

Hechos, evidencia y estándares de prueba

Ensayos de epistemología jurídica

Andrés Páez
(coordinador)

Universidad de los Andes
Facultad de Derecho
Facultad de Ciencias Sociales
Departamento de Filosofía

Hechos, evidencia y estándares de prueba: ensayos de epistemología jurídica / Andrés Páez, coordinador.
– Bogotá: Universidad de los Andes, Facultad de Derecho, Facultad de Ciencias Sociales, Departamento de Filosofía, Ediciones Uniandes, 2015.

240 páginas; 17 x 24 cm

Otros autores: Danny Marrero, Amalia Amaya, Daniela Accatino, Eleonora Cresto, Pamela Lastres, César Higa, Simone Trento, Douglas Niño, Sandro R. D'Onofrio.

ISBN 978-958-774-216-9

1. Filosofía del derecho 2. Prueba (derecho) I. Páez, Andrés II. Universidad de los Andes (Colombia).
Facultad de Derecho. Facultad de Ciencias Sociales. Departamento de Filosofía

CDD 345.06

SBUA

Primera edición: noviembre del 2015

© Andrés Páez (coordinador)

© Universidad de los Andes, Facultad de Derecho, Facultad de Ciencias Sociales, Departamento de Filosofía

Ediciones Uniandes
Calle 19 n.º 3-10, oficina 1401
Bogotá, D. C., Colombia
Teléfono: 339 49 49, ext. 2133
<http://ediciones.uniandes.edu.co>
infeduni@uniandes.edu.co

Facultad de Derecho
Carrera 1.ª n.º 18A-12, Bloque RGC, piso 2
Bogotá, D. C., Colombia
Teléfono: 339 49 49, ext. 4860
<http://derecho.uniandes.edu.co/es/publicaciones>

Departamento de Filosofía
Publicaciones Facultad de Ciencias Sociales
Carrera 1.ª n.º 18A-12, Bloque G, piso 6
Bogotá, D. C., Colombia
Teléfono: 339 49 49, ext. 4819
<http://publicacionesfaciso.uniandes.edu.co>
publicacionesfaciso@uniandes.edu.co

ISBN: 978-958-774-216-9
ISBN *e-book*: 978-958-774-217-6
DOI: <http://dx.doi.org/10.7440/2015.63>

Corrección de estilo: Astrid Paola Molano
Diagramación interior: Karina Betancur Olmos
Diseño de cubierta: Victor Gómez
Imagen de cubierta: El experto Richard Souviron presenta pruebas dentales en el juicio de Ted Bundy, en junio de 1979.
Foto de Mark T. Foley, State Archives of Florida

Impresión:
Editorial Kimpres SAS
Calle 19 sur n.º 69C-17
Teléfono: 413 68 84
Bogotá, D. C., Colombia

Impreso en Colombia – *Printed in Colombia*

Todos los derechos reservados. Esta publicación no puede ser reproducida ni en su todo ni en sus partes, ni registrada en o transmitida por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea mecánico, fotoquímico, electrónico, magnético, electro-óptico, por fotocopia o cualquier otro, sin el permiso previo por escrito de la editorial.

5

Estándares múltiples de prueba en medicina y derecho*

ANDRÉS PÁEZ

I. Introducción

DAVID SCHUM (2001, CAP. 2) ha demostrado de manera muy elocuente las dificultades a las que se debe enfrentar quienquiera que asuma la tarea de ofrecer una definición aceptable del término *evidencia*¹. Una de las principales virtudes del bayesianismo es que proporciona una definición muy simple de qué es la evidencia. Todas las teorías bayesianas en la filosofía de la ciencia, tanto subjetivas como objetivas², comparten el siguiente principio: para que un hecho e sea evidencia a favor de una hipótesis h en un contexto epistémico K , es necesario y suficiente que e aumente la probabilidad de h . Formalmente,

e es evidencia a favor de h en K si y sólo si $p(h | K \& e) > p(h | K)$

* Este texto es uno de los resultados del proyecto *Umbrales de evidencia científica*, financiado por el Departamento Administrativo de Ciencia, Tecnología e Innovación de la República de Colombia (Colciencias).

1 A lo largo de este ensayo usaré el término *evidencia* para referirme a lo que los juristas llaman *pruebas* o *premisas fácticas*. La razón es que la gran mayoría de la bibliografía discutida aquí proviene de la filosofía de la ciencia, área en donde ese uso del término *prueba* es desconocido y podría generar confusión. En esos contextos, una *prueba* se refiere a una demostración formal, no a una pieza de evidencia.

2 Este principio se puede encontrar en las obras de bayesianos tan diferentes como Howson y Urbach (2006), que defienden una aproximación subjetiva, y Maher (1996), que defiende una versión objetiva del bayesianismo.

Para citar este artículo: <http://dx.doi.org/10.7440/2015.69>

Para que e sea evidencia *en contra* de h , sólo necesitamos reversar el símbolo de desigualdad. En su simplicidad, esta definición sufre de un defecto fatal. Aunque es innegable que la relevancia positiva es una condición *necesaria* para que un hecho constituya evidencia para una hipótesis³, es muy dudoso que sea una condición *suficiente*. La relevancia positiva en sí misma no logra capturar las connotaciones explicativas, justificativas e instrumentales que tiene la evidencia, tanto en contextos científicos como jurídicos (Páez, 2013). La evidencia es tanto un concepto epistemológico como un concepto práctico. Esta última dimensión queda completamente soslayada por la definición tradicional.

Mi tesis principal en este ensayo es que la evidencia es un concepto umbral con respecto a la probabilidad: debe haber una masa crítica de probabilidad, por así decirlo, para que un hecho se convierta en evidencia a favor de una hipótesis. Aunque en la literatura hay otras teorías de la evidencia que postulan la existencia de umbrales, aquí presento una teoría diferente, que he llamado la teoría de los umbrales múltiples. A partir de la manera en que la evidencia es, de hecho, utilizada en contextos científicos en general, y en contextos médicos y jurídicos en particular, mi propósito es defender la idea de que la evidencia viene en grados que corresponden a intervalos de valores de probabilidad para la hipótesis puesta a prueba, y que estos intervalos pueden ser determinados de una manera que no es arbitraria.

El ensayo se estructura de la siguiente manera. En la segunda sección discuto la necesidad de introducir umbrales para la evidencia y examino las diferentes formas en que los umbrales de probabilidad han sido entendidos en la literatura. En la tercera sección presento la teoría de umbrales múltiples. En las secciones cuarta y quinta muestro cómo se puede aplicar esta teoría en los procesos de decisión en la medicina y el derecho.

II. La insuficiencia de la relevancia positiva

Peter Achinstein (1983, 2001) ha utilizado contraejemplos similares a los siguientes para refutar la relevancia positiva como una condición suficiente para la evidencia. En todos ellos se asume que el lector no tiene información adicional acerca de la situación, más allá de la que está contenida en el ejemplo:

1. Cuando Michael Phelps entra a una piscina, aumenta su probabilidad de ahogarse; pero el hecho de que haya entrado a una piscina no es evidencia de que se ahogó.

³ Achinstein (2001) ha puesto en duda que la relevancia positiva sea una condición *necesaria*. Roush (2004) ofrece una respuesta a sus argumentos que, a mi juicio, es completamente satisfactoria. No discutiré este problema en este artículo.

2. Cuando compro un billete de lotería, aumento mi probabilidad de ganármela; pero el hecho de haber comprado un billete no es evidencia de que me la gané.
3. Cuando mi esposa sale a la calle, aumenta su probabilidad de ser atropellada por un Cadillac Eldorado; pero el hecho de que haya salido a la calle no es evidencia de que fue atropellada.

Aunque estos ejemplos son convincentes, es difícil decir por qué. Los hechos que aumentan la probabilidad, ciertamente, hacen posibles los hechos hipotéticos y, en ese sentido, son relevantes. Pero saber que algo es posible en sí mismo no conduce a un cambio en nuestra conducta o en nuestras decisiones acerca de los hechos involucrados. El razonamiento práctico requiere un mínimo de información relevante, que no es proporcionada por la simple posibilidad de un hecho.

Esto no significa que los hechos que aumentan la probabilidad no sean evidencia potencial para la confirmación de la hipótesis. La misma información e puede no ser evidencia a favor de h en un contexto, y comenzar a serlo en otro, aunque aumente la probabilidad de h en ambos. Tomemos el primer ejemplo. Puede haber otro contexto K^* en el que nos enteramos de que Michael Phelps entró a una piscina y en el que esa información es clara evidencia a favor de la hipótesis de que se ahogó. Supongamos que K^* incluye los siguientes hechos:

1. Michael Phelps es el único exmiembro del equipo olímpico de natación de los Estados Unidos que está en la ciudad.
2. La piscina sólo puede ser utilizada por exmiembros del equipo olímpico de natación de los Estados Unidos.
3. Hay un gran alboroto alrededor de la piscina.
4. Una ambulancia se acerca a gran velocidad con la sirena encendida.

Ahora bien, el hecho de que $p(h | e \& K^*) > p(h | e \& K)$ puede ser explicado, en parte, porque la probabilidad inicial de la hipótesis era mayor en K^* : $p(h | K^*) > p(h | K)$. Sin embargo, la razón principal por la que e es evidencia a favor de h en K^* es el efecto de e sobre la probabilidad de h . Supongamos que

$$p(h | e \& K) - p(h | K) = r$$

$$p(h | e \& K^*) - p(h | K^*) = r^*$$

y supongamos que

$$r = 0.01 - 0.001 = 0.099$$

$$r^* = 0.7 - 0.1 = 0.6$$

Como r^* es mucho mayor que r , hay algún número n tal que $r^* > n \geq r$, el cual sirve como umbral para que e sea evidencia. Este umbral establece la diferencia entre permanecer indiferente ante la información e y tomar la decisión de actuar cuando e hace a h lo suficientemente probable para algún propósito práctico.

Existen dos teorías del umbral en la literatura filosófica. La primera, que llamaré la teoría del *umbral absoluto*, establece que para que e cuente como evidencia a favor de h , e debe elevar la probabilidad de h desde un valor menor o igual a un umbral r hasta un valor que lo supere. Más precisamente, e es evidencia a favor de h si y sólo si $p(h) \leq r$ y $p(h | e) > r$. La segunda teoría, que llamaré la teoría del *umbral relativo*, establece que para que e cuente como evidencia a favor de h , e debe elevar la probabilidad de h en al menos un valor umbral r . En otras palabras, e es evidencia a favor de h si y sólo si $p(h | e) - p(h) > r$ ^[4].

Una teoría del umbral debe enfrentar varias posibles objeciones. La más obvia es que una hipótesis puede tener una probabilidad inicial tan alta que impida que cualquier información adicional se convierta en evidencia a su favor. De manera más precisa, en una teoría de umbral absoluto, si $p(h) > r$, nada nuevo cuenta como evidencia a favor de h . En una teoría de umbral relativo, si $p(h)$ es tan alta que $p(h | e) - p(h) < r$ para cualquier e , nada nuevo cuenta como evidencia a favor de h .

En respuesta a una objeción similar, formulada en el contexto del famoso problema bayesiano de la evidencia antigua, Colin Howson (1991) arguye que la solución es emprender un razonamiento contrafáctico. La probabilidad de h dado e debe ser relativizada, no a K , sino a una situación epistémica K^* en la cual $p(h) \leq r$, en la teoría del umbral absoluto, o a una en la que $p(h | e) - r > p(h)$, en la teoría del umbral relativo⁵. Aunque la solución de Howson restauraría las condiciones necesarias para que e pueda constituir evidencia a favor de h , es difícil ver cómo se debe implementar su propuesta. Se nos pide que eliminemos información del contexto epistémico original hasta que h alcance el grado de probabilidad deseado, pero no existe un algoritmo o una manera única de proceder, tal y como se puede apreciar al examinar la extensa literatura sobre contracciones en la teoría de revisión de creencias⁶. En la siguiente sección presentaré la que considero es la respuesta adecuada a esta objeción.

4 Achinstein (2001) ejemplifica la teoría del umbral absoluto, mientras la medida de relevancia D propuesta por Carnap (1962, p. xv) guarda alguna semejanza con la teoría del umbral relativo.

5 Asumiendo que $p(h)$ existe en K^* y que $p(h) \neq 0$.

Varios autores han formulado una objeción del todo diferente a las teorías de umbral. En su opinión, incluso si $p(h | e)$ o $p(h | e) - p(h)$ es inferior a cualquier umbral determinado, e aún tiene la connotación de “hacer [a h] más firme” (Carnap, 1962, p. xviii); “ e hace que sea más racional confiar en que h es verdadera” (Maher, 1996, p. 162); y e hace a h “más aceptable o mejor fundamentada de lo que sería en ausencia de esta evidencia” (Salmon, 1975, p. 5). La objeción puede ser enfrentada con una respuesta muy simple: la firmeza, la confianza, la aceptabilidad y la fundamentación, junto con conceptos como la justificación y la confirmación, también son conceptos umbral. La tesis según la cual la evidencia es un concepto umbral es de una sola pieza con la tesis de que todos éstos son conceptos umbral. Mi estrategia será, entonces, mostrar que el primero es un concepto umbral, para darle plausibilidad a la tesis de que los segundos también lo son.

Una objeción relacionada con la anterior es que la probabilidad asociada con $p(h | e)$ es una medida *continua* de grados de creencia (Carnap, 1962). Así, incluso un aumento muy pequeño en $p(h)$ corresponde a un aumento muy pequeño en nuestra creencia en h . Una vez más, la pregunta es si el concepto mismo de creencia es un concepto umbral con respecto a la probabilidad. Achinstein (2001, p. 77) afirma que lo es y yo concuerdo con su afirmación. Pero en ese caso la función de probabilidad no puede representar grados de creencia y es, entonces, necesario buscar una interpretación alternativa. Como veremos en breve, Achinstein ofrece una interpretación de la función de probabilidad que será inaceptable. En la siguiente sección proporcionaré una interpretación alternativa.

Debemos pasar ahora a considerar la pregunta acerca de cuál teoría es preferible: una teoría de umbral absoluto o una de umbral relativo. En lo que resta de la sección mostraré que ninguna de las dos opciones es satisfactoria en sí misma y en la siguiente sección presentaré una teoría de umbral muy diferente. Comenzaré con la versión de Achinstein de una teoría de umbral absoluto. En *The book of evidence*, el autor presenta una teoría de la confirmación que evita el subjetivismo del bayesianismo ortodoxo y, al mismo tiempo, ofrece una manera alternativa de entender qué es una creencia razonable. La teoría de Achinstein está basada en los siguientes dos supuestos (2001, pp. 115-116):

1. Para cualquier hipótesis h y evidencia putativa e , hay algún número k mayor o igual a cero tal que si e es una buena razón para creer h , entonces $p(h | e) > k$.

6 El problema es explorado en detalle por Rott y Pagnucco (1999). Para una aproximación alternativa, véase Levi (2004).

2. Para cualquier e y h , si e es una buena razón para creer h , entonces e no puede ser una buena razón para creer la negación de h ($\sim h$).

De este segundo supuesto se sigue que e es una buena razón para creer h sólo si $p(h | e) > \frac{1}{2}$. También, se sigue que la evidencia es un concepto absoluto: e es evidencia a favor de h sólo si e no es evidencia para la negación de h ^[7].

La aproximación de Achinstein a la evidencia y a la confirmación es muy problemática. Para comenzar, ésta no concuerda con el uso establecido del concepto de evidencia. Su teoría nos obligaría a descartar el siguiente caso: supongamos que hay evidencia en contra de dos (o más de dos) sospechosos de haber cometido un crimen que sólo pudo haber sido cometido por una persona. Es decir, existe alguna evidencia de que A lo cometió y alguna otra evidencia de que B lo cometió. Sin embargo, esta situación sería imposible si requerimos que, para contar como evidencia, los hechos del caso deben elevar la probabilidad de que cualquiera de los sospechosos haya cometido el crimen por encima del umbral de $\frac{1}{2}$. Una aproximación al concepto de evidencia que distorsiona nuestra comprensión común del concepto es muy cuestionable. Sin embargo, los problemas de la aproximación de Achinstein son más graves aún. Una de las consecuencias de su posición es que si $p(h | e) < \frac{1}{2}$, e sería una buena razón para creer $\sim h$. En el ejemplo anterior, la poca evidencia disponible nos daría una buena razón para creer en la inocencia de los dos sospechosos. Si bien debe haber una presunción de inocencia en todo proceso judicial, sería un error conceptual equipararla con la posesión de buenas razones para creer en la inocencia de los sospechosos.

Estas consecuencias indeseables de la propuesta de Achinstein se desprenden de su novedosa interpretación de la función de probabilidad. Achinstein rechaza tanto el subjetivismo del bayesianismo ortodoxo como el poco realista supuesto de interpretar las probabilidades como frecuencias relativas. Su interpretación alternativa es concebir las probabilidades como *probabilidades epistémicas objetivas*. Estas probabilidades miden el grado de razonabilidad de una creencia. Son objetivas debido a que están completamente soportadas por hechos físicos y no dependen en absoluto de las creencias o el conocimiento de alguien: “Para los subjetivistas un enunciado de probabilidad debe ser entendido como una probabilidad para *alguien*. En esta propuesta no hay tal relativización”. Estas probabilidades sólo miden qué tan razonable es la creencia en h a la luz de ciertos hechos, sin importar “qué tan razonable sea creer esto para alguna persona o grupo en particular” (pp. 98-99).

7 Para Achinstein que la probabilidad de la hipótesis sea mayor a un medio es una condición necesaria pero no suficiente para tener una creencia razonable. Para que e sea evidencia a favor

Achinstein dice que estas probabilidades son *epistémicas* porque representan grados de razonabilidad de creencias, en lugar de propensiones o frecuencias relativas. Pero este sentido de *epistémico* es tan débil que es difícil no estar de acuerdo con Roush cuando afirma: “Es fácil ver qué tiene de objetiva esta forma de entender la probabilidad, pero es difícil ver qué tiene de epistémica, aparte de las connotaciones de la palabra ‘razonable’” (2003, p. 204). Si hemos de considerar la noción de evidencia como una noción epistemológica, que en mi opinión es la forma correcta de entenderla, ser “una buena razón para creer *h*” en el sentido de Achinstein no puede ser lo mismo que ser evidencia a favor de *h*. En la próxima sección, esbozaré una aproximación a la probabilidad epistémica que hace honor a sus credenciales epistemológicas.

El rechazo de la teoría del umbral absoluto propuesta por Achinstein no significa que todas las teorías de umbral absoluto deban ser rechazadas. Es poco probable, sin embargo, que sea posible establecer *a priori* un umbral absoluto que no sea enteramente arbitrario. Lo mismo vale para los umbrales relativos. Es poco probable que haya un umbral *r* que sea apropiado para todos los posibles contextos donde se lleve a cabo la evaluación de una hipótesis. La aproximación que defiendo en este ensayo usará umbrales tanto en el sentido absoluto como en el relativo, pero intentaré evitar la arbitrariedad al permitir que sea la práctica científica y jurídica la que determine sus valores.

III. La teoría de umbrales múltiples

Diferentes ciencias y diferentes especialidades dentro de una misma ciencia utilizan estándares de prueba muy diferentes para tomar decisiones prácticas y teóricas. En las ciencias naturales es común utilizar umbrales estadísticos para describir cuánto soporte ha recibido una hipótesis a la luz de la evidencia disponible. Los físicos de CERN que anunciaron el descubrimiento del elusivo bosón de Higgs tuvieron que superar un umbral de significación estadística de 5 sigmas. Este umbral significa que si la partícula no existe, hay una probabilidad de 1 en 3.5 millones de que el experimento produzca un resultado que parezca confirmar su existencia. Previamente, con el fin de afirmar que había evidencia a favor de la existencia de la partícula, y emprender la tarea de construir el acelerador de partículas para buscarla, los físicos tuvieron que superar el umbral de 3 sigmas, que corresponde a una probabilidad de 1 en 741 de obtener un falso positivo. En medicina, en contraste, las pruebas clínicas sólo

de *h* también debe haber una conexión explicativa entre *e* y *h*. Aquí sólo me interesa el umbral de un medio como condición necesaria.

necesitan superar un umbral de 2 sigmas, que corresponde a una probabilidad de 1 en 20 de obtener un falso positivo. Los umbrales usados en el diagnóstico y tratamiento son mucho más flexibles, como veremos más adelante.

El uso de múltiples umbrales cualitativos es muy común en diferentes disciplinas, desde el derecho hasta la psicología. En los Estados Unidos, por ejemplo, la base jurídica para realizar una pesquisa depende del grado de sospecha justificado por la evidencia⁸, a la luz de la Cuarta Enmienda a la Constitución de los Estados Unidos. Se contemplan, así, los siguientes grados:

1. Corazonada o simple sospecha (insuficiente para una pesquisa).
2. Bases concretas para una *sospecha razonable* de que el sospecho cometió, está cometiendo o está a punto de cometer un crimen.
3. *Causa probable* para creer que el sospechoso es culpable.
4. Prueba de culpabilidad *más allá de una duda razonable*.

Muchas disciplinas usan escalas de Likert para clasificar hipótesis y afirmaciones. Una escala de Likert es una escala psicométrica usada comúnmente en cuestionarios. Al responder una pregunta en un cuestionario que utiliza estas escalas, las personas especifican su nivel de acuerdo o desacuerdo con una afirmación o su nivel de sospecha o de confianza acerca de la ocurrencia de un hecho. El tipo más común usa una escala de 5 puntos. Por ejemplo, para expresar el nivel de acuerdo o desacuerdo con una afirmación se usa la siguiente escala: totalmente en desacuerdo, en desacuerdo, ni de acuerdo ni en desacuerdo, de acuerdo, totalmente de acuerdo. También hay escalas de Likert de 3, 6, 7 y hasta 10 puntos. Una escala de Likert es usada, por ejemplo, para reportar casos de abuso infantil. Las leyes de los Estados Unidos obligan a los trabajadores de la salud a reportar cualquier caso sospechoso de abuso infantil. Usan un escala de Likert de 5 puntos para identificar su nivel de sospecha, con base en evidencia tal como laceraciones, quemaduras de cigarrillo, huesos rotos, comportamiento inusual. Otro ejemplo del uso de la escala de Likert es el diagnóstico del cáncer. Los radiólogos y patólogos, a partir de resultados radiológicos y patológicos, expresan independientemente su grado de sospecha de malignidad usando una escala de 4 puntos: no sospechoso, bajo, moderado y alto.

El uso de múltiples umbrales o de expresiones del grado de soporte que la evidencia le proporciona a una hipótesis es la base para la siguiente propuesta sobre la naturaleza de la evidencia. Para cada nivel de soporte proporcionado por la evidencia, y para cada contexto específico, es posible definir un intervalo de valores de probabilidad. Por ejemplo:

8 *Terry v. Ohio*, 392 U. S. 1 (1968).

e proporciona evidencia *muy débil* para h si y sólo si $p(h | e) \in (a, b]$

e proporciona evidencia *débil* para h si y sólo si $p(h | e) \in (b, c]$

e proporciona evidencia *moderada* para h si y sólo si $p(h | e) \in (d, e]$

e proporciona evidencia *fuerte* para h si y sólo si $p(h | e) \in (f, g]$

e proporciona evidencia *muy fuerte* para h si y sólo si $p(h | e) \in (h, i]$

En general, e proporciona evidencia de nivel X para h si y sólo si $p(h | e) \in [n, m]$ ⁹

Es claro que esto no es suficiente. Supongamos que e proporciona evidencia muy fuerte para h , porque $p(h | e) \in (n, m]$ y $(n, m]$ es el intervalo más alto. Ahora supongamos que $p(h) \in (n, m]$, es decir, la hipótesis ya estaba en el intervalo más alto posible. En tal caso, e por supuesto no logró mejorar el intervalo de probabilidad de la hipótesis. En aquellos casos en que la hipótesis permanece en el mismo intervalo, diré que e no proporciona evidencia alguna a favor de h , o simplemente, que e no es evidencia en absoluto.

Para que e proporcione evidencia de nivel X para h , debe elevar la probabilidad de h de un intervalo inferior al intervalo asociado con el nivel de evidencia X ¹⁰. Sea X el intervalo asociado con el nivel de evidencia X , en tal caso,

e es evidencia de nivel X para h si y sólo si

1. $p(h) \notin X$
2. $p(h | e) \in X$
3. $p(h | e) > p(h)$

Generalizando,

e es evidencia para h si y sólo si, para algún intervalo X ,
 e proporciona evidencia de nivel X para h .

Así, si $p(h | e) \notin X$ para el X más bajo, entonces e no es evidencia en absoluto. Este es el único sentido en el que es posible utilizar la noción de un umbral absoluto.

9 Los intervalos no tienen que cubrir todo el intervalo unidad $[0, 1]$, y generalmente no lo harán. No hay traslapes o vacíos entre los intervalos porque son continuos, abiertos a la izquierda y cerrados a la derecha.

10 Debemos establecer una distinción entre evidencia positiva y negativa, es decir, evidencia a favor y en contra de una hipótesis. Si e disminuye la probabilidad de h a un intervalo más bajo, es evidencia para la negación de h , y por lo tanto, califica como evidencia. En lo que sigue, me enfocaré en el concepto positivo de evidencia, pero cabe señalar que un análisis similar se aplica a la evidencia negativa.

Tras haber visto en qué consiste la teoría de los umbrales múltiples, ahora sí podemos responder a la objeción formulada en la sección anterior, según la cual en una teoría de umbrales, una hipótesis puede tener una probabilidad inicial tan alta que impide que cualquier información adicional sea evidencia a su favor. En lugar de emprender un análisis contrafáctico, *à la* Howson, mi respuesta es que estoy de acuerdo con la objeción: en tales casos e no sería evidencia en absoluto. En lugar de verla como una objeción, creo que esta situación refuerza la idea de que la evidencia es un concepto umbral. Una vez que una hipótesis ha alcanzado el estándar de prueba requerido para algún fin práctico o teórico, cualquier aumento ulterior de su probabilidad no hará ninguna diferencia.

Negarle a e el estatus de evidencia, ya sea porque no aumenta la probabilidad de h , al menos hasta el intervalo más bajo, o porque no logra cambiar su intervalo actual, no es equivalente a decir que la información que e proporciona carece de valor. El valor de la información proporcionada por e puede ser analizada utilizando una distinción introducida por Isaac Levi (1991, p. 82) entre el *contenido* informacional y el *valor* informacional de una proposición. Es claro que e tiene algún contenido informacional, puesto que $p(h | e) > p(h)$. Pero, su *valor* informacional está determinado por la diferencia práctica o teórica que marca en un contexto epistémico dado. Dada una hipótesis h , dos evidencias potenciales e_1 y e_2 y dos contextos epistémicos K_1 y K_2 , que $p(h | e_1 \& K_1) - p(h | K_1) = p(h | e_2 \& K_2) - p(h | K_2)$, sólo muestra que tienen el mismo contenido informacional, pero no necesariamente el mismo valor informacional. Los valores de $p(h | K_1)$ y $p(h | e_1 \& K_1)$ en el lado izquierdo de la ecuación podrían ser tales que $p(h | K_1) \notin X$ y $p(h | e_1 \& K_1) \in X$, mientras que los valores del lado derecho podrían ser tales que $p(h | K_2) \in X$ y $p(h | e_2 \& K_2) \in X$. Así, e_1 sería evidencia, pero e_2 no. El contenido informacional proveído por e_2 sería incorporado al contexto epistémico en el que se va a evaluar la siguiente evidencia potencial, y su principal contribución sería acercar a $p(h)$ al límite en el que comienza el siguiente intervalo de probabilidad.

Yo concibo las probabilidades en mi teoría como probabilidades epistémicas, en el siguiente sentido. Los valores de probabilidad representan la expresión numérica de un juicio a través del cual un sujeto expresa su valoración acerca de qué tan bien soportada está la hipótesis por la evidencia disponible, incluyendo el conocimiento de fondo. Las probabilidades epistémicas, y en esto concuerdo con Achinstein, no expresan grados de creencia. Sin embargo, diferentes intervalos de probabilidad representan los diferentes grados de soporte que una hipótesis debe tener para determinar decisiones prácticas a la luz de las creencias de un grupo o comunidad epistémica. En lugar de medir grados de razonabilidad de creencias, como Achinstein propone, se podría decir que las probabilidades epistémicas miden la razonabilidad de las decisiones prácticas y teóricas tomadas a partir del grado de soporte que tenga la

hipótesis involucrada en dichas decisiones. Volviendo al ejemplo del bosón de Higgs, la comunidad de físicos estuvo de acuerdo en que superar el umbral de 3 sigmas era todo el soporte epistémico necesario para invertir 4000 millones de dólares en la construcción del Gran Colisionador de Hadrones. No obstante, para poder anunciar la existencia del bosón de Higgs era necesario superar el umbral de 5 sigmas. Ejemplos menos dramáticos abundan en las ciencias sociales y naturales. En las siguientes dos secciones presentaré ejemplos de umbrales cuantitativos en medicina y derecho. Mi propósito es mostrar que la teoría de los umbrales múltiples puede acomodar naturalmente estos y otros ejemplos, y que los umbrales o estándares de prueba pueden ser determinados de una manera no arbitraria.

IV. Umbrales cuantitativos en medicina

Aunque los umbrales cualitativos descritos en la sección anterior cumplen el propósito para el que fueron creados, en muchos contextos epistémicos es deseable tener una medida más precisa del nivel de confirmación de una hipótesis. En medicina, varios estudios indican que a la mayoría de las personas les cuesta mucho trabajo asignar valores numéricos a las expresiones cualitativas de probabilidad¹¹. Los filósofos del derecho se han encontrado con un problema similar cuando han tratado de implementar métodos bayesianos en la corte, como lo ilustra el bien conocido caso de *R v Adams*¹².

Durante las últimas décadas, la medicina basada en evidencia (MBE) ha surgido como una nueva aproximación a la toma de decisiones en el ámbito clínico. Brevemente, la MBE usa herramientas matemáticas e investigación de alta calidad para calcular los costos, riesgos y beneficios del diagnóstico y manejo de pacientes individuales (Greenhalgh, 2010). La MBE reconoce que muchas decisiones médicas dependen de factores que son difíciles de cuantificar como los aspectos éticos o aquellos que tienen que ver con la calidad de vida de los pacientes. Así, el tratamiento en cada caso depende en gran medida de las preferencias del paciente.

El punto de partida de la MBE es la obtención sistemática y la evaluación crítica de la mejor evidencia disponible. La evidencia más fuerte para una intervención terapéutica es proporcionada por la revisión sistemática de pruebas clínicas aleatorias, de triple ciego, controladas por placebo en una población homogénea de pacientes con la misma condición médica. En contraste, los

11 Véase Bryant y Norman (1980), Shaw y Dear (1990) y Ohnishi *et al.* (2002).

12 *R v Denis Adams* [1996] 2 Cr App R 467.

testimonios de pacientes, los estudios de caso e incluso la opinión de expertos tienen un valor muy limitado, debido al efecto placebo, los sesgos inherentes a la observación y el reporte de casos, las dificultades en la atribución de experticia, entre otros problemas. Tan importante como la correcta evaluación de la calidad de la evidencia es el cálculo correcto de la probabilidad inicial de una enfermedad, antes de que se haya llevado a cabo prueba alguna. La probabilidad preprueba de una enfermedad, denotada $p(h)$, se establece con criterios tales como:

1. Estadísticas regionales o nacionales de la prevalencia de la enfermedad en cuestión en la población en general o en algún subconjunto relevante de ella.
2. Experiencia clínica.
3. Bases de datos sobre la práctica médica.
4. Probabilidades preprueba usadas en el estudio que soporta la prueba diagnóstica apropiada para la enfermedad.
5. Reportes de investigación sobre casos similares.

La tercera piedra angular de la MBE es la evaluación correcta de las pruebas diagnósticas. La información de estas es reportada en términos de sensibilidad, especificidad y valor predictivo. Infortunadamente, estas medidas tienen graves limitaciones. Por ejemplo, sólo pueden ser usadas en pruebas dicotómicas, en las que sólo se pueden obtener cuatro resultados: verdadero positivo, verdadero negativo, falso positivo y falso negativo. En estas pruebas, la sensibilidad se define como la proporción de pacientes con la enfermedad que tienen un resultado positivo (verdaderos positivos). La especificidad es la proporción de pacientes sin la enfermedad que tienen un resultado negativo (verdaderos negativos). El valor predictivo puede ser positivo o negativo. El valor predictivo positivo de una prueba es la proporción de pacientes con un resultado positivo que tienen la enfermedad, mientras el valor predictivo negativo es la proporción de pacientes con un resultado negativo que no tienen la enfermedad. A menudo, la información de las pruebas se expresa en términos de coeficientes de probabilidad (*likelihood ratios*). Éstos tienen la ventaja de ser derivables de pruebas diagnósticas con múltiples niveles o categorías de resultados. Además, los coeficientes de probabilidad de pruebas diferentes e independientes se pueden usar de manera secuencial (Friedland *et al.*, 1998, p. 23).

El proceso de decisión en la MBE sigue esencialmente el método tradicional bayesiano. Para determinar la probabilidad de la enfermedad (E+), dado el resultado positivo de una prueba diagnóstica (R+), se utiliza la siguiente fórmula (Hunink & Glasziou, 2001, p. 142):

$$p(E+ | R+) = \frac{p(R+ | E+) p(E+)}{p(R+ | E+) p(E+) + p(R+ | E-) p(E-)}$$

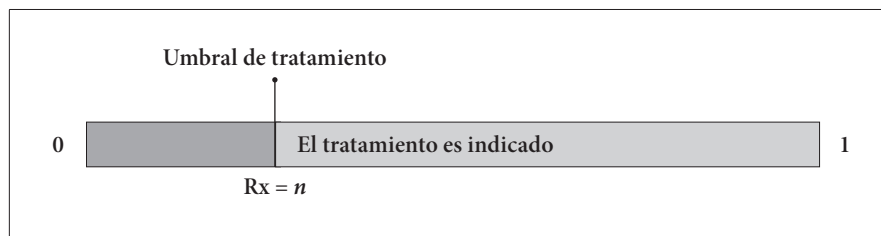
La ecuación también puede ser expresada en términos de la sensibilidad y la especificidad de la prueba diagnóstica:

$$p(E+ | R+) = \frac{\text{sensibilidad} \times \text{probabilidad preprueba}}{\text{sensibilidad} \times \text{probabilidad preprueba} + (1 - \text{especificidad}) \times (1 - \text{probabilidad preprueba})}$$

Esta expresión se puede generalizar para múltiples resultados de pruebas diagnósticas y para múltiples enfermedades.

Con este marco teórico, podemos pasar a considerar dos tipos de umbrales cuantitativos en medicina. El primero es el umbral de tratamiento. Un umbral de tratamiento (R_x) es la probabilidad por encima de la cual se debe comenzar el tratamiento de la enfermedad. Si la probabilidad de que el paciente tenga la enfermedad es mayor que el umbral de tratamiento, éste es indicado.

Figura 1



Los umbrales de tratamiento están determinados por los costos y beneficios de cada tratamiento particular. El uso de antibióticos para una laringitis estreptocócica, por ejemplo, tienen un umbral de tratamiento muy bajo. El tratamiento es generalmente con penicilina v, que es barata, segura y efectiva. El uso del antibiótico reduce la duración promedio de los síntomas de 3 a 5 días a un día y reduce la probabilidad de contagio. La quimioterapia para el cáncer, por otra parte, tiene un umbral de tratamiento muy alto. Dados los resultados inciertos y los efectos adversos de la quimioterapia, como la fatiga, la depresión del sistema inmunológico, las molestias gastrointestinales, la pérdida de pelo, la infertilidad y el posible daño del corazón, hígado, riñones, oídos y el cerebro, se requiere evidencia muy fuerte de que el paciente tiene cáncer.

La mayoría de casos no son tan simples¹³. Determinar el umbral de tratamiento en muchos casos depende de cómo se evalúen los costos y beneficios y por quién (Tsevat *et al.*, 1994). Se deben tener en cuenta tanto la perspectiva individual como la social. Cada una requiere un tipo diferente de evaluación. Decisiones sobre asuntos referentes a la calidad de vida de un paciente serán muy diferentes a las decisiones acerca de cómo utilizar los limitados recursos públicos destinados a la salud. Más aún, múltiples investigaciones empíricas muestran que las evaluaciones individuales de costos y beneficios son a menudo inconsistentes con los supuestos de la teoría de la utilidad esperada (Kahneman & Tversky, 1979). Más adelante volveré sobre los problemas involucrados en la determinación de los umbrales de tratamiento, cuando examine el problema de los estándares de prueba en el derecho.

Entonces, ¿cómo se relaciona la teoría de los umbrales múltiples con los umbrales de tratamiento en medicina? Consideremos el siguiente ejemplo. McIsaac *et al.* (2000) desarrollaron un protocolo muy simple para determinar rápidamente si un paciente que llega a la sala de urgencias con dolor de garganta tiene una laringitis estreptocócica¹⁴. Supongamos que K es el conocimiento de fondo, que incluye la información sobre el dolor de garganta del paciente, y h la hipótesis para la cual buscamos evidencia:

K : el paciente tiene dolor de garganta.

h : el paciente tiene una faringitis estreptocócica

La siguiente información constituye evidencia potencial para la hipótesis:

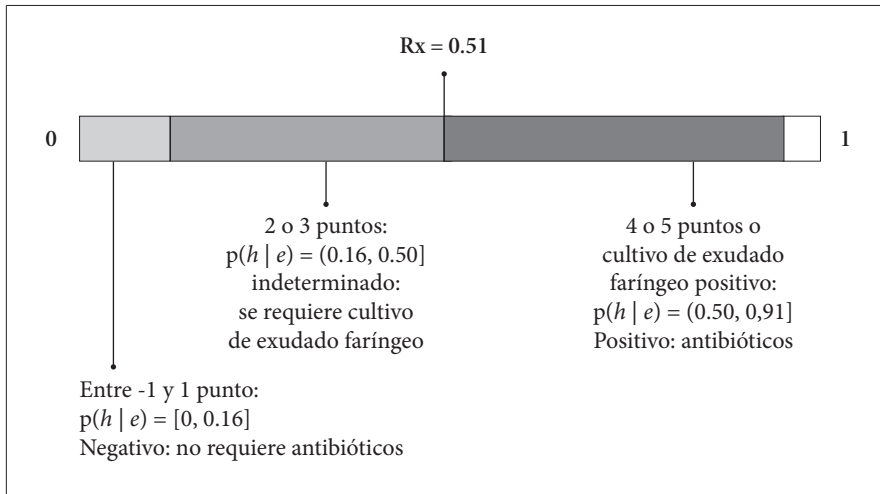
1. Temperatura mayor a 38°C (100.4°F).
2. Ausencia de tos.
3. Adenopatía cervical anterior.
4. Amígdalas rojas e inflamadas.
5. Edad inferior a 15 años.

Cada evidencia potencial vale un punto. Se debe restar un punto si el paciente es mayor de 45. Con esta información presente, los siguientes umbrales son utilizados para determinar el tratamiento apropiado:

¹³ Véase Hunink y Glasziou (2001, pp. 161 y ss.) para un análisis detallado de umbrales de tratamiento más complejos.

¹⁴ Protocolos como éste son muy comunes en las salas de urgencias. Existen dos tipos de protocolos: aquellos que agilizan y aquellos que estandarizan el cuidado de los pacientes. Las guías para el tratamiento de enfermedades también son muy comunes. Véase, por ejemplo, la guía de

Figura 2



Examinemos este ejemplo a la luz de la teoría de los umbrales múltiples presentada en la sección anterior. Consideremos la siguiente situación:

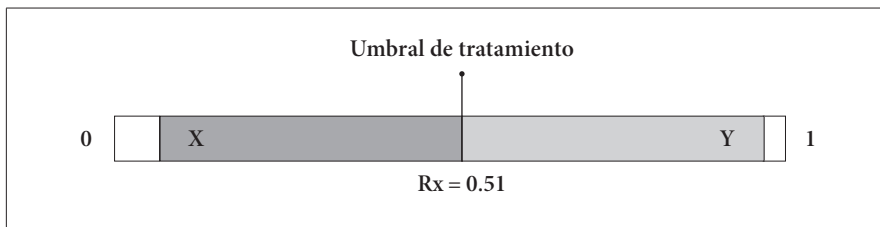
K : el paciente tiene fiebre y amígdalas inflamadas (2 puntos).

h : el paciente tiene una faringitis estreptocócica.

e_i : el cultivo de exudado faríngeo dio positivo para una faringitis estreptocócica.

$$X = p(h) = 0.17 \quad Rx = 0.51 \quad Y = p(h | e_i) = 0.91$$

Figura 3



Como e_i aumentó la probabilidad de la hipótesis al siguiente intervalo de probabilidad, conforme a la teoría de los umbrales múltiples, e_i es evidencia para h .

Jackson (2000) para manejar el riesgo cardiovascular, la cual incluye seis factores de riesgo: edad, colesterol, presión arterial, tabaquismo, diabetes y sexo.

Ahora consideremos un caso en el que el conocimiento de fondo es exactamente el mismo, pero donde e corresponde a un dato diferente:

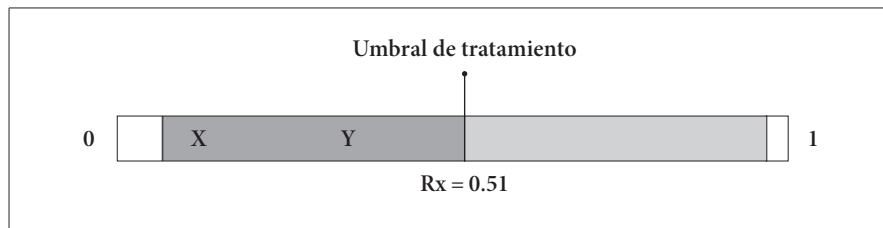
K : el paciente tiene fiebre y amígdalas inflamadas (2 puntos).

h : el paciente tiene una faringitis estreptocócica.

e_2 : el paciente tiene menos de 15 años.

$$X = p(h) = 0.17 \quad R_x = 0.51 \quad Y = p(h | e_2) = 0.35$$

Figura 4



Como e_2 no logró aumentar la probabilidad de la hipótesis al siguiente intervalo de probabilidad, según la teoría, e_2 no es evidencia para h .

Por último, cambiemos el conocimiento de fondo y mantengamos la misma evidencia potencial e_2 que en el ejemplo anterior:

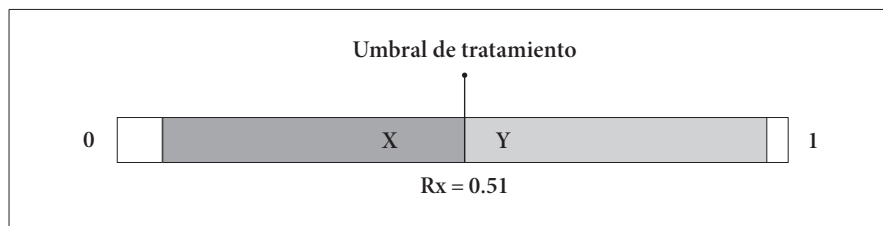
K : el paciente tiene fiebre, amígdalas inflamadas y ausencia de tos (3 puntos).

h : el paciente tiene una faringitis estreptocócica.

e_2 : el paciente tiene menos de 15 años.

$$X = p(h) = 0.35 \quad R_x = 0.51 \quad Y = p(h | e_2) = 0.53$$

Figura 5



Aunque el aumento de probabilidad, 0.18, es exactamente el mismo que en el ejemplo anterior, e_2 ahora sí puede ser considerada evidencia para h , porque

mueve la probabilidad de la hipótesis al siguiente intervalo. El aumento de probabilidad es suficiente para cambiar al menos una decisión acerca del tratamiento del paciente. La diferencia entre los casos representados en las figuras 5 y 6 ilustra por qué una teoría de umbral relativo por sí sola no es suficiente para capturar la relevancia de e_2 en cada caso.

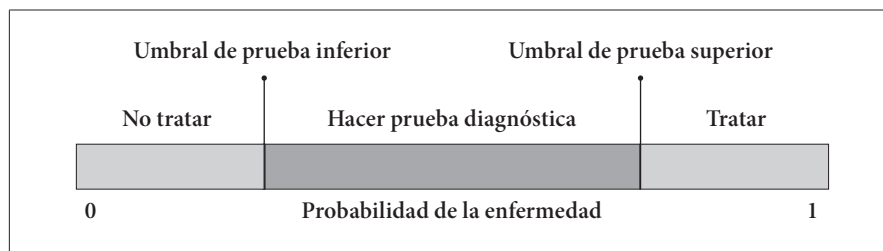
Otro umbral importante en medicina es el *umbral de prueba diagnóstica*, que no debe ser confundido con los umbrales de prueba usados en derecho. La decisión de realizar una prueba diagnóstica está determinada por consideraciones prácticas. Según Hunink y Glasziou,

Llevar a cabo una prueba diagnóstica para obtener información adicional vale la pena sólo si al menos una decisión cambiaría con base en los resultados de la prueba, y si los riesgos para el paciente asociados a la prueba son menores que el beneficio esperado que se obtendría del cambio en esa decisión. (2001, p. 80)

Un cambio en una decisión sólo ocurriría, a su vez, si la probabilidad de la hipótesis que afirma que el paciente sufre de la enfermedad cambia de un intervalo de probabilidad a otro, según los umbrales de tratamiento establecidos en los protocolos y las guías para esa enfermedad. “El umbral de tratamiento es el pivote alrededor del cual giran las pruebas diagnósticas. Hacer una prueba diagnóstica puede ser visto como un intento de situar a un paciente claramente de un lado u otro del umbral de tratamiento” (p. 161).

Uno de los principales problemas implicados en las pruebas diagnósticas es la posibilidad de obtener falsos positivos y falsos negativos. Si $p(h)$, la probabilidad inicial de la enfermedad, es muy baja, un falso positivo no convencerá al médico de que debe tratar al paciente; y si $p(h)$ es muy alta, un falso negativo no convencerá al médico de que no debe tratar al paciente. Las pruebas diagnósticas, por lo tanto, sólo son útiles cuando la probabilidad de la hipótesis está en la zona gris de las probabilidades intermedias, y más aún cuando están muy cerca de los umbrales de tratamiento. La siguiente figura resume estas posibilidades:

Figura 6



Los umbrales de prueba diagnóstica se calculan con base en las utilidades esperadas de las pruebas y el tratamiento. El umbral de prueba inferior es la probabilidad de h , en la que estamos en un estado de indiferencia entre no tratar y hacer la prueba, porque sus utilidades esperadas son iguales; el umbral superior es la probabilidad de h , en la que estamos en un estado de indiferencia entre hacer la prueba y tratar, porque sus utilidades esperadas son iguales. Los umbrales de prueba diagnóstica ilustran una vez más la idea de que no todo hecho relevante desde el punto de vista estadístico califica como evidencia. Una prueba diagnóstica que no conduce a ningún cambio de decisión, porque la probabilidad inicial de la hipótesis es muy baja o muy alta, simplemente no es evidencia a favor o en contra de esa hipótesis, incluso si la sensibilidad y especificidad de la prueba son conocidas a cabalidad. Ese es exactamente el resultado deseado, a la luz de la teoría de los umbrales múltiples y porque queremos que nuestro concepto de evidencia incluya falsos positivos y falsos negativos.

Los dos tipos de umbrales discutidos en esta sección ilustran cómo la teoría de los umbrales múltiples captura un elemento esencial de la práctica científica. En la siguiente sección, esbozaré cómo puede también darse una transformación radical en los procesos de decisión en el derecho.

V. Estándares múltiples de prueba en el derecho

La relevancia probabilística positiva y negativa se utiliza en el derecho no para definir qué es prueba o evidencia¹⁵, sino para definir qué es una prueba *relevante*. La regla 401 de las *Federal Rules of Evidence*, por ejemplo, afirma:

La prueba es relevante si:

- (a) tiene la tendencia a hacer que un hecho sea más o menos probable de lo que sería sin la prueba; y
- (b) el hecho es importante en la determinación de la acción.

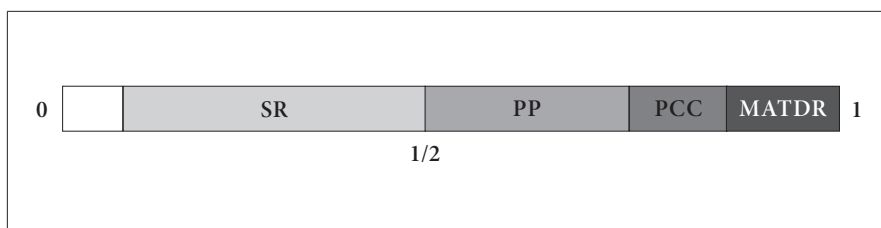
En la literatura de la epistemología jurídica encontramos definiciones similares de *prueba relevante*. Laudan, por ejemplo, dice que “si una cierta pieza de información, incluso si fuera creíble, no alteraría nuestra confianza a favor o en contra de una hipótesis, la consideramos irrelevante para la hipótesis” (2006, p. 17). Aquí la probabilidad involucrada parece medir grados de confianza en la

15 En esta sección me referiré a la evidencia como *prueba*, siguiendo el uso tradicional del término en el derecho.

hipótesis, según el cálculo de probabilidades. El problema con estas definiciones de prueba relevante es que tratan cualquier hecho que aumente la probabilidad de la hipótesis como prueba. Esa es justamente la tesis que he venido rechazando a lo largo de estas páginas. Podríamos adaptar fácilmente al contexto jurídico los contraejemplos planteados al comienzo del ensayo. Consideremos la afirmación de Laudan según la cual la declaración de doscientos testigos que corroboren lo que dos o tres testigos confiables ya han dicho es un ejemplo de una prueba relevante pero redundante (2006, p. 19n). Es difícil ver cómo dicho testimonio sería relevante a la luz de su propia definición de relevancia. Bajo las circunstancias descritas, la adición de docenas de testigos no alteraría nuestra confianza en la hipótesis bajo consideración. Más aún, es difícil ver por qué la declaración de los doscientos testigos es una prueba en absoluto. Es, precisamente, porque la confianza es un concepto umbral, que también debemos usar una teoría de umbrales para la prueba.

Una reformulación de la teoría de umbrales múltiples es facilitada por la existencia de múltiples estándares de prueba en el derecho. La *sospecha razonable* (SR), la *preponderancia de la prueba* (PP), la *prueba clara y convincente* (PCC) y la *prueba más allá de toda duda razonable* (MATDR) son los más comunes en el derecho anglosajón, pero hay muchos otros estándares que son utilizados en contextos específicos y en diferentes países¹⁶. La figura 7 presenta los estándares de prueba que acabo de mencionar, que serán utilizados aquí sólo con fines de ilustración.

Figura 7



¹⁶ Otros estándares usados en los Estados Unidos, por ejemplo, incluyen *razonable de creer*, que se aplica a las pesquisas de automóviles, y *alguna prueba creíble*, que se usa en el derecho administrativo. Por otra parte, algunos autores han expresado su insatisfacción con la existencia de sólo un estándar de prueba en el derecho penal. Laudan, por ejemplo, afirma: “Quiero preguntar (sin resolver definitivamente el asunto) si tiene sentido utilizar el mismo estándar —ya sea MATDR o algún otro— para *todos* los crímenes. En materia de estándares de prueba, no estoy convencido de que un mismo estándar sea adecuado para todos” (2006, p. 55). Cuando se trata de castigar un crimen, se acepta universalmente la idea de que el castigo debe ser proporcional al crimen. ¿Por qué no defender la idea, se pregunta Laudan, de que el estándar de prueba también deba ser proporcional al crimen? Volveré sobre este problema al final del ensayo.

Consideremos un ejemplo de aplicación de la teoría de umbrales múltiples. En un juicio de paternidad encontramos la siguiente situación:

K : la demandante ha testificado que tuvo relaciones sexuales con el demandado durante el periodo probable de concepción del menor.

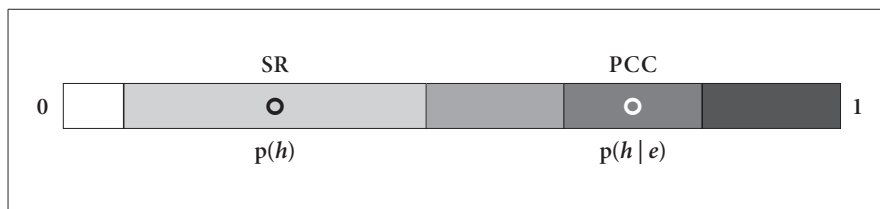
h : el demandado es el padre del menor.

e_1 : prueba de ADN positiva.

$$p(h) = 0.25 \quad p(h | e_1) = 0.8$$

Asumiremos que el estándar de prueba para un juicio de paternidad es la prueba clara y convincente, como es el caso en muchos estados de los Estados Unidos.

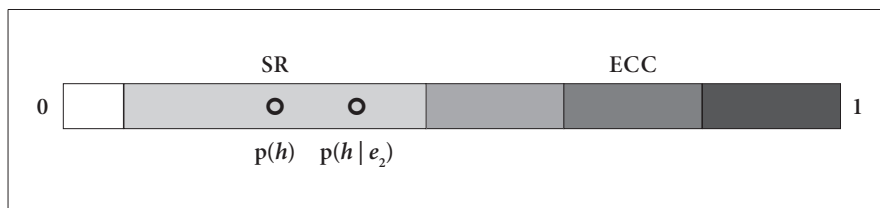
Figura 8



Como $p(h) \notin \text{CCE}$, $p(h | e_1) \in \text{CCE}$ y $p(h | e_1) > p(h)$, podemos concluir que e_1 es una prueba a favor de h .

Si cambiamos el ejemplo y en lugar de una prueba de ADN consideramos e_2 , el testimonio de un testigo confiable que declara que el demandado y la demandante eran novios durante el tiempo probable de concepción del menor, obtenemos el siguiente escenario:

Figura 9



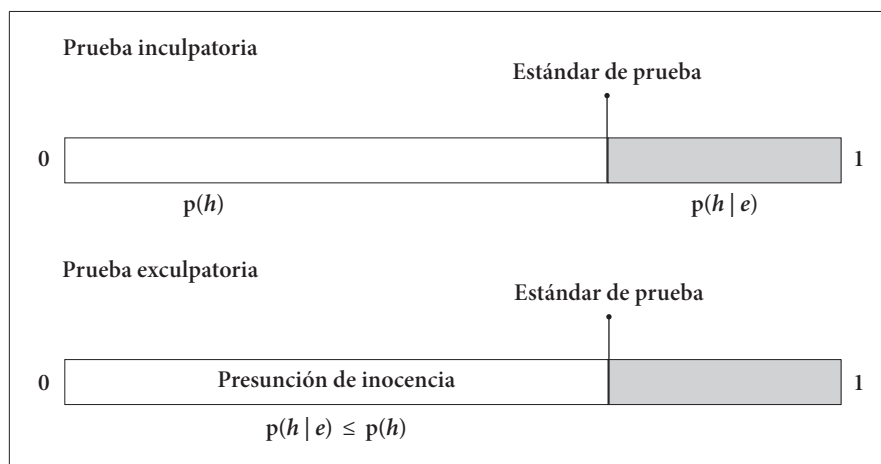
Como e_2 no cambia el intervalo de probabilidad de h al siguiente nivel, e_2 no es una prueba a favor de h . Por supuesto, la declaración de la testigo se puede añadir al conocimiento de fondo a la luz del cual será evaluada la siguiente pieza

de información. De esta manera, será más fácil para la parte demandante superar el estándar de prueba. Pero, la prueba potencial siempre es evaluada paso a paso, con base en las condiciones epistémicas que cambian constantemente.

Hay dos problemas potenciales para la utilización de la teoría de umbrales múltiples en contextos jurídicos, problemas que no surgen en el caso médico. El primer problema reside en una característica esencial del sistema adversarial que predomina en los países del *common law*, y que es utilizado cada vez con más frecuencia en América Latina. Es fácil ver cómo se pueden aplicar los criterios establecidos para catalogar un hecho introducido por la fiscalía como prueba en la teoría de umbrales múltiples. Como el propósito de la fiscalía es aumentar la probabilidad de la hipótesis de culpabilidad, más allá de un estándar de prueba preestablecido, cualquier hecho que lo consiga califica como prueba. Sin embargo, es difícil ver cómo se podría aplicar la teoría de umbrales múltiples a los hechos presentados por la defensa. Como el propósito de presentar estos hechos es mantener la probabilidad de la hipótesis por debajo del estándar de prueba, en el intervalo asignado a la presunción de inocencia, tales hechos nunca calificarían como pruebas, lo cual es absurdo.

La solución es establecer una diferencia entre pruebas inculpatorias y exculpatorias. La prueba inculpatoria es el concepto primitivo en el derecho; la prueba exculpatoria, su contrario. Si el propósito de la prueba inculpatoria es lograr cambiar el intervalo inicial de la hipótesis a uno superior, el de la prueba exculpatoria es mantenerlo dentro de algún intervalo inferior al estándar de prueba. La teoría de los umbrales múltiples es, así, una teoría de la prueba inculpatoria, y por extensión, una teoría de su concepto opuesto. La siguiente figura ilustra la diferencia entre ambos tipos de pruebas:

Figura 10



El segundo desafío al que se enfrenta la teoría de los umbrales múltiples en el contexto jurídico es la determinación del valor de los múltiples umbrales o estándares de prueba¹⁷. Como vimos en la sección anterior, en medicina tanto el umbral de tratamiento como el de prueba diagnóstica son calculados de acuerdo con las utilidades esperadas del diagnóstico y el tratamiento, lo que genera un sistema de umbrales muy flexible. A continuación argüiré que el análisis de estos dos umbrales puede ser replicado en el caso del derecho. En particular, el umbral de prueba diagnóstica utilizado en medicina puede servir como base para el análisis de los juicios de admisibilidad usados para determinar los hechos de un caso, y el umbral de tratamiento puede arrojar alguna luz sobre el problema de cómo determinar los estándares de prueba para decidir la culpabilidad del imputado.

Tanto en el caso jurídico como en el médico, los conceptos clave son los beneficios y daños involucrados en el proceso de toma de decisión. Recordemos que los umbrales de tratamiento están parcialmente determinados por la evaluación formal de los diferentes atributos de los resultados de los tratamientos. Estas evaluaciones se obtienen a partir de la valoración de la utilidad, que es un método para medir los niveles de discapacidad o inconveniencia que la gente asocia con resultados específicos del tratamiento. Pero la toma de decisiones en medicina va más allá de la perspectiva individual. Generalmente, se distinguen tres niveles: (1) el nivel micro, que implica decisiones acerca de pacientes específicos según sus circunstancias personales y sociodemográficas; (2) el nivel meso, en el que se establecen directrices intrahospitalarias para grupos de pacientes con la misma enfermedad; y (3) el nivel macro o social, en el que se toman las decisiones que involucran políticas públicas y el uso de los limitados recursos públicos asignados a la salud (Stiggelbout, 2000, p. 290). Aunque sólo el tercer nivel se refiere de manera directa a los aspectos económicos de la decisión, las limitaciones presupuestales de las instituciones y las personas han llevado este tipo de consideraciones a los otros dos niveles. Como la utilidad de un tratamiento y de su resultado están determinados por las características particulares de una enfermedad, habrá una enorme variación en los umbrales de tratamiento de una enfermedad a otra. Además, con respecto a una enfermedad dada habrá valoraciones de utilidad en conflicto en los diferentes niveles.

17 Varios autores (Tribe, 1971; Cohen, 1977; Allen, 1986; Haack, 2013) han argüido que no es factible ni deseable entender las probabilidades mencionadas en las definiciones jurídicas de la prueba como probabilidades matemáticas. Aceptar sus argumentos implicaría negar que sea posible construir un marco probabilístico para la teoría de los umbrales múltiples en el contexto jurídico. La pregunta acerca del papel de las probabilidades matemáticas en el derecho ha sido materia de amplio debate y es imposible tratar a fondo el asunto en este ensayo. Los problemas que he venido discutiendo aquí surgen en el contexto de las teorías probabilísticas de la evidencia y continuaré la discusión dentro de ese marco.

Por ejemplo, un tratamiento que es altamente preferible en el nivel micro puede tener una gran desutilidad en el nivel macro.

Hay tres métodos estándar para valorar utilidades en la toma de decisiones en el ámbito médico: la apuesta estándar, la ponderación de tiempo y la escala análoga visual (Roberts & Sonnenberg, 2000; Miyamoto, 2000). Sin entrar en detalles, en todos estos métodos la meta es valorar la utilidad del resultado de un estado de salud. El fin último es obtener *años de vida ajustados por calidad* (AVAC o QALY, por su sigla en inglés), que son el resultado de combinar las utilidades asignadas a un resultado con la duración esperada de supervivencia para cada resultado. A cada año de salud plena se le asigna el valor 1, mientras que a la muerte se le asigna el valor 0. Si los años adicionales no son vividos con salud plena, se asigna un valor intermedio, que es multiplicado por la duración esperada de supervivencia. Los QALY pueden ser usados directamente por el paciente para valorar la utilidad de diferentes alternativas de tratamiento, o por aquellos que diseñan políticas públicas para valorar el uso más eficiente de los limitados recursos destinados a la salud.

En lo que sigue, voy a esbozar cómo las valoraciones de utilidad en los tres niveles, los estándares variables que generan, y los conflictos que se producen, pueden ser replicados en el caso de la determinación del estándar de prueba en el derecho. El paralelo tiene varias limitaciones en el nivel micro, donde hay varias diferencias obvias entre el contexto médico y el jurídico: (1) el estándar de prueba no es determinado por las circunstancias socioeconómicas del imputado; (2) el imputado no tiene el derecho a decidir si acepta la sentencia impuesta por el juez o el jurado; y (3) en muchos casos médicos la decisión es tomada usando el método de la ponderación de probabilidades entre dos tratamientos alternativos, en lugar de valoraciones de tratamientos individuales (Llewellyn-Thomas *et al.*, 1996). Las circunstancias que rodean el proceso de toma de decisiones en el nivel micro son replicadas de manera más natural, en el contexto jurídico, durante el proceso de negociación de penas (*plea bargaining*), muy común en los sistemas jurídicos adversariales.

Un estándar de prueba variable en el derecho, análogo al umbral de tratamiento variable que encontramos en medicina, también puede ser formulado en términos de funciones de utilidad. Basado en el trabajo de Kaplan (1968), Cullison (1969) y Tribe (1971), Erik Lillquist (2002) ha propuesto un estándar variable para el derecho penal, que según él es más benéfico para la sociedad. De acuerdo con la teoría de las decisiones, una persona que tome una decisión en condiciones de incertidumbre debe escoger la opción que tenga la mayor utilidad esperada. En general, la utilidad esperada de una decisión de condenar a un imputado (EUC) es la utilidad de una condena acertada (Ucg), sopeada por la probabilidad de que el imputado es en efecto culpable (Pg), más la

desutilidad de una convicción errónea (U_{ci}), sopesada por la probabilidad de que el imputado es en efecto inocente ($1 - P_g$). Formalmente,

$$EUC = P_g U_{cg} + (1 - P_g) U_{ci}$$

Por otra parte, si U_{ag} es la desutilidad de absolver un imputado culpable y U_{ai} es la utilidad de absolver un imputado inocente, la utilidad esperada de absolver a un imputado será:

$$EUA = P_g U_{ag} + (1 - P_g) U_{ai}$$

El estándar óptimo deberá ser fijado en el punto en el que $EUC = EUA$. Si se resuelve para P_g , que corresponde al estándar de prueba SP , obtenemos la siguiente ecuación:

$$SP = \frac{1}{1 + \frac{(U_{cg} - U_{ag})}{(U_{ai} - U_{ci})}}$$

Dada esta ecuación, tenemos dos opciones. Podemos intentar asignar valores tentativos a las cuatro utilidades y desutilidades y examinar el valor resultante para SP ; o podemos seleccionar un valor para SP y explorar qué rangos de valuaciones de utilidad son consistentes con él. El problema, como muchos autores lo han mostrado, es que si comenzamos con un SP razonable —por ejemplo, 0.9 para BARD— terminamos con utilidades y desutilidades poco razonables. Si comenzamos con utilidades y desutilidades razonables, terminamos con un estándar de prueba mucho más bajo¹⁸. Lillquist arguye que la fuente del problema es que no hay ninguna teoría que pueda explicar los costos y beneficios de los varios posibles veredictos en todos los casos. “En cambio, los costos de los veredictos erróneos y los beneficios de los veredictos acertados variarán con seguridad dependiendo de la naturaleza del crimen y del carácter del imputado” (2002, p. 91). Lillquist ofrece tanto evidencia empírica como una teoría normativa para defender la idea de un estándar de prueba variable.

Muchos de los estudios empíricos citados por Lillquist muestran que los jurados y los jueces, en los casos penales, le asignan al estándar de *más allá de toda duda razonable* valores numéricos que varían ampliamente y que son mucho más bajos de lo que en general se presume (Simon, 1970; Hastie, 1994). Aún más relevante para nuestro propósito es que hay estudios que muestran

18 Connolly (1999) ofrece una discusión muy clara sobre la tensión que existe entre utilidades razonables y estándares de prueba aceptables. El problema ya había sido detectado por Tribe (1971) y en la respuesta ofrecida por Milanich (1981).

que las variaciones numéricas del estándar de prueba están determinadas por el tipo de crimen involucrado. El asesinato y el homicidio involuntario obtuvieron el estándar más alto, mientras que el hurto simple obtuvo el más bajo: 0.75 (Simon & Mahan, 1971). El tipo de pruebas presentadas durante el juicio también afecta el estándar de prueba. Las pruebas acerca de la personalidad del imputado o acerca de sus antecedentes judiciales afectan su probabilidad de ser condenado (Hans & Vidmar, 1986). La severidad del castigo y lo impresionante del crimen también son factores que afectan el estándar de prueba (Elwork *et al.*, 1981).

Estos resultados empíricos pueden ser interpretados como el producto de la variabilidad en las utilidades y desutilidades que cada miembro del jurado le asigna a cada uno de los cuatro posibles resultados de un juicio, y como una indicación de que estas valoraciones de utilidad son muy sensibles a las circunstancias de cada caso particular. El paralelo con las valoraciones de utilidad altamente individuales que se realizan en el nivel micro en medicina es evidente.

Así como las decisiones médicas y los umbrales de tratamiento en el nivel meso y macro no dependen de las valoraciones de utilidad de los doctores y pacientes individuales, las utilidades asignadas por los legisladores y reguladores del aparato judicial para determinar las leyes y las directrices jurídicas no pueden depender de las valoraciones individuales de jueces y jurados. La consideración de procesos de toma de decisiones en torno a la distribución de recursos fiscales y humanos limitados de inmediato nos lleva a tener en cuenta el uso del dinero en el sistema judicial. Al respecto señala Laudan: “Vale la pena que nos preguntemos si es un uso eficiente del dinero y de otros recursos requerir que el estado organice el mismo tipo de prueba para enviar a un malversador a la cárcel durante un año que para condenar a un asesino en serie” (2006, p. 56). Las decisiones que se toman al nivel social también nos deben llevar a considerar los valores sociales y culturales que deben contribuir a darle forma a un sistema jurídico. En muchos estados de los Estados Unidos, por ejemplo, un menor de edad puede ser juzgado como adulto en ciertas circunstancias. Se podría argumentar que la desutilidad de la condena errónea de un menor es mayor que la desutilidad de la condena errónea de un adulto y que, por lo tanto, el estándar de prueba debería ser más alto para menores que para adultos. Este argumento no está basado en el interés de encontrar la verdad de los hechos, sino en la intención de ofrecerles a los menores un mínimo de protección ante fiscales despiadados. Muchos otros factores sociales y culturales tendrán una enorme relevancia en la valoración de utilidades en el nivel meso y macro, generando de esa manera estándares diferentes de un país a otro, e incluso de un estado federal a otro.

La pregunta más importante es cómo conectar los procesos de decisión en el nivel micro con la adopción de leyes y políticas públicas en el nivel macro.

Es imposible hacer justicia a esta pregunta en este ensayo. No obstante, de lo que he dicho se sigue que en ausencia de un consenso social acerca de la valoración de utilidades para los cuatro valores involucrados, una política jurídica sensata debería permitir una determinación más flexible y contextualizada del estándar de prueba para tipos de delitos. El caso médico muestra que la flexibilidad y la sensibilidad frente al contexto no llevan a la anarquía, siempre y cuando las decisiones sean tomadas de la manera más informada posible. El caso de la medicina también muestra, sin embargo, que los conflictos entre los diferentes niveles de decisión son inevitables, y que los pacientes no siempre podrán acceder al mejor tratamiento posible, según sus propias valoraciones de utilidad. En el derecho, un conflicto similar es posible entre un conjunto de lineamientos más flexibles para establecer el estándar de prueba y las preferencias individuales de jueces y jurados, determinadas por diferentes ideologías.

Concluiré examinando brevemente el paralelo entre los umbrales de prueba diagnóstica y la decisión de admitir una pieza de información como prueba en un juicio. Hay que admitir que el paralelo está limitado tan sólo a algunos casos específicos. Recordemos que está permitido llevar a cabo una prueba diagnóstica sólo si al menos una decisión acerca del tratamiento del paciente cambiará a raíz de los resultados de la prueba. El umbral de tratamiento de esa enfermedad determina si habrá un cambio en alguna decisión. Del mismo modo, si al final de un juicio hay pruebas abrumadoras para asegurarnos de que la hipótesis de culpabilidad está muy por encima del estándar de prueba, no tiene mucho sentido en ese punto admitir como prueba una pieza de información que no conduciría a un cambio en el intervalo de probabilidad de la hipótesis. La información ni siquiera calificaría como prueba según la teoría de los umbrales múltiples. Sin embargo, la mayoría de los juicios no se deciden a partir de una sola prueba. La naturaleza dinámica de un juicio aumentará y disminuirá la probabilidad de la hipótesis, empujándola a un lado y al otro del límite establecido para el estándar de prueba. Por lo tanto, sólo será posible imponer umbrales de aceptabilidad muy cercanos a los límites del intervalo $[0, 1]$, los que corresponderán a aquellos casos en los que es extremadamente implausible que el imputado sea culpable o inocente a la luz de las pruebas presentadas en cualquier momento durante el juicio¹⁹.

19 Otro umbral importante en el derecho tiene que ver con la admisibilidad de la prueba a la luz de consideraciones constitucionales. Aunque en muchas ocasiones las cortes han exigido prueba más allá de toda duda razonable de que la admisión de un cierto tipo de prueba no pondrá en peligro los derechos constitucionales del imputado, existe un consenso creciente en el que el estándar que se debe utilizar es la preponderancia de la prueba. Algunos autores que apoyan este punto de vista aún prefieren un estándar más alto para las confesiones. No discutiré este umbral aquí.

VI. Conclusiones

La teoría de los umbrales múltiples presentada en este ensayo puede ser vista como una alternativa viable a la ortodoxia bayesiana y como un puente entre el proceso de toma de decisiones en la medicina y el derecho. Como alternativa al bayesianismo, la teoría requiere una reinterpretación epistémica de la función de probabilidad que la disocie de su papel usual de representación de grados de creencia. Esta reinterpretación se necesita con el fin de defender la idea de que la creencia y nociones asociadas como justificación, aceptación y confianza también son conceptos umbrales. Así, la evidencia puede ser caracterizada en términos de su potencial para aumentar (o disminuir) la probabilidad de una hipótesis, más allá de alguno de los umbrales preestablecidos.

La teoría de los umbrales múltiples deriva parte de su fortaleza de las prácticas científicas reales. En muchas ciencias existen umbrales cualitativos y cuantitativos que determinan decisiones prácticas y teóricas. En el caso de la medicina, los dos umbrales principales, el umbral de tratamiento y el de prueba diagnóstica, están íntimamente conectados y se determinan uno al otro a partir de la utilidad de sus respectivos resultados. La flexibilidad de los umbrales de tratamiento puede ser replicada en el caso jurídico. Si se adoptaran múltiples umbrales o estándares dependientes de la severidad del crimen y de otros factores que afectan la utilidad y la desutilidad de los posibles resultados, el caso médico nos podría dar algunas luces acerca de cómo proceder. Podemos buscar, entonces, en la investigación empírica básica que se lleva a cabo en la psicología, la sociología y la antropología un apoyo para entender la sensibilidad contextual del proceso de toma de decisiones en el derecho. La meta final es alcanzar una etapa en los estudios jurídicos orientados por la investigación empírica, en la que sea posible hablar del derecho basado en evidencia.

Referencias

- ACHINSTEIN, Peter. (1983). Concepts of evidence. En P. Achinstein (Ed.), *The concept of evidence* (pp. 145-174). Oxford: Oxford University Press.
- (2001). *The book of evidence*. Nueva York: Oxford University Press.
- ALLEN, Ronald J. (1986). A reconceptualization of civil trials. *Boston University Law Review*, 66, 401-437.
- BRYANT, G. D., & Norman G. R. (1980). Expressions of probability words and numbers. *New England Journal of Medicine*, 302, 411.
- CARNAP, Rudolf. (1962). *Logical foundations of probability*. Chicago: University of Chicago Press.

- COHEN, L. Jonathan. (1977). *The probable and the provable*. Oxford: Oxford University Press.
- CONNOLLY, Terry. (1999). Decision theory, reasonable doubt, and the utility of erroneous acquittals. En T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making. An interdisciplinary reader*. Nueva York: Cambridge University Press.
- CULLISON, Alan D. (1969). Probability analysis of judicial fact-finding: Preliminary outline of the subjective approach. *The University of Toledo Law Review*, 1, 538-598.
- ELWORK, Amiram, Sales, Bruce D., & Suggs, David. (1981). The trial: A research review. En B. D. Sales (Ed.), *The trial process*. Nueva York: Springer.
- FRIEDLAND, Daniel J., Go, Alan S., J. Davoren, Ben, Shlipak, Michel G., Bent, Stephen W., Subak, Leslee L., & Mendelson, Terrie. (1998). *Evidence-based medicine. A framework for clinical practice*. Nueva York: McGraw Hill.
- GREENHALGH, Trisha. (2010). *How to read a paper: The basics of evidence-based medicine*. Hoboken: Wiley-Blackwell.
- HAACK, Susan. (2013). El probabilismo jurídico: una disención epistemológica. En C. Vásquez (Ed.), *Estándares de prueba y prueba científica. Ensayos de epistemología jurídica*. Madrid: Marcial Pons.
- HANS, Valerie P., & Vidmar, Neil. (1986). *Judging the jury*. Nueva York: Plenum Press.
- HASTIE, Reid. (1994). Algebraic models of decision processes. En R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making*. Nueva York: Cambridge University Press.
- HOWSON, Colin. (1991). The 'old evidence' problem. *British Journal for the Philosophy of Science*, 42, 547-555.
- HOWSON, Colin, & Urbach, Peter. (2006). *Scientific reasoning. The Bayesian approach*. La Salle: Open Court.
- HUNINK, Myriam, & Glasziou, Paul. (2001). *Decision making in health and medicine*. Nueva York: Cambridge University Press.
- JACKSON, Rodney. (2000). Updated New Zealand cardiovascular disease risk-benefit prediction guide. *BMJ*, 320, 709-710.
- KAHNEMAN, Daniel, & Tversky, Amos. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 276-287.
- KAPLAN, John. (1968). Decision theory and the factfinding process. *Stanford Law Review*, 20, 1065-1092.
- LAUDAN, Larry. (2006). *Truth, error, and criminal law. An essay in legal epistemology*. Nueva York: Cambridge University Press.

- LEVI, Isaac. (1991). *The fixation of belief and its undoing*. Nueva York: Cambridge University Press.
- (2004). *Mild contraction. Evaluating loss of information due to loss of belief*. New York: Cambridge University Press.
- LILLQUIST, Erik. (2002). Recasting reasonable doubt: Decision theory and the virtues of variability. *University of California Davis Law Review*, 36, 146-176.
- LLEWELLYN-THOMAS, Hilary A., Williams, J. Ivan, Levi, Linda, & Naylor, C. D. (1996). Using a trade-off technique to assess patients' treatment preferences for benign prostatic hyperplasia. *Medical Decision Making*, 16, 262-272.
- MAHER, Patrick. (1996). Subjective and objective confirmation. *Philosophy of Science*, 63, 149-174.
- MCISAAC, Warren J., Goel, Vivel, To, Teresa, & Low, Donald E. (2000). The validity of a sore throat score in family practice. *Canadian Medical Association Journal*, 163(7), 811-815.
- MILANICH, Patricia G. (1981). Decision theory and standards of proof. *Law and Human Behavior*, 5, 87-96.
- MIYAMOTO, John M. (2000). Utility assessment under expected utility and rank-dependent utility assumptions. En G. B. Chapman, & F. A. Sonnenberg (Eds.), *Decision making in health care. Theory, psychology, and applications*. Nueva York: Cambridge University Press.
- OHNISHI, Motoki, Fukui, Tsuguya, Matsui, Kunihiko, Hira, Kenji, Shinozuka, Masaya, Ezaki, Hironori, Otaki, Junji, Kurokawa, Wataru, Imura, Hiroshi, Koyama, Hiroshi, & Shimbo, Takuro. (2002). Interpretation of and preference for probability expressions among Japanese patients and physicians. *Family Practice*, 19, 7-11.
- PÁEZ, Andrés. (2013). Probability-lowering causes and the connotations of causation. *Ideas y Valores*, 151, 43-55.
- ROBERTS, Mark S., & Sonnenberg, Frank A. (2000). Decision making techniques. En G. B. Chapman, & F. A. Sonnenberg (Eds.), *Decision making in health care. Theory, psychology, and applications*. Nueva York: Cambridge University Press.
- ROTT, Hans, & Pagnucco, Maurice. (1999). Severe withdrawal (and recovery). *Journal of Philosophical Logic*, 28, 501-547.
- ROUSH, Sherrilyn. (2003). Review of *The book of evidence* by Peter Achinstein. *Isis*, 94, 203-204.
- (2004). Positive relevance defended. *Philosophy of Science*, 71, 110-116.
- SALMON, Wesley C. (1975). Confirmation and relevance. En G. Maxwell, & R. M. Anderson (Eds.), *Induction, probability, and confirmation. Minnesota studies in the philosophy of science* (Vol. vi) (pp. 3-36). Minneapolis: University of Minnesota Press.

- SCHUM, David. (2001). *The evidential foundation of probabilistic reasoning*. Evanston: Northwestern University Press.
- SHAW, N. J., & Dear, P. R. (1990). How do parents of babies interpret qualitative expressions of probabilities? *Archives of Disease in Childhood*, 65, 520-523.
- SIMON, Rita J. (1970). "Beyond a reasonable doubt" - An experimental attempt at quantification. *The Journal of Applied Behavioral Science*, 6, 203-209.
- SIMON, Rita J., & Mahan, Linda. (1971). Quantifying burdens of proof: A view from the bench, the jury and the classroom. *Law & Society Review*, 5, 319-330.
- STIGGELBOUT, Anne M. (2000). Assessing patients' preferences. En G. B. Chapman, & F. A. Sonnenberg (Eds.), *Decision making in health care. Theory, psychology, and applications*. Nueva York: Cambridge University Press.
- TRIBE, Laurence. (1971). Trial by mathematics: Precision & ritual in the legal process. *Harvard Law Review*, 84, 1329-1375.
- TSEVAT, J., Weeks, J. C., Guadagnoli, E., *et al.* (1994). Using health-related quality-of-life information: Clinical encounters, clinical trials, and health policy. *Journal of General Internal Medicine*, 9, 576-82.