del riesgo*, Organización de Estados Iberoamericanos/Biblioteca Nueva, Madrid, pp. 99–142.
- Epstein, Steven, 1996, *Impure Science. Aids and the Politics of Knowledge*, University of California Press, Berkeley/Los Ángeles.
- Fisher, R.A., 1971, *The Design of Experiments*, 8a. edición, Hafner, Nueva York.
- Freedman, B., 1987, “Equipose and the Ethics of Clinical Research”, *New England Journal of Medicine*, vol. 317, no. 3, pp. 141–145.
- Gigerenzer, G., 2003, *Reckoning with Risk*, Penguin, Londres.
- Gillies, D., 2000, *Philosophical Theories of Probability*, Routledge, Londres.
- Hill, A.B., 1952, “The Clinical Trial”, *New England Journal of Medicine*, vol. 247, pp. 113–119.
- Kjaergard, L., y B. Als-Nielsen, 2002, “Association Between Competing Interests and Author’s Conclusions: Epidemiological Study of Randomised Clinical Trials Published in the BMJ”, *British Medical Journal*, vol. 325, pp. 249–252.
- Krimsky, Sheldon, 2003, *Science in the Private Interest*, Rowman and Littlefield, Lanham.

especialmente las observaciones de F. de Abajo, V. Camps, D. Gillies, V. Iranzo, T. López de la Vieja, C. Velayos, J. Zamora y dos evaluadores anónimos.

- Levine, Robert J., 1988, *Ethics and Regulation of Clinical Research*, Yale University Press, New Haven/Londres.
- Marks, H.M., 2003, "Rigorous Uncertainty: Why RA Fisher Is Important", *International Journal of Epidemiology*, vol. 32, pp. 932–937.
- , 2000, "Trust and Mistrust in the Marketplace: Statistics and Clinical Research, 1945–1960", *History of Science*, vol. 38, pp. 343–355.
- , 1997, *The Progress of Experiment. Science and Therapeutic Reform in the United States, 1900–1990*, Cambridge University Press, Nueva York.
- Spiegelhalter, D., L. Freedman y M. Parmar, 1994, "Bayesian Approaches to Randomized Trials", *Journal of the Royal Statistical Society A*, vol. 157, no. 3, pp. 357–416.
- Urbach, P. 1993, "The Value of Randomization and Control in Clinical Trials [with discussion]", *Statistics in Medicine*, vol. 12, nos. 15–16, pp. 1421–1442.
- , 1985, "Randomization and the Design of Experiments", *Philosophy of Science*, vol. 52, no. 2, pp. 256–273.
- Whitehead, J., 1993, "The Case for Frequentism in Clinical Trials [with discussion]", *Statistics in Medicine*, vol. 12, nos. 15–16, pp. 1405–1420.
- Worrall, J., 2002, "What Evidence in Evidence-Based Medicine?", *Philosophy of Science*, vol. 69, pp. 316–S30.
- Zamora Bonilla, J., 2003, *La lonja del saber. Introducción a la economía del conocimiento científico*, UNED, Madrid.

Recibido el 31 de octubre de 2005; revisado el 19 de junio de 2006; aceptado el 5 de septiembre de 2006.