

# EVIDENCE-BASED SENTENCING Y EVIDENCIA CIENTÍFICA. A LA VEZ, ALGUNAS CONSIDERACIONES SOBRE LAS «POLÍTICAS BASADAS EN LA EVIDENCIA» Y EL DERECHO PENAL<sup>\*</sup>

*EVIDENCE-BASED SENTENCING AND SCIENTIFIC EVIDENCE. AT THE SAME TIME, SOME REFLECTIONS ON «EVIDENCE-BASED POLICIES» AND CRIMINAL LAW*

*Lucía Martínez Garay*

*Profesora Titular de Derecho penal*

*Universidad de Valencia*

## RESUMEN

El evidence-based sentencing (EBS) es una denominación reciente para una pretensión que tiene hondas raíces en el Derecho penal: aplicar las penas más adecuadas al nivel de riesgo de reincidencia que presenta cada condenado para intentar reducir ese riesgo al mínimo posible. Esta versión moderna de la prevención especial se inserta en el enfoque más amplio de las «políticas públicas basadas en la evidencia». Considera que la mejor evidencia existente para el control de la reincidencia son las modernas herramientas de valoración estructurada del riesgo, y aduce ser capaz de lograr varios objetivos a la vez: reducir la reincidencia, mantener un elevado nivel de seguridad ciudadana, hacer un uso más eficiente de los recursos públicos y alejar la política criminal de luchas ideológicas al fundamentarla en evidencia científica objetiva. De los muchos problemas que suscita el EBS, este trabajo analiza solo uno: la calidad de la evidencia científica en que se basa, y los datos disponibles sobre en qué medida alcanza (o no) los objetivos que pretende.

## PALABRAS CLAVE

Evidence-based sentencing, políticas basadas en la evidencia, valoración de riesgo de reincidencia, evidencia científica, Derecho penal, prevención especial.

## ABSTRACT

Evidence-based sentencing (EBS) is a recent name for a claim that has deep roots in criminal law: to apply the sentence that is most appropriate to the risk of recidivism posed by each offender, in order to reduce this risk as much as possible. This modern version of special prevention inserts itself in the broader approach of the so-called «evidence-based public policies». It considers that the best existing evidence for reducing recidivism are modern structured risk assessment tools, and it claims to be able to achieve several objectives at the same time: reduce recidivism, maintain a high level of public safety, make a more efficient use of public resources, and move criminal policy away from ideological struggles by basing it on the objective knowledge provided by the best available scientific evidence. Of the many problems that surround EBS, this work focuses only on one: the quality of the scientific evidence on which it is based, and the available data on the extent to which it achieves (or not) its intended objectives.

## KEY WORDS

Evidence-based sentencing, evidence-based policies, violence risk assessments, scientific evidence, Criminal Law, selective incapacitation and rehabilitation.

DOI: [doi.org/10.36151/td.2020.018](https://doi.org/10.36151/td.2020.018)

\* Este trabajo se ha realizado en el marco de los proyectos de investigación «Derecho penal de la peligrosidad: tutela y garantía de los derechos fundamentales» (DER2017-86336-R, financiado por FEDER/Ministerio de Ciencia, Innovación y Universidades-Agencia Estatal de Investigación) y «Derecho Penal y Comportamiento Humano» (RTI2018-097838-B-I00, financiado por el Ministerio de Ciencia, Innovación y Universidades). Una versión algo distinta y más breve de este trabajo se publicará en el libro colectivo *El Derecho penal ante «lo empírico». Sobre el acercamiento del Derecho penal y la Política Criminal a la realidad empírica*, dirigido por los profs. Miró Llinares y Fuentes Osorio, para el cual fue inicialmente concebido. Agradezco a Tomás Vives Antón sus acertados consejos y comentarios sobre sucesivos borradores, que han contribuido a mejorar notablemente la estructura y el contenido del trabajo.

# **EVIDENCE-BASED SENTENCING Y EVIDENCIA CIENTÍFICA. A LA VEZ, ALGUNAS CONSIDERACIONES SOBRE LAS «POLÍTICAS BASADAS EN LA EVIDENCIA» Y EL DERECHO PENAL**

**Lucía Martínez Garay**

Profesora Titular de Derecho Penal  
Universidad de Valencia

**Sumario:** 1. Introducción. 2. Qué es el *evidence-based sentencing* (EBS). Precisiones terminológicas y delimitación del objeto del trabajo. 3. Antecedentes y contexto del *evidence-based sentencing*. 3.1. Los principios de la intervención correccional efectiva y el modelo «riesgo-necesidad-responsividad». 3.2. La justicia gerencial-actuarial y la gestión efectiva del riesgo. 3.3. Las «políticas basadas en la evidencia». 3.3.1. La medicina basada en la evidencia. 3.3.2. La aplicación en otros ámbitos del modelo «basado en la evidencia». 4. El *evidence-based sentencing* y la evidencia científica. 4.1. La calidad de la evidencia sobre la que se basa el *evidence-based sentencing*. 4.2. ¿Hay evidencia empírica de que el *evidence-based sentencing* alcance sus objetivos? 4.3. Discusión. 5. Conclusiones. Notas. Bibliografía

## 1. INTRODUCCIÓN

La evidencia suministrada por las ciencias sociales sobre los efectos de las penas ha sido sistemáticamente obviada por la política criminal en los EE. UU. desde hace al menos 40 años<sup>1</sup>. La segunda mitad de la década los ochenta y toda la de los noventa se caracterizaron por el predominio del eslogan del *tough on crime* y la «guerra contra las drogas»: aumentó el número de delitos castigados con cadena perpetua sin posibilidad de libertad condicional (*parole*), se derogó la *parole* en muchos estados y también a nivel federal, se aprobaron leyes que imponen duras condenas mínimas de cumplimiento obligatorio, se introdujeron las *three-strikes laws*, se previeron penas más severas o internamientos indefinidos para *sexual violent predators* y se incrementó en general la severidad de los castigos, también para los jóvenes. Los resultados de esta política criminal son conocidos: las tasas de encarcelamiento se dispararon hasta alcanzar en 2008 un pico de 2.310.300 personas encarceladas, lo que representa una tasa de 760 por cada 100.000 residentes<sup>2</sup>, con tremendas disparidades raciales y un coste billonario para la Federación y los estados, entre otros problemas<sup>3</sup>.

Ninguna de estas iniciativas contaba con evidencia que demostrara su eficacia en la disminución de la delincuencia (Tonry, 2013b; y Haggerty, 2004: 220). En apariencia (pues al menos igual de importantes que las razones político-criminales eran las razones políticas a secas), estas políticas pretendían reducir la criminalidad a través de los efectos intimidatorios e inocuizadores de sanciones más duras<sup>4</sup> (Tonry, 2013b:159), pero los estudios disponibles ya entonces, y también los que se han ido publicando con posterioridad, demostraban y demuestran que dichos efectos son mínimos, inexistentes o incluso contraproducentes (cfr. Tonry, 2013b: 175 ss.; y National Research Council, 2014: 337-339).

Desde finales de la década de los noventa, sin embargo, y en un contexto en el que la política criminal estadounidense ha abandonado las formas más extremas del *tough on crime*<sup>5</sup>, ha ido aumentando la popularidad de lo que ha dado en denominarse *evidence-based practices*. Se trata de una expresión aplicable a muchos de los campos involucrados en el control del delito: la actividad policial, la instrucción del procedimiento y la adopción de medidas cautelares, y la fase final de imposición y ejecución de las penas. En relación con este último ámbito, el único que se analizará en este trabajo, la expresión hace referencia a un conjunto de técnicas y estrategias que utilizan los tribunales, las instituciones penitenciarias y los servicios de supervisión en la comunidad cuya aplicación pretende ajustar la intervención punitiva a las características específicas de diferentes individuos y grupos para hacerlas más efectivas (Klingele, 2015: 539). En concreto, se trataría de «[...] políticas, procedimientos y programas que la evidencia científica ha demostrado que reducen la reincidencia para grupos específicos de delincuentes, como personas sometidas a *probation*<sup>6</sup>, en *parole*, o delincuentes adictos a drogas» (National Conference of State Legislatures, 2011: 28). Este enfoque está cosechando un éxito notable en los EE. UU., pues está siendo implementado —con mayor o menor profundidad— en muchas de sus jurisdicciones<sup>7</sup>.

Considerando el trasfondo del panorama que hemos resumido en las líneas anteriores, llama inmediatamente la atención el nombre que reciben estas prácticas, «basadas en la evidencia»: ¿podría ser que estuviéramos ante un cambio de tendencia, y que, tras más

de 40 años de desprecio por la evidencia científica, la política criminal estadounidense en materia de tipificación, imposición y ejecución de penas se esté diseñando realmente con base en el conocimiento científico que las ciencias sociales han ido acumulando sobre los efectos de las penas? Considero que, por desgracia, no es así, y que, aunque en algunos casos esté justificada, la apelación a la «evidencia científica» corre el riesgo de convertirse en un recurso para legitimar prácticas que no han demostrado su eficacia en la reducción de la delincuencia y para presentar como meras cuestiones técnicas lo que no dejan de ser opciones políticas.

## 2. QUÉ ES EL *EVIDENCE-BASED SENTENCING* (EBS). PRECISIONES TERMINOLÓGICAS Y DELIMITACIÓN DEL OBJETO DEL TRABAJO

El *evidence-based sentencing* (en adelante, EBS)<sup>8</sup> consiste en la aplicación, en el momento de la imposición de la pena y también durante la ejecución de la misma, de una serie de pautas y criterios que derivan de los resultados de investigaciones científicas rigurosas sobre las prácticas que han demostrado ser más efectivas para reducir la reincidencia<sup>9</sup>. Las investigaciones científicas a las que se refieren los partidarios de este enfoque —que no dudan en calificar de «las mejores» (*the best research evidence*)— son, por un lado, los estudios sobre programas de rehabilitación y técnicas de intervención que han demostrado buenos resultados en grupos específicos de delincuentes, y, por otro, las investigaciones sobre la capacidad predictiva de las modernas herramientas estructuradas de estimación de riesgo de reincidencia o de violencia.

En el enfoque del EBS hay una relación muy estrecha entre el momento de la imposición de la pena y el de la posterior ejecución de la misma, ya que de lo que se trata es de imponer condenas que, por sus características, permitan ejecutar luego los programas que se consideran más idóneos para reducir el riesgo de reincidencia (Casey, Warren y Elec, 2011: 7). Por ello, y aunque algunas definiciones del EBS atienden sobre todo al momento de elección de la pena adecuada en la sentencia<sup>10</sup>, es común entender que incluye también (o incluso especialmente) algunas decisiones que tienen que adoptarse en el curso de su ejecución, especialmente las relacionadas con la concesión o no de *probation*, su eventual revocación y el régimen de supervisión/tratamiento que ha de acompañarla.

Aunque en este trabajo utilizaré la expresión *evidence-based sentencing*, existen otras denominaciones que, en esencia, se refieren a la mismas técnicas y estrategias, aunque cada una de ellas destaca una característica distinta del fenómeno. Así, la evidencia científica sobre la que descansa el EBS son las herramientas actariales o estructuradas de estimación de riesgo de reincidencia (o de reincidencia violenta) que tanto se han desarrollado en la Criminología a lo largo de las últimas décadas (*actuarial sentencing*, Hannah-Moffat, 2013)<sup>11</sup>. Esas herramientas permiten estimar el nivel de riesgo de cada delincuente, esto es, predecir la probabilidad de que siga delinquiendo (*predictive sentencing*, de Keijser, Roberts y Ryberg, 2019). Lo que el EBS propone es utilizar ese nivel de riesgo como el criterio central para decidir cuál es la condena más adecuada para el sujeto (*risk-based sentencing*,

Slobogin, 2019). Se afirma que ello permitirá un uso más eficiente de los recursos públicos (*effective sentencing*, National Conference of State Legislatures, 2011), dado que los concentraría en los sujetos que realmente los necesitan y evitaría la imposición de penas largas e intervenciones intensivas a los de bajo riesgo. Todo ello supone, se dice, una gestión mucho más inteligente del problema penal (*smart sentencing*, Marcus, 2006)<sup>12</sup>, pues con los mismos —o con menos— recursos públicos se conseguirá una reducción de la reincidencia y una mejor protección de la comunidad<sup>13</sup>.

La razón de que en este trabajo se haya optado por la expresión *evidence-based sentencing* en lugar de alguna de las otras es doble. Por un lado, esta denominación es la que expresa la pretensión del enfoque que en este estudio se va a analizar críticamente: al autodenominarse «basado en la evidencia», el EBS parece asumir que existe más o mejor evidencia científica para apoyar una orientación preventivo-especial del sistema penal construida alrededor de la noción central de riesgo que en otros modelos posibles de justicia penal, asunción que en este trabajo trataré de rebatir. Por otro lado, al utilizar esta denominación el EBS se alinea deliberadamente con un movimiento de alcance más general, el de las políticas basadas en la evidencia o *evidence-based policymaking*. Este movimiento persigue un objetivo a mi juicio loable (tener en cuenta la evidencia científica a la hora de diseñar cualesquiera políticas públicas), pero por vías que suscitan numerosas dudas, como tendrá ocasión de exponer más adelante. El EBS ofrece, por tanto, una ocasión para analizar no solo el concreto aspecto de la relación entre evidencia empírica y ejecución de las penas, sino también la relación más general entre el Derecho penal y la evidencia empírica en los términos en que la concibe el enfoque de las políticas basadas en la evidencia.

En definitiva, este trabajo se centra en el análisis de la calidad de la evidencia científica sobre la que se basa el EBS, y en la evaluación de la evidencia empírica que eventualmente acrecita que el EBS alcanza los objetivos que pretende conseguir. Quedan fuera del objeto de estudio otros muchos aspectos problemáticos del EBS sobre los cuales existe un intenso debate doctrinal y que en absoluto son de menor relevancia que el que aquí se va a tratar<sup>14</sup>, pero que, por elementales razones de espacio, no es posible abordar en este momento. En todo caso, el objeto de estudio elegido no es tan estrecho como pudiera parecer, pues ofrece la oportunidad de abordar tres cuestiones a mi juicio muy relevantes en el debate penal y criminológico del presente y, probablemente, del futuro. Por un lado, la relación entre evidencia empírica y Derecho penal en un momento en el que, desde diversas perspectivas, parece estar produciéndose un acercamiento entre ambas esferas (Miró Llinares, 2017). Por otro lado, permite también hacer algunas consideraciones sobre un tipo concreto de evidencia empírica —el análisis estadístico de grandes conjuntos de datos— que va a plantear importantes retos a la tradicional administración de la justicia penal<sup>15</sup>. Por último, ofrece la oportunidad de introducir en el debate jurídico-penal español la cuestión de las «prácticas basadas en la evidencia», que hasta ahora, y salvo contadas excepciones, no se ha tematizado<sup>16</sup>.

### 3. ANTECEDENTES Y CONTEXTO DEL *EVIDENCE-BASED SENTENCING*

Para situar el problema, conviene ofrecer en primer término alguna pincelada sobre el contexto (histórico, político y epistemológico) en el que ha surgido el *evidence-based sentencing*.

#### 3.1. LOS PRINCIPIOS DE LA INTERVENCIÓN CORRECCIONAL EFECTIVA Y EL MODELO «RIESGO-NECESIDAD-RESPONSIVIDAD»

En el ámbito de la justicia criminal, los primeros programas y procedimientos que se autodenominaron «basados en la evidencia» nacieron en el campo de la ejecución de las penas, en el contexto del descrédito en que había caído el ideal rehabilitador desde la década de los ochenta. Enfrentados al ambiente hostil que representaban las políticas del *tough on crime* dominantes en aquel periodo, los profesionales que seguían creyendo en la prevención especial positiva, en los efectos criminógenos de la prisión y en las posibilidades de mejora de muchos condenados se esforzaron por reunir pruebas de que algunos programas e intervenciones de corte resocializador y rehabilitador sí funcionaban. Así, se inició un debate sobre cuáles eran los programas de tratamiento efectivos o más efectivos y, a partir de la publicación de los trabajos de Andrews y Bonta, entre otros, fue construyéndose un enfoque teórico orientado a sistematizar y explicar los «principios de la intervención correccional efectiva» que cristalizó en el denominado modelo de «Riesgo, Necesidad y Responsividad», o RNR (Velásquez Valenzuela, 2014: 70 ss.; y Klingele, 2015: 552 ss.). Este modelo se construye alrededor de la noción central de riesgo, y se basa —muy resumidamente— en las siguientes premisas: *i*) que el riesgo de reincidencia es medible y que hoy día esta medición puede llevarse a cabo con métodos actuariales o estructurados, mucho más fiables que el método clínico tradicional; *ii*) que las intervenciones deben ajustarse al nivel de riesgo de cada sujeto (siendo desaconsejables intervenciones intensivas en sujetos de bajo riesgo, que pueden incluso tener efectos contraproducentes, y debiendo concentrarse el esfuerzo y los recursos en los individuos de mayor nivel de riesgo); *iii*) que cada sujeto tiene unas necesidades específicas (médicas, económicas, cognitivo-conductuales, educativas, etc.) que favorecen la comisión de nuevos delitos, y que, por tanto, la actuación sobre esas necesidades puede reducir su nivel de riesgo; y *iv*) que las intervenciones deben ajustarse a la receptividad de cada sujeto o grupos de sujetos, pues no todos responden de la misma manera a los mismos programas o técnicas (Andrews y Bonta, 2010: 44 ss.).

Estos estudios consiguieron construir una «teoría sobre la rehabilitación efectiva» (Velásquez Valenzuela, 2014: 78) que, además, logró hacerse un importante hueco en el debate político criminal a partir de la década de los noventa. Sin embargo, la parte de dicho enfoque que posiblemente está resultando más decisiva en la política criminal actual no es la de la rehabilitación, sino la de la efectividad, como enseguida veremos.

### 3.2. LA JUSTICIA GERENCIAL-ACTUARIAL Y LA GESTIÓN EFECTIVA DEL RIESGO

El desarrollo de los principios de intervención efectiva que hemos descrito en el epígrafe anterior fue paralelo en el tiempo a otra línea de evolución en la política criminal estadounidense: la incorporación progresiva de técnicas de gestión del riesgo criminal basadas en el modelo gerencial-actuarial. En las últimas décadas ha ido abriéndose paso una nueva racionalidad en el control penal que se caracteriza por abordarlo desde la perspectiva de una gestión eficiente de los recursos del sistema. El crimen se percibe desde este enfoque no tanto como un problema grave que debe ser eliminado o contra el que hay que luchar cuanto como un fenómeno intrínseco a cualquier sociedad que puede gestionarse con mayor o menor eficiencia teniendo en cuenta los recursos disponibles. Según esta perspectiva, el riesgo es también un concepto central, aunque no exactamente en el mismo sentido en que lo es en el paradigma del modelo RNR al que nos hemos referido en el epígrafe anterior. El modelo gerencial-actuarial está poco preocupado por las causas del delito y no contempla la rehabilitación como un objetivo principal; pueden aplicarse programas rehabilitadores si dan buenos resultados en términos de coste-beneficio, pero el objetivo central es mantener el nivel de desviación controlado en márgenes tolerables. Para ello se usan indicadores de niveles de riesgo medidos con instrumentos actariales porque resultan útiles para gestionar grandes poblaciones de manera objetiva y con un mínimo de costes. La preocupación central para este enfoque es la eficiencia en el uso de recursos y la optimización de costes (Brandariz García, 2016: 93 ss. y 111 ss.; Klingele, 2015: 545 ss. y 572 y ss.; y Velásquez Valenzuela, 2014).

En este contexto hay que situar el relativo éxito de los programas de rehabilitación «basados en la evidencia». Como destaca Klingele (2015: 552), los profesionales y criminólogos que querían convencer a gestores y políticos poco predisuestos a destinar recursos económicos a programas de rehabilitación y tratamiento se esforzaron por demostrar que merecía la pena invertir en ellos porque producían beneficios medibles en términos de reducción de la reincidencia y, para ello, documentaron con exactitud los resultados y reunieron datos objetivos que avalaran su efectividad<sup>17</sup>. El énfasis en la efectividad en términos de coste-beneficio pretendía obtener financiación y oportunidades para unos programas que no parecía posible llevar adelante apelando solo al valor intrínseco del ideal rehabilitador como objetivo legítimo y justo de las sanciones penales.

El objetivo se consiguió, al menos en parte. Hoy en día se considera superado el viejo aforismo del «nothing works»<sup>18</sup> y se acepta prácticamente sin discusión que determinados programas y tratamientos —especialmente, los basados en el modelo RNR— han demostrado ser eficaces en la reducción de la reincidencia y que ese conocimiento debe ser utilizado por la Administración de justicia. Esta es la premisa fundamental de todos los documentos elaborados por los organismos oficiales, las asociaciones profesionales, los *think tanks* y otros grupos de interés relacionados con la Administración de justicia que promueven el uso de *evidence-based sentencing*. Ahora bien, lo que se ha incorporado al discurso oficial del EBS es el énfasis en la posibilidad de reducción *efectiva* de la reincidencia, pero no —o no tanto— la idea de que esa reducción tenga que conseguirse a través de la *rehabilitación* de los condenados. Esto es fácilmente constatable cuando se

analizan los fines, que según dichos documentos, debe perseguir el EBS: el objetivo esencial que se proclama en todos ellos es la reducción de la reincidencia, acompañada siempre de una referencia a la mejora de la seguridad pública (*public safety*) y la utilización más efectiva de los recursos públicos; en contraste, las referencias a la rehabilitación de los condenados ocupan, si es que están presentes, un muy difuminado segundo plano<sup>19</sup>. Tampoco el objetivo de reducir la población penitenciaria, que también aparece en casi todas las iniciativas que apoyan el EBS<sup>20</sup>, debe confundirse con la rehabilitación: la razón principal por la que se persigue la reducción de las tasas de encarcelamiento es que suponen un elevadísimo coste para las Administraciones, pero no se subraya suficientemente que el encarcelamiento masivo no ha demostrado ser un medio efectivo de reducción de la criminalidad. Es cierto que el EBS busca evitar los efectos criminógenos de la prisión, pero la razón principal por la que se busca reducir el número de internos es el ahorro de costes<sup>21</sup>, y además a través de medios que demuestren ser igualmente eficaces que la privación de libertad para la protección de la seguridad pública (más efectivos, por tanto, en términos de coste-beneficio)<sup>22</sup>.

La «rehabilitación efectiva» diseñada por Andrews y Bonta y otros profesionales ha sido, sin duda alguna, un pilar fundamental para el mantenimiento y la expansión de programas dirigidos a la resocialización de los condenados, tanto en prisión como fuera de ella, pero en el enfoque del EBS prima la dimensión de la efectividad sobre la de la rehabilitación, y las estimaciones de riesgo de violencia están, ante todo, al servicio de un control de la población condenada efectivo en términos de coste-beneficio. Cuando los programas de rehabilitación resulten la opción más efectiva para conseguir ese objetivo, se destinarán recursos para financiarlos, pero cuando no sea así, el control se implementará a través de otros mecanismos.

### 3.3. LAS «POLÍTICAS BASADAS EN LA EVIDENCIA»

Como hemos comentado al inicio de este trabajo, al incorporar la expresión «basado en la evidencia» el EBS no se refiere a cualquier tipo de conocimiento que las ciencias sociales hayan proporcionado sobre los efectos de las sanciones criminales, sino que se inserta conscientemente en un movimiento de alcance mucho más general conocido como *evidence-based policies* o *evidence-based policymaking*<sup>23</sup>. Este enfoque parte de los éxitos y del prestigio que alcanzó la «medicina basada en la evidencia» a principios de la década de los noventa y, muy resumidamente, considera que: *i)* hay una jerarquía en la calidad de la evidencia científica, en la cúspide de la cual se encuentran los ensayos controlados aleatorizados (*randomized control trials*) y los metaanálisis, mientras que la mera experiencia clínica del profesional se encuentra en el nivel más bajo; *ii)* no solo en medicina, sino también en el diseño de todo tipo de políticas públicas, las decisiones deben apoyarse en la mejor evidencia científica disponible, entendida según la jerarquía a la que acabamos de referirnos; y *iii)* las decisiones basadas en esa clase de evidencia científica son mejores porque, de un lado, han demostrado ser eficaces para lograr el objetivo propuesto, y por otro lado, no están ya motivadas (o no solo) por objetivos políticos —en el sentido de intereses

ideológicos o partidistas—, sino fundamentadas en «datos objetivos» sobre lo que funciona y lo que no («what works») para resolver los problemas. Desde este enfoque, se considera que estar «basado en la evidencia» es condición necesaria para una mayor transparencia, responsabilidad y mejor gobernanza (Strassheim y Kettunen, 2014: 259) lo que supone, en definitiva, un plus de legitimidad<sup>24</sup>.

Es indudable que la idea de desarrollar políticas basadas en evidencia científica resulta inmediatamente atractiva, también para el Derecho penal: «Igual que nadie defendería que los médicos deberían practicar su profesión de espaldas a la evidencia, también parece una afirmación incontestable, obvia, que los políticos deberían basar sus decisiones en la evidencia» (Klein, 2000: 65)<sup>25</sup>. Sin embargo, a pesar del indudable predicamento de que goza en muchos ámbitos la idea del *evidence-based policymaking*<sup>26</sup>, llevar a cabo políticas realmente basadas en evidencia científica es bastante más complejo de lo que podría parecer. Un buen ejemplo lo proporciona, precisamente, el campo en el que surgieron las prácticas basadas en la evidencia: la medicina.

### 3.3.1. La medicina basada en la evidencia

De acuerdo con la conocida definición de uno de los fundadores de esta corriente, la medicina basada en la evidencia es «[...] el uso consciente, explícito y juicioso de la mejor evidencia científica disponible en la toma de decisiones sobre el cuidado de los pacientes individuales» (Sackett *et al.*, 1996: 71)<sup>27</sup>. La medicina basada en la evidencia enfatiza la importancia de la experiencia personal y del juicio clínico del profesional, que es quien conoce las características y las preferencias de cada paciente individual (Richardson, 2017). Pero parte de la base de que la práctica diaria de la medicina se basa muchas veces en costumbres, intuiciones y conocimientos que han ido quedando obsoletos; los médicos siguen administrando a veces tratamientos que creen que funcionan aunque la evidencia científica no los respalde o incluso haya demostrado su ineffectividad. Los médicos han de dedicar la mayor parte de su tiempo al cuidado de sus pacientes y difícilmente pueden estar al día de los cientos de artículos que se publican cada semana en decenas de revistas especializadas<sup>28</sup>.

Por eso, la medicina basada en la evidencia proporciona estándares con los que medir la calidad científica de los tratamientos, terapias y diagnósticos, y además sistematiza y ordena el conocimiento existente sobre los mismos con el fin de facilitar a los profesionales el acceso a la información más rigurosa y actualizada que necesitan para ayudar a sus pacientes. A tal efecto, se establece una jerarquía en la calidad científica de los estudios, en la cúspide de la cual se sitúan los ensayos clínicos aleatorizados (*randomized control trials* o RCT), seguidos de los ensayos clínicos sin asignación aleatorizada, los estudios de cohortes, los estudios de casos y controles, etc.<sup>29</sup>. También se consideran de la máxima calidad científica los metaanálisis o revisiones sistemáticas y críticas realizadas con métodos estandarizados de los estudios existentes sobre una cuestión y que resumen sus resultados. Con los metaanálisis, a su vez, se elaboran Guías de Práctica Clínica que trasladan el conocimiento existente a recomendaciones concretas de acción para el diagnóstico y tratamiento de los pacientes.

La medicina basada en la evidencia ha tenido un enorme éxito. Sus logros más importantes, reconocidos incluso por sus críticos, son haber contribuido a desterrar prácticas clínicas ineficaces o dañinas, y haber conseguido —a través de los ensayos clínicos randomizados— descubrir importantes factores de riesgo y distinguir tratamientos eficaces de los que no lo son, prestando con ello un enorme servicio a los pacientes en enfermedades graves como el SIDA, la hepatitis, el cáncer o el asma (Fava, 2017; Horwitz y Singer, 2017: 14; y Greenhalgh, Howick y Maskrey, 2014: 1). La medicina basada en la evidencia ha concienciado, además, sobre la necesidad de cuestionarse continuamente la eficacia y los beneficios de los tratamientos que se aplican, sobre el reconocimiento de que el juicio clínico del profesional puede estar sesgado u obsoleto, y sobre la pertinencia de que la medicina se abra a nuevos descubrimientos.

Sin embargo, este enfoque también ha recibido importantes críticas. Una de las principales es que el desarrollo de la nueva evidencia científica no se produce de manera equilibrada ni está libre de sesgos porque la financiación de la investigación está muy ligada a intereses particulares: los ensayos clínicos aleatorizados son caros y necesitan una importante infraestructura, y en muchas ocasiones son las propias empresas farmacéuticas interesadas en vender el tratamiento las que organizan y financian los ensayos, con lo que surgen evidentes conflictos de interés (Knottnerus y Tugwell, 2017; Fava, 2017; Ioannidis, 2016; Greenhalgh, Howick y Maskrey, 2014: 1-2; y Berguer y Berguer, 2003)<sup>30</sup>. Por otro lado, la necesidad de contar con evidencia científica para defender la aplicación de cualquier terapia, técnica o fármaco estimula la publicación acelerada de resultados no siempre suficientemente contrastados en revistas que no siempre aplican criterios suficientemente rigurosos de admisión, práctica que con el tiempo ha producido una multiplicación de «basura» publicada (Knottnerus y Tugwell, 2017: 1) que, en lugar de mejorar, entorpece el avance de la medicina: «Demasiados trabajos de investigación están mal diseñados o ejecutados. Demasiados resultados de las investigaciones no son dados a conocer o se diseminan de manera fragmentada. Conforme ha ido aumentando el volumen de la actividad de investigación clínica, la calidad de la evidencia muchas veces ha empeorado»<sup>31</sup>.

También se critica a la medicina basada en la evidencia que la investigación tiende a centrarse en trastornos y poblaciones en las que el potencial beneficio comercial se estima grande, mientras que otros problemas importantes de salud pública quedan desatendidos; que minusvalora la relevancia de otros tipos de evidencia que no encajan en los estrechos márgenes de los ensayos clínicos randomizados; que olvida las necesidades y preferencias del paciente individual prescribiendo tratamientos estandarizados que solo han demostrado su efectividad para grupos; que desprecia la experiencia profesional del clínico y el necesario conocimiento de las características individuales de cada paciente; y que ha favorecido la «hipermedicalización» de la sanidad, aumentando el número de pruebas, tratamientos y fármacos que se administran, con un elevadísimo coste y sin evidencia sólida de su eficacia<sup>32</sup>.

Es muy importante destacar que los defensores de la medicina basada en la evidencia no niegan la pertinencia de muchas de estas objeciones y reconocen que se trata de pro-

blemas reales y graves. La diferencia entre los críticos y los defensores radica en que estos últimos consideran que los problemas señalados no son inherentes a este enfoque, sino que se deben al mal uso que se ha hecho de sus principios y de su metodología. Ioannidis, por ejemplo, afirma que la medicina basada en la evidencia ha sido «secuestrada» por intereses espurios, básicamente económicos y académicos corporativos, hasta el punto de que más que de medicina basada en la evidencia habría que hablar de medicina basada en el dinero (*finance-based medicine*, Ioannidis, 2016: 85; y 2017: 13). Otros autores subrayan como virtud de la medicina basada en la evidencia que han sido sobre todo sus propios partidarios los que han denunciado los problemas que la aquejan, y que, siempre que se siga de manera rigurosa, el propio enfoque sigue siendo el mejor medio para combatirlos (Guyatt, 2017). Se ha reconocido también que, a lo largo de su desarrollo, la medicina basada en la evidencia ha olvidado algunos de los principios básicos que establecieron sus fundadores, entre ellos la relevancia de la experiencia y del juicio clínico sobre la situación y preferencias de cada paciente individual, y la importancia de otras formas de conocimiento además de los ensayos clínicos aleatorizados cuando las características de la enfermedad, del paciente o del conocimiento existente así lo aconsejen (Richardson, 2017). Con todo, la situación ha llegado a un punto en el que los problemas denunciados son de tal magnitud que resulta dudoso si será posible corregir las malas prácticas<sup>33</sup>.

### 3.3.2. La aplicación en otros ámbitos del modelo «basado en la evidencia»

Tener presente la evolución en la medicina ayuda a comprender algunas de las críticas que se han formulado a la pretensión de construir todo tipo de políticas públicas sobre el modelo «basado en la evidencia».

Se ha objetado, por ejemplo, que parte de un entendimiento excesivamente linear de la relación entre el conocimiento científico y la práctica política, según el cual para cada problema social la ciencia sería capaz de detectar «la mejor» solución y, una vez identificada, simplemente habría que implementarla (Greenhalgh y Russell, 2009: 305). Esta manera de presentar las cosas desconoce que muchas veces la evidencia existente sobre determinada cuestión no es unívoca, pero además refleja un entendimiento positivista del conocimiento científico para el que los problemas sociales estarían «ahí fuera», esperando a que alguien los detecte y les dé una solución, cuando en realidad la identificación de un problema, su definición y la prioridad que se le asigne en la agenda política dependen de una construcción social en la que se enfrentan valores e intereses contrapuestos de distintos grupos y diversas concepciones ideológicas (Greenhalgh y Russell, 2009: 315)<sup>34</sup>. Acabamos de ver cómo incluso sobre un campo en principio estrictamente objetivo y técnico, como la medicina clínica y la investigación médica, confluyen intereses muy diversos: médicos que quieren curar a sus pacientes, industrias farmacéuticas que quieren ganar dinero, gestores de sistemas sanitarios que quizá priorizan el ahorro de costes, investigadores preocupados por mejorar sus *curricula* y promocionar académicamente, revistas científicas que quieren posicionarse como las de mayor impacto, etc. Los

diversos intereses de todos estos actores influyen en lo que se investiga y lo que no, en lo que se publica y lo que no, en lo que se financia y lo que no e incluso en la determinación de lo que es una enfermedad y lo que no lo es<sup>35</sup>. En el ámbito de la política todavía es mucho más acusada la influencia que ejercen las ideologías dominantes y los equilibrios de poder entre diversos grupos sociales a la hora de identificar un problema y establecer prioridades de actuación.

Por otro lado, el énfasis que ponen los partidarios del *evidence-based policymaking* en la necesidad de adoptar las políticas «que funcionan» («what works») es engañoso al menos por dos razones. Por un lado, porque tiende a presentar los problemas como meras cuestiones técnicas, oscureciendo la naturaleza política de muchas de las cuestiones a resolver: del hecho de que para algo exista una solución sobre cuya eficacia disponemos de evidencia científica no significa que ese algo sea un problema importante que hay que resolver; la investigación científica puede proporcionar evidencia sobre cuáles son las consecuencias de determinadas políticas, pero no es capaz por sí sola de establecer cuál de esas políticas debe considerarse preferible<sup>36</sup>. En las decisiones políticas hay otros factores que legítimamente condicionan las decisiones, además del conocimiento científico (Klein, 2000; y Parkhurst, 2016) y sobre ellas debe haber un debate abiertamente político, no (solo) técnico.

Por otro lado, la insistencia en implementar las políticas respecto a las que existe evidencia científica de que «funcionan» opera con generalizaciones que pueden ser legítimas en ciencias como la medicina, pero que no tienen por qué serlo en otros ámbitos. Los cuerpos de todos los seres humanos comparten unas características anatómicas y fisiológicas esencialmente idénticas. Por ello, un medicamento que ha demostrado ser efectivo para bajar la fiebre en una muestra de pacientes de Wyoming producirá básicamente el mismo efecto en pacientes de Berlín. Sin embargo, no tiene por qué ocurrir lo mismo respecto de intervenciones cuyos efectos están mediados por factores contextuales que pueden diferir radicalmente en cada momento y lugar (Parkhurst, 2016: 20-22). Un programa de innovación educativa para reducir el fracaso escolar en Berlín no tiene por qué ser igualmente efectivo en Wyoming si las características académicas y socioeconómicas de los estudiantes y de sus familias no son similares. E incluso si lo fueran, dos individuos que comparten un perfil parecido no reaccionan igual ante determinados estímulos o tratamientos ni toman las mismas decisiones porque media siempre la personalidad, las preferencias y las experiencias particulares de cada cual. Dicho de otra manera: desde el punto de vista epistemológico, no es correcto trasladar al ámbito de las políticas públicas el tipo de generalizaciones que sí pueden estar justificadas en medicina.

Otro problema también relacionado con cuestiones de tipo epistemológico es el siguiente: el *evidence-based policymaking* privilegia un determinado tipo de «evidencia científica» (ensayos aleatorizados, mediciones cuantitativas, métodos estadísticos); sin embargo, en las ciencias sociales hay muchos otros tipos de análisis que pueden proporcionar conocimiento relevante sobre la complejidad de las relaciones y tensiones propias de la realidad social<sup>37</sup>. No se trata solo de que no existan ensayos clínicos aleatorizados sobre la mayor parte de políticas sociales, sino también de que la mayor parte de los fenómenos sociales

son imposibles de medir con la exactitud matemática con que, por ejemplo, se mide el efecto de determinado antibiótico sobre la cantidad de bacterias presentes en un organismo. Si, pese a ello, se considera válida (o de calidad superior) solo la evidencia que se haya obtenido con métodos cuantitativos y estadísticos y se excluyen las demás, se produce una monopolización del conocimiento considerado relevante y una sobresimplificación de la realidad (Strassheim y Kettunen, 2014: 263). Como afirman Saltelli y Giampietro: «Una vez que el análisis ha excluido todas las fuentes de conocimiento incómodo, el problema se reduce a uno que puede ser abordado con la usual combinación de análisis coste-beneficio y metodologías de estimación de riesgo, y la solución optimizada al deseado nivel de precisión, aunque a estas alturas la solución pueda haber perdido toda su relevancia para el problema original» (2017: 66).

A todo ello hay que añadir que, ante la situación no infrecuente de que sobre determinada cuestión existan estudios científicos con resultados no coincidentes, el responsable de la adopción de una decisión o de la implementación de una política puede escoger deliberadamente solo aquella evidencia científica que apoye la decisión que previamente ya había decidido adoptar en función de otros criterios e intereses (*cherry picking*).

En definitiva, el (loable) objetivo de diseñar e implementar políticas más transparentes y eficaces por estar basadas en conocimientos sólidos sobre sus efectos y consecuencias no es en absoluto sencillo de conseguir<sup>38</sup>. Los fenómenos sociales y el comportamiento humano —y, por tanto, la política criminal, que es el ámbito que interesa especialmente en este trabajo— son esencialmente diferentes de los fenómenos físicos, por lo que ni pueden ser estudiados (solo) con la misma metodología que es apropiada para estos últimos ni tampoco cabe esperar que una misma política o decisión produzca en contextos o sobre individuos diferentes (siempre) el mismo efecto que produjo en otros contextos o sobre otros individuos<sup>39</sup>.

La relación entre ciencia y política es compleja y bidireccional, y existe el peligro de hacer pasar por «políticas basadas en la evidencia» lo que en realidad no es más que «construcción de evidencia basada en la política» (*policy-based evidence making*), es decir, la utilización selectiva de los datos para legitimar con esta supuesta evidencia científica objetivos políticos preestablecidos (Strassheim y Kettunen, 2014: 262; Parkhurst, 2016: 48 ss.). Además, existe también el peligro de suplantar el debate abierto y legítimo sobre opciones ideológicas y morales contrapuestas ocultando la dimensión política de los argumentos en liza tras la aparente neutralidad de mejores o más consistentes apoyos científicos para alguna de las opciones en disputa.

Teniendo en cuenta lo expuesto hasta aquí, interesa analizar hasta qué punto el *evidence-based sentencing* realmente incorpora «la mejor evidencia disponible» sobre el control del crimen, y si está científicamente acreditado que el diseño de las sanciones penales sobre la base de las estimaciones de riesgo de reincidencia es una política «que funciona».

## 4. EL EVIDENCE-BASED SENTENCING Y LA EVIDENCIA CIENTÍFICA

### 4.1. LA CALIDAD DE LA EVIDENCIA SOBRE LA QUE SE BASA EL EVIDENCE-BASED SENTENCING

La evidencia científica sobre la que se basa el EBS consiste fundamentalmente en las estimaciones estructuradas de riesgo de reincidencia<sup>40</sup>. Y la opinión probablemente mayoritaria considera que dichas estimaciones ofrecen información sólida y contrastada. Se afirma repetidamente que las estimaciones de riesgo modernas, realizadas con métodos estructurados (ya sean estos puramente actuariales o de juicio clínico estructurado)<sup>41</sup>, son mucho más fiables que las que se hacían con el mero juicio clínico del profesional, y que actualmente las tasas de acierto son más elevadas que hace treinta o cuarenta años<sup>42</sup>. Sin embargo, si bien es cierto que ha habido avances importantes en la estimación del riesgo de violencia desde los años 80, la mera afirmación de que la «capacidad predictiva» de los modernos instrumentos estructurados es «mejor» que los juicios clínicos de los profesionales simplifica demasiado la cuestión y, por ello, transmite una imagen excesivamente optimista que puede inducir a confusión.

Para situar el problema conviene recordar que lo que contribuyó decisivamente al des crédito de los viejos juicios clínicos de peligrosidad en las décadas de los setenta y los ochenta fue la constatación de que producían un determinado tipo de errores, los falsos positivos, en una proporción que en aquel momento fue considerada excesiva. Diversos estudios demostraron que solo uno de cada tres sujetos considerados peligrosos por un psiquiatra o un psicólogo cometía actos violentos después (Monahan, 1981: 77). El hecho de que dos de cada tres predicciones de conducta violenta resultaran desmentidas por el comportamiento posterior del sujeto se consideró una evidencia de calidad insuficiente como para fundamentar sobre ella restricciones importantes de derechos como decretar o prolongar internamientos psiquiátricos o penitenciarios.

Pues bien, en la actualidad esa clase de errores no ha disminuido de manera significativa. El indicador estadístico que mide el porcentaje de sujetos que, respecto del conjunto de los que se habían valorado como de riesgo alto, efectivamente reinciden se denomina valor predictivo positivo. En todos los estudios sobre reincidencia violenta que se han llevado a cabo a nivel internacional, dicho valor suele estar por debajo del 50 % y, con frecuencia, muy por debajo (Douglas *et al.*, 2017). En un relevante y conocido metaanálisis, que revisó la aplicación de nueve herramientas de valoración del riesgo en 73 estudios que incluían a más de 24.000 sujetos, el valor predictivo positivo para la estimación de riesgo de reincidencia violenta fue, de media, un 41 %; es decir, que de cada 10 sujetos a los que se había considerado de riesgo alto solo 4 cometieron nuevos hechos violentos. En ese mismo estudio, el valor predictivo positivo de las herramientas que estiman el riesgo de reincidencia sexual fue aún más bajo: del 23 % en promedio (Fazel *et al.*, 2012: 10). Estudios publicados más recientemente siguen evidenciando valores predictivos positivos para la probabilidad de reincidencia violenta que ni siquiera alcanzan el 40 %<sup>43</sup>.

Teniendo en cuenta estos datos, no parece que, al menos en este aspecto, las modernas estimaciones estructuradas de riesgo hayan mejorado mucho en cuanto a su capacidad

predictiva<sup>44</sup>. ¿De dónde proviene, entonces, el convencimiento generalizado de que son mucho mejores que los juicios clínicos de peligrosidad de hace cuarenta años? Entre las principales razones cabe destacar, a mi juicio dos: por un lado, que ha cambiado la forma en que se interpreta lo que es una buena capacidad de estimación del riesgo, y por otro, que también ha cambiado la manera en que se comunica la información sobre la calidad de las estimaciones.

En relación con la primera cuestión, antiguamente las predicciones de peligrosidad se consideraban binarias, en el sentido de que tanto el evento futuro como la predicción tenían únicamente dos posibilidades: al sujeto se le consideraba peligroso o no peligroso y, o bien cometía nuevos delitos, o bien no lo hacía. En este tipo de predicciones solo hay 4 resultados posibles: verdadero negativo (se le consideró no peligroso y no reincidió), falso negativo (se le consideró no peligroso pero sí reincidió), verdadero positivo (se le consideró peligroso y reincidió) y falso positivo (se le consideró peligroso pero no reincidió).

A partir de la década de los noventa, sin embargo, empezó a diferenciarse entre la ocurrencia o no del evento y los diferentes grados de confianza que uno puede tener en que dicho evento acaezca (Mossman, 2006: 549 ss. y 555 ss.). Es decir, aunque la reincidencia sea un evento binario<sup>45</sup>, el juicio que uno hace sobre la probabilidad de su ocurrencia no lo es, ya que el evaluador puede considerar que dicha probabilidad es no solo alta o baja, sino también muy baja, baja, media, alta, muy alta, extrema, etc. Por ejemplo, si una herramienta de valoración del riesgo clasifica como de riesgo alto a un grupo de individuos que posteriormente evidencian una tasa de reincidencia del 30 %, como de riesgo medio a un grupo en el que la reincidencia alcanza el 20 %, y como de riesgo bajo a otros en que esta se queda en el 10 %, parece evidente que sí es capaz de distinguir qué grupos de individuos presentan un riesgo mayor que otros. En ese sentido, puede afirmarse que su capacidad predictiva es «buena», al menos por lo que hace a su capacidad para discriminar el mayor o menor riesgo relativo de unos grupos respecto de otros. Y ello aunque en el grupo de riesgo más alto el porcentaje de los sujetos que realmente reincidieron solo llegue al 30 %.

Pues bien, en este aspecto, el de la estimación del riesgo relativo, es donde se han producido los avances más significativos de la investigación criminológica. Cuando los niveles de riesgo posibles ya no son solo dos, resulta insuficiente analizar la capacidad predictiva solo en términos de falsos positivos o falsos negativos. En este sentido, se han desarrollado otros indicadores más sofisticados, entre los que el más utilizado es probablemente la llamada área bajo la curva ROC<sup>46</sup>. Se trata de un indicador de riesgo relativo que mide la probabilidad de que un sujeto reincidente, escogido al azar, haya recibido en la herramienta una valoración de riesgo superior a la de un individuo no reincidente, igualmente escogido al azar (Singh, 2013). Es decir, informa sobre cuán adecuadamente diferencia la herramienta entre los sujetos de mayor y de menor riesgo, pero sin decir nada sobre cuáles son las probabilidades de riesgo asociadas a cada nivel. Por ejemplo, en una determinada herramienta el grupo de bajo riesgo puede tener asociada una probabilidad de reincidencia del 5 % y el de alto riesgo del 15 %, y tener un área bajo la curva ROC de 0.75 (que convencionalmente se considera un valor elevado); en otra herramienta, las probabilidades de reincidencia pueden ser del 10 % y del 50 %, respectivamente, y tener el mismo valor

de área bajo la curva ROC de 0.75. En ambas hay un 75 % de probabilidad de que un individuo reincidente seleccionado al azar haya tenido una clasificación de riesgo superior a un individuo no reincidente seleccionado también al azar. Pero ser de alto riesgo tiene asociada una probabilidad de reincidencia muy diferente en cada una de estas herramientas y, desde el punto de vista de un juez que quisiera imponer una sentencia adaptada al nivel de riesgo del sujeto, tiene un significado muy diferente que la probabilidad de reincidencia asociada al nivel de «riesgo alto» sea del 15 % o del 50 %<sup>47</sup>.

Por otro lado, el área bajo la curva sintetiza en una sola cifra (por ejemplo, 0.75) la capacidad de discriminación del instrumento, pero las tasas de falsos positivos y de falsos negativos pueden variar drásticamente según cuál sea el umbral de discriminación que se utilice a efectos de adoptar una concreta decisión. Por ejemplo, supongamos que para conceder o no la suspensión de la pena de prisión utilizáramos una herramienta de valoración del riesgo que clasifica a los sujetos en cinco niveles de riesgo (muy bajo, bajo, medio, alto, muy alto) y que tiene un área bajo la curva ROC de 0.75. Si optamos por conceder la libertad condicional solo a los clasificados como de riesgo muy bajo, tendremos muchos falsos positivos y muy pocos falsos negativos; si, por el contrario, concedemos a todos la libertad salvo a los del grupo de riesgo muy alto, tendremos muchos más falsos negativos y menos falsos positivos. Ninguna de dichas tasas de errores tiene por qué coincidir con la cifra del área bajo la curva (75 %), pudiendo estar tanto los falsos positivos como los falsos negativos muy por encima o muy por debajo del 75 % en cada una de nuestras decisiones, en función del criterio con el que hayamos utilizado en cada caso la herramienta<sup>48</sup>. Por ello, cuando algunos autores partidarios del EBS exigen que las evaluaciones de riesgo sobre las que se vaya a basar una sentencia se realicen con herramientas que alcancen valores de área bajo la curva de 0.75 para que pueda considerarse que son suficientemente fiables y que miden bien lo que pretenden medir (así, *v. gr.* Slobgin, 2018: 589), están, a mi juicio, algo desencaminados, porque ese valor de área bajo la curva ROC no significa ni que el 75 % de las clasificaciones de riesgo se hayan visto corroboradas por el comportamiento posterior del sujeto ni que los sujetos clasificados como de alto riesgo tengan un 75 % de probabilidades de reincidir<sup>49</sup>.

Por todo ello, diversos autores han advertido que el valor de este indicador estadístico por sí mismo es muy poco informativo a la hora de enjuiciar la utilidad de una herramienta de valoración del riesgo para la toma de decisiones legales (Fazel, 2019: 198; Szmuckler, Everitt y Leese, 2011; y Shepherd y Sullivan, 2017), y recomiendan que se complemente el análisis de la capacidad predictiva incluyendo también información sobre otros indicadores estadísticos, y que se informe siempre de las limitaciones de todos ellos (Singh, 2013; Rossegger *et al.* 2014; o Douglas *et al.*, 2017, que aconsejan que se informe con claridad sobre los números de falsos positivos y falsos negativos).

Esta última observación enlaza con la otra razón que, como hemos avanzado más arriba, explica el optimismo generalizado en torno a la capacidad predictiva de las herramientas estructuradas de estimación del riesgo: la forma en que se comunica la información sobre su rendimiento. La capacidad predictiva de estas herramientas puede expresarse con muchos indicadores estadísticos diferentes, cada uno de los cuales mide una dimensión

distinta de esa capacidad predictiva (Singh, 2013; Loinaz, 2017: 87 ss.; y Muñoz Vicente y López-Ossorio, 2016). Por ejemplo, un mismo instrumento puede tener una sensibilidad muy elevada pero una especificidad baja, o un área bajo la curva ROC aceptable y, sin embargo, un valor predictivo positivo muy débil<sup>50</sup>. En otras palabras, en ocasiones la capacidad predictiva de un instrumento de valoración del riesgo puede ser calificada como «buena» y «mala» a la vez si algunos de los indicadores alcanzan valores muy satisfactorios y, sin embargo, otros se quedan en niveles mucho más modestos.

Si esto es así, resulta de primordial importancia que la información que se ofrezca sobre la capacidad predictiva de estas herramientas abarque las diversas dimensiones posibles, pues solo de esta manera se proporciona un cuadro completo de su rendimiento; de otro modo, si la información incluye solo aquellos parámetros o indicadores que arrojan valores más elevados y omite otros que resultan menos satisfactorios, se fomenta la impresión de que su funcionamiento global es mejor de lo que realmente es. Sin embargo, es muy frecuente que los artículos y estudios científicos que se publican sobre esta materia informen solo de indicadores de riesgo relativo (y, especialmente, del área bajo la curva ROC) y señalen expresamente, además, que estos alcanzan valores aceptables o satisfactorios, pero no den información sobre el riesgo absoluto, es decir, sobre la probabilidad de reincidencia o de reincidencia violenta asociada a cada nivel de riesgo ni ofrezcan (normalmente) los valores predictivos positivo y negativo asociados a cada umbral de discriminación.

En relación con lo que acaba de señalarse, hay otra cuestión importante desde el punto de vista de la justicia penal: hasta qué punto las probabilidades de reincidencia asociadas a cada nivel de riesgo se corresponden verdaderamente con las tasas de reincidencia reales observadas en la práctica (lo que se conoce como correcta calibración del instrumento). En este punto, la literatura científica publicada hasta la fecha presenta importantes carencias. En primer lugar, muchas herramientas de valoración del riesgo no han sido validadas externamente en una muestra diferente de la que se utilizó para construirlas (Douglas *et al.*, 2019: 153). Además, cuando sí existen estudios de validación, estos no siempre respetan las características esenciales del estudio original (en cuanto a perfil de la muestra, periodo de seguimiento, variable criterio, etc.), circunstancia que dificulta la interpretación de los resultados como confirmación o refutación del mismo (Rossegger *et al.*, 2013). La investigación existente hasta ahora sobre indicadores de calibración es aún muy escasa (Fazel, 2019: 201; y Hanson, 2017: 27), pero sí hay evidencia al menos de dos fenómenos preocupantes: por una parte, que las tasas de reincidencia correspondientes a cada nivel de riesgo varían notablemente entre unas muestras y otras (Singh *et al.*, 2014; y Rossegger *et al.*, 2014) y, por otra, que cuando las valoraciones de riesgo se aplican en la práctica del día a día de la justicia penal los niveles de acierto son más bajos que los evidenciados en los estudios académicos (Fazel, 2019: 203). Por todo ello, es muy dudoso que la evidencia científica disponible hasta el momento permita generalizar las probabilidades de reincidencia obtenidas en uno o unos pocos estudios científicos a poblaciones diferentes de aquellas en las que estos se han llevado a cabo.

Por todos estos problemas, y algunos otros que también presentan las valoraciones de riesgo<sup>51</sup>, desde la Criminología se ha reconocido que «[...] no solo es que la capacidad

predictiva de las herramientas de valoración del riesgo sea imperfecta, es que también se presenta de manera imperfecta en la literatura. Esta base de evidencia limitada y sesgada crea el peligro de que quienes deben tomar las decisiones confíen en las puntuaciones de riesgo más de lo que su precisión garantiza» (Douglas *et al.*, 2017: 135)<sup>52</sup>.

En definitiva, las modernas estimaciones estructuradas de riesgo consiguen identificar con elevados niveles de acierto a los sujetos de bajo riesgo y también discriminan bastante bien dentro de un grupo a los sujetos con mayor riesgo que otros. Sin embargo, a pesar del aumento de las publicaciones sobre la materia y de la mayor sofisticación de los indicadores y cálculos estadísticos, hoy por hoy no ha mejorado sustancialmente la capacidad para identificar a los sujetos con mayor probabilidad de cometer nuevos delitos violentos sin incurrir en elevadas cifras de falsos positivos. Y aunque a veces son incluso los propios criminólogos quienes alertan de este problema, y de las implicaciones que de él derivan si se pretende utilizar estas estimaciones para fundamentar en ellas medidas gravemente restrictivas de derechos<sup>53</sup>, lo cierto es los defensores del *evidence-base sentencing* tienden a obviarlo cuando afirman de manera genérica que las estimaciones estructuradas de riesgo de reincidencia son «mucho mejores» que el antiguo juicio clínico estructurado y que, por ello, deben ser utilizadas a la hora de elegir el tipo de sanción y la forma de ejecutarla.

¿Cabe esperar que una mayor sofisticación de los cálculos matemáticos favorecida por las posibilidades que ofrece la inteligencia artificial produzca una mejora apreciable de este estado de cosas en un futuro cercano? Aunque soy consciente de las cautelas con las que es necesario responder a este tipo de cuestiones<sup>54</sup>, tiendo a pensar que no, dado que ningún algoritmo, por complejo que sea, puede dar resultados de mayor calidad que la de los datos con que se lo «alimenta». Si los datos con que trabaja la máquina son incompletos o incorrectos, las estimaciones que con ellos se hagan también lo serán<sup>55</sup>, y este es un problema importante en las valoraciones de riesgo de reincidencia, pues si bien algunos de los factores pueden comprobarse de manera directa y con un margen de error mínimo (por ejemplo, la edad o el sexo del sujeto), hay muchos otros sobre los que solo se dispone de información incompleta, indirecta, o previamente «contaminada» con un criterio subjetivo.

Así ocurre con uno de los factores que se consideran más decisivos en cualquier herramienta de valoración del riesgo: el historial criminal. El número real de delitos cometidos en el pasado por una persona no tiene por qué coincidir con el número de antecedentes penales que consten en los registros oficiales, bien porque algunos nunca hayan llegado a descubrirse, bien porque haya sido condenado por algunos que nunca cometió (por ejemplo, debido a conformidades<sup>56</sup>). Se trata de un problema difícil de resolver porque sobre este punto ni siquiera una entrevista directa con el sujeto es una fuente fiable, ya que, al tratarse de una materia tan sensible, es probable que al preguntarle mienta para mostrar una imagen que cree que le favorecerá en la valoración<sup>57</sup> o haya olvidado algún incidente. También otro tipo de factores —por ejemplo, si tiene o no amigos en prisión o cuántas relaciones de pareja estables ha habido en su vida— son igualmente difíciles de comprobar. Y otros —por ejemplo, si tiene o no expectativas viables de futuro, mucho o poco autocontrol, su grado de tolerancia a la frustración, etc.— dependen de valoraciones que tiene

que hacer un tercero y que, por tanto, incorporan en mayor o menor grado apreciaciones subjetivas.

Estos problemas aquejan ya en la actualidad a las valoraciones convencionales de riesgo, y no desaparecerán aunque sean cada vez más sofisticados los algoritmos que se usen para procesar la información si esta sigue presentando las deficiencias de partida señaladas. Se trata de una cuestión que merece la pena destacar porque, cuando se discute sobre las potencialidades de la inteligencia artificial (o, en general, del tratamiento masivo de datos de manera cada vez más automatizada) para el control del delito, se tiende a dar por supuesto que proporcionará una información mejor, más completa y más precisa, que permitirá, a su vez, mejorar la eficacia de las intervenciones. Se reconoce después que esa ventaja técnica, que se da por descontada, puede generar fricciones y problemas en otros ámbitos, especialmente en todo lo relacionado con los derechos a la intimidad y al control sobre el tratamiento de datos personales<sup>58</sup>. Planteado así el problema, la cuestión se presenta entonces como una ponderación entre el coste que supone asumir restricciones en estos (y otros) derechos fundamentales a cambio de disponer de herramientas más eficaces para el control del delito. Lo que me gustaría subrayar es que esa mayor eficacia —que no niego que sea posible en algunos casos, quizás incluso en muchos— no puede darse por descontada a la vista del tipo de información con la que operan los algoritmos cuando se trata de delincuencia. Por ello, antes de empezar a ponderar y a plantearnos la posibilidad de asumir restricciones en el ámbito de aplicación de algunas garantías y derechos, sería conveniente exigir evidencia científica de que realmente la inteligencia artificial alcanza mejores resultados. Por lo que hace a las valoraciones de riesgo de reincidencia, creo que hoy en día no puede afirmarse que se haya aportado (¿aún?) evidencia de esa clase por las razones expuestas en este epígrafe y las que a continuación se añadirán.

#### 4.2. ¿HAY EVIDENCIA EMPÍRICA DE QUE EL *EVIDENCE-BASED SENTENCING* ALCANCE SUS OBJETIVOS?

Es relativamente frecuente que se esgriman como evidencia del éxito del EBS argumentos tales como que los jueces y el resto de operadores del sistema jurídico consideran que la información aportada por las valoraciones de riesgo es «útil», o que como consecuencia de su introducción se ha reducido el número de condenas a prisión y ha aumentado el de penas alternativas<sup>59</sup>. Todos estos resultados pueden merecer sin duda una valoración positiva. Sin embargo, no aportan información directa sobre los dos objetivos principales del EBS que son, recordemos, la reducción de la reincidencia y la de los costes asociados al uso excesivo de la prisión, junto con el control de las tasas de delincuencia<sup>60</sup>. En cuanto «política basada en la evidencia» que dice ser, el *evidence-based sentencing* debería poder demostrar con evidencia empírica que esos resultados efectivamente se consiguen; en otras palabras, que «funciona» («it works»).

Sin embargo, y como hemos visto en el epígrafe anterior, existen abundantes estudios acerca de la validez predictiva de las herramientas de valoración del riesgo, mientras que son muchos menos los que han estudiado su utilidad para reducir las tasas de reincidencia

o de criminalidad (Viljoen, Cochrane y Jonnson, 2018: 184). Y los que existen no han generado resultados muy alentadores para el EBS. Tras una revisión sistemática que incluyó estudios publicados hasta 2017, Viljoen, Cochrane y Jonnson concluyeron que no hay suficiente evidencia empírica para afirmar que el uso de valoraciones estructuradas de riesgo reduzca la violencia o la reincidencia, ya que los estudios publicados al respecto, además de presentar numerosas e importantes limitaciones metodológicas, arrojan resultados dispares: si, en algunos casos, tras el uso de valoraciones de riesgo se observa una reducción en las tasas de violencia o delincuencia, en otros no es así (2018: 200, 204). Merece la pena destacar, además, que 8 de los 12 estudios analizados en esta revisión se habían realizado sobre muestras de pacientes psiquiátricos (Viljoen, Cochrane, Jonnson, 2018: 198), lo que plantearía importantes problemas para extraer los resultados a la población estándar de delincuentes, aun en el caso de que sí se hubiese demostrado una asociación entre el uso de valoraciones estructuradas de riesgo y descensos en la delincuencia.

Recientemente se han publicado también los resultados de una investigación empírica sobre la aplicación del EBS en el estado de Virginia<sup>61</sup>. Virginia fue el primer estado en los EE. UU. en implementar de manera sistemática valoraciones de riesgo en el momento de dictar la sentencia: las introdujo en 2003 para delitos no violentos y para delincuentes sexuales. Las estimaciones de riesgo se integraron en las *sentencing guidelines* del estado con los siguientes objetivos: *i)* imponer penas alternativas a la prisión a un número significativo de delincuentes no violentos de bajo riesgo, y *ii)* permitir la imposición de penas más largas a los delincuentes sexuales de alto riesgo. Y en ambos supuestos, las estimaciones de riesgo se introdujeron solo como una recomendación para el juez, que decide si quiere seguirlos o no. Esta política estaba presidida, tal como es característico en el EBS, por consideraciones de mayor efectividad en términos de coste-beneficio: para reservar las costosas plazas en prisión a los delincuentes más violentos y mantener al mismo tiempo elevados niveles de seguridad pública<sup>62</sup>.

La investigación a la que nos referimos analiza cuál ha sido el impacto de estos cambios en las tasas de población penitenciaria y de reincidencia, y los resultados son dignos de consideración. Por un lado, en el grupo de delincuentes no violentos no se ha conseguido que disminuya la tasa de población penitenciaria ni el número de penas de prisión impuestas. Esto no significa, sin embargo, que los jueces no hayan tenido en cuenta las recomendaciones derivadas de los niveles de riesgo. Según el estudio, sí cabe apreciar diferencias entre la probabilidad de ser condenado a prisión, y también en la duración de la pena, entre los delincuentes no violentos que quedan por debajo y por encima del punto de corte que marca la frontera entre riesgo alto y bajo. Lo que ocurre es que la disminución del número de penas de cárcel y de la duración de las mismas para el grupo de bajo riesgo se ha visto compensada por el aumento de ambos factores en el grupo de alto riesgo, con lo que el efecto neto de la implementación de estas políticas sobre el número total de internos acaba siendo nulo (Stevenson y Doleac, 2019: 2, 3 y 19).

Cabría pensar que, aunque no se haya conseguido reducir la población penitenciaria, al aumentar la dureza de las penas para los delincuentes de mayor riesgo al menos sí se habría logrado la segunda de las metas perseguidas: la reducción de la reincidencia. Los

datos evidencian, sin embargo, que tampoco es así (*Ibid.*: 2 y 20): las tasas de reincidencia no experimentaron ningún cambio significativo.

Por otro lado, también llaman la atención los resultados en el grupo de los delincuentes sexuales: si para estos casos el propósito explícito de la reforma era permitir solo incrementos de la dureza de las penas por encima de lo recomendado por las *guidelines* cuando el riesgo fuera alto, lo que el estudio evidencia es que, tras introducir las valoraciones de riesgo, se produjo una disminución del 5 % en la probabilidad de ser encarcelado y aproximadamente un descenso del 24 % en la duración de las condenas (*Ibid.*: 19).

El estudio apunta diversas explicaciones para estos sorprendentes resultados. Una muy relevante es que los jueces no siguieron las recomendaciones en todos los casos (que, recordese, no eran vinculantes): además de las *guidelines* y de las valoraciones de riesgo, tuvieron en cuenta también otros criterios, entre los que sobresale la edad. Ser joven es uno de los factores que más incrementa el riesgo en cualquier instrumento de valoración, también en el aplicado en Virginia. Sin embargo, los tribunales han valorado tradicionalmente la juventud como una atenuante de la responsabilidad y, con ella, de la pena. El estudio demuestra que en Virginia los jueces sí modificaron las sentencias en este punto para adaptarlas a lo que recomendaban las valoraciones de riesgo, pues se constató un aumento relativo de la gravedad de las penas impuestas a los jóvenes y también de la probabilidad de que estos recibieran una pena de prisión. Pero la adaptación fue solo parcial, dado que si los jueces hubieran seguido de manera consecuente las recomendaciones en todos los casos, dichos incrementos se habrían multiplicado (*Ibid.*: 3 y 4).

En cuanto al grupo de los delincuentes sexuales, el estudio sugiere dos posibles explicaciones. La primera es que los jueces tuvieran la idea preconcebida de que este grupo de delincuentes presenta más riesgo de reincidencia del que realmente tiene y gracias a las valoraciones de riesgo hayan reparado en que son menos peligrosos de lo que pensaban, lo que explicaría la reducción en la gravedad de las condenas. Sin embargo, los autores del trabajo consideran más plausible otra hipótesis<sup>63</sup>: que las valoraciones de riesgo bajo hayan sido utilizadas como un «escudo» por los jueces para imponer castigos —a su juicio más adecuados— que antes no se atrevían a imponer, habida cuenta de los tremendos costes para su prestigio y su carrera profesional que suponen los falsos negativos. Disponer de una evaluación que indica riesgo bajo permitiría descargar en ella parte de la responsabilidad en caso de nueva comisión de delitos graves por quien podría haber estado en prisión si hubiera recibido una pena más larga (*Ibid.*: 19-20).

En cuanto al hecho de que tras la implementación de las valoraciones de riesgo la reincidencia se haya mantenido constante, los autores del estudio descartan como posible explicación que la herramienta usada en Virginia sea deficiente o esté mal diseñada, y apuntan simplemente al hecho de que la reincidencia es un fenómeno muy difícil de predecir en cualquier circunstancia. A su juicio, las estimaciones de riesgo explican solo un porcentaje mínimo de la reincidencia, por lo que la adopción de unas u otras medidas sobre la base de esas estimaciones solo puede producir efectos muy limitados en la variación de las tasas de reincidencia (*Ibid.*: 33 ss.)<sup>64</sup>.

#### 4.3. DISCUSIÓN

Hasta ahora no hay evidencia científica que demuestre que el *evidence-based sentencing* logre los objetivos que pretende alcanzar: reducir la reincidencia y la tasa de ocupación penitenciaria. Podría ser, no obstante, que esta ineeficacia no derivase de insuficiencias intrínsecas al EBS, sino de la mala aplicación que de él se hace en la práctica. El estudio de Stevenson y Doleac podría apuntar en este sentido: dado que en el caso de Virginia las recomendaciones de riesgo no eran vinculantes, los jueces no siempre las siguieron; si lo hubieran hecho, los resultados serían mejores. Por tanto, lo que hay que hacer es eliminar la discreción judicial y obligar a configurar siempre las sanciones en función de lo que aconsejen los niveles de riesgo.

Sin embargo, entre los partidarios del EBS es generalizada la opinión de que las valoraciones de riesgo no deben ser vinculantes para los jueces, sino que estos han de conservar su autonomía para valorar todas las circunstancias concurrentes y tomar la decisión que les parezca más apropiada<sup>65</sup>. Desde este enfoque suele reconocerse que, además del nivel de riesgo, puede haber otros factores importantes a tener en cuenta en la ejecución de la pena, entre ellos la disponibilidad de recursos para llevar a cabo los tratamientos o el tipo de supervisión indicado. Y también se afirma mayoritariamente que el propósito de reducir la reincidencia es uno más entre los diversos fines que persigue el Derecho penal, por lo que criterios tales como la gravedad de la culpabilidad, la gravedad del hecho o las necesidades de prevención general pueden legitimar la imposición de penas no ajustadas al riesgo de reincidencia del sujeto (Casey, Warren y Elek, 2011: 11-14). Por lo que hace al ámbito académico, la mayoría de los partidarios del EBS lo encuadran en un retribucionismo limitado que establece límites máximos (y en ocasiones mínimos) más allá de los cuales no podrían imponerse sanciones, por mucho que pudieran resultar adecuadas en función de los niveles de riesgo (*v. gr.* Monahan y Skeem, 2014: 158; y Slobogin, 2019: 108 ss.).

Ello plantea la siguiente paradoja: si para lograr los beneficios que pretende el EBS (en términos de reducción del crimen y de la reincidencia) hay que seguir estrictamente las recomendaciones basadas en el nivel de riesgo, pero al mismo tiempo se reconoce que es imposible hacer esto último porque hay buenas razones que desaconsejan hacerlo y que hay que respetar, parece que los propios defensores de la práctica están reconociendo la imposibilidad de que resulte exitosa<sup>66</sup>.

Por otro lado, es muy dudoso que, aun diseñando las penas única y exclusivamente en atención al nivel de riesgo, pudieran alcanzarse los objetivos perseguidos. Por un lado, porque, como ya hemos visto, la evidencia científica a la que apela el *evidence-based sentencing* es bastante menos sólida de lo que a primera vista puede parecer. Las altas tasas de falsos positivos ponen seriamente en cuestión la eficiencia de un sistema que destinaría sistemáticamente un exceso de recursos al control penal intensivo de sujetos que no lo necesitan. Y la preocupación por la reincidencia que preside todo este enfoque refleja una intolerancia a los falsos negativos que probablemente también favorecería, respecto a los grupos de bajo riesgo, una respuesta penal no acorde con el riesgo real que representan. Por otro lado, además, si estimar con un nivel de acierto elevado el riesgo de reincidencia no es fácil, reducir

ese riesgo es más difícil aún. Como afirman Monahan y Skeem, exceptuando quizá los casos de los programas de desintoxicación y los programas que tratan las distorsiones cognitivas, actualmente no hay evidencia concluyente de que la modificación de determinados factores de riesgo genere un descenso de la reincidencia<sup>67</sup>.

A ello hay que añadir que las valoraciones de riesgo no son el único conocimiento sobre la delincuencia que ha producido la Criminología a lo largo de sus muchos años de historia: por el contrario, en la investigación criminológica podemos encontrar también evidencia científica que apoyaría la implementación de políticas criminales bastante distintas a las que preconiza el EBS. Tomemos el ejemplo de la edad: ser joven es, como ya hemos indicado, uno de los factores que más directamente correlacionan con un mayor riesgo de reincidencia y de reincidencia violenta, y que, por ello, incorporan prácticamente todas las herramientas de valoración. El factor edad puede llegar a explicar casi el 50 % de la puntuación de riesgo en las herramientas estructuradas, y su peso en la puntuación total es, en muchas ocasiones, igual o incluso mayor que el del historial criminal<sup>68</sup>. No obstante, la evidencia empírica sobre carreras criminales y curvas de edad también muestra que, si bien la proporción de personas que cometen delitos en la adolescencia y la primera edad adulta es muy grande, la inmensa mayoría de ellas dejan de hacerlo unos pocos años después, a principios de la veintena. Asimismo, los datos indican que, incluso dentro del grupo de quienes pueden considerarse criminales de carrera, muchos desisten a una edad relativamente temprana (en la treintena)<sup>69</sup>. Siendo esto así, la imposición de condenas largas a personas jóvenes, e incluso condenas muy largas a personas relativamente jóvenes con un amplio historial delictivo, no puede tener más que un muy limitado efecto preventivo-incapacitador, ya que muchos de estos sujetos habrían desistido igualmente del crimen, no habrían continuado delinquiendo y, por tanto, los cuantiosos recursos destinados a su custodia no pueden considerarse en modo alguno una inversión eficiente desde el punto de vista del coste-beneficio. Es decir, incluso manteniéndonos dentro de la lógica utilitaria y «eficientista» propia del EBS, hay evidencia empírica que pone seriamente en cuestión la pretensión de que la imposición de un control penal más intenso (y más caro) a quienes por su edad e historial criminal presentan un alto riesgo de reincidencia, sea en realidad una inversión eficiente (Brandariz, 2016: 236; y Tonry, 2013b: 182).

Adicionalmente, la evidencia que aporta la investigación criminológica sobre carreras criminales y sobre las curvas de edad explica en gran medida los bajos valores predictivos positivos que, como hemos visto *supra*, son un problema endémico de las predicciones de peligrosidad (Tonry, 2013b:183). Cuando se toma una muestra de sujetos condenados y se analiza retrospectivamente su pasado, es fácil que aparezca un largo historial delictivo iniciado, además, a temprana edad, y por ello ambos factores aparecen empíricamente relacionados con la delincuencia. Y algo parecido cabría decir de otros factores de riesgo como los de tipo socioeconómico. Sin embargo, al intentar hacer predicciones prospectivas, los mismos factores producen muchos falsos positivos porque, aunque muchas personas hayan cometido delitos en su juventud, la inmensa mayoría desisten de hacerlo poco después; aunque muchos delincuentes provengan de entornos socioeconómicamente

desfavorecidos, la mayoría de las personas que viven en esos entornos no cometan delitos, y así sucesivamente<sup>70</sup>.

Hay todavía una consideración más que merece la pena hacer en relación con la edad como factor de riesgo y la evidencia científica. Ya hemos comentado antes que en Derecho penal existe una tradición muy asentada de valorar la juventud como un factor de atenuación de la responsabilidad. En todos los países de nuestro entorno existen sistemas de justicia juvenil específicos que funcionan con criterios tendentes a la reeducación y prevén castigos comparativamente más suaves que los correspondientes a los adultos que cometen el mismo delito. Ese estado de cosas se justifica en el dato, evidenciado por las ciencias sociales y, en particular, por la psicología, de que, si bien a partir de cierta edad pueden conocer las normas y en principio están obligados a obedecerlas, los menores son por lo general aún inmaduros, muy influenciables, impulsivos y tienen menor capacidad de tolerancia a la frustración. Esas características psicológicas justifican, si se adopta una perspectiva retributiva, la aplicación de castigos más suaves. En tiempos más recientes, la neurociencia ha confirmado que entre los jóvenes y los adultos no hay solo diferencias psicológicas, sino también que el grado de desarrollo del cerebro es distinto en unos y otros, y que esta diferencia persiste no solo hasta la edad legal que en muchos países marca la frontera para ser tratado como adulto (alrededor de los 18 años), sino también hasta el inicio de la veintena (Sánchez Vilanova, 2017: 208 ss.).

Existe, por tanto, evidencia empírica que justifica la imposición de regímenes penales más intensos a los jóvenes adultos debido al mayor riesgo de reincidencia que presenten, si es que consideramos que la prevención especial negativa es el objetivo central del sistema penal. No obstante, también hay evidencia científica que demuestra que esta estrategia es muy poco efectiva en términos de reducción de la criminalidad debido a la dificultad de identificar al pequeño grupo de los verdaderos delincuentes de carrera. Y también existe evidencia empírica para justificar una respuesta penal más suave para el mismo grupo de personas si consideramos que la retribución proporcionada a la gravedad de la culpabilidad debe ser el criterio central para la imposición de las penas<sup>71</sup>.

## 5. CONCLUSIONES

En el diseño de las políticas criminales, a veces la evidencia científica importa y a veces no. Hay áreas que se han mostrado impermeables al conocimiento contrastado a lo largo de años en las ciencias sociales sobre la nula o muy limitada eficacia y los enormes efectos secundarios de ciertas prácticas (por ejemplo la persistencia de la pena de muerte, la «guerra contra las drogas» o el continuo incremento de la dureza de las penas). Sin embargo, hay otras áreas donde la evidencia científica sí ha logrado penetrar en mayor o menor medida (por ejemplo, en la actividad policial y en algunos sectores de la ejecución penitenciaria de corte rehabilitador). Las principales razones para que el conocimiento criminológico sea o no sea tenido en cuenta no tienen que ver principalmente con su calidad científica, con la solidez de los datos o con la consistencia de las teorías sobre las que descansa, sino que los

factores decisivos suelen ser otros: que se adecue o no a las finalidades políticas perseguidas por los gobiernos en un momento concreto, que coincida con una ventana de oportunidad adecuada para ser bien recibido por la opinión pública, que haya presiones de ciertos grupos de interés a favor o en contra de la incorporación de esa evidencia, etc. (Tonry, 2013a)<sup>72</sup>. Con el *evidence-based sentencing* no está ocurriendo nada distinto: no hay más ni mejor evidencia científica para apoyar una orientación predominantemente preventivo especial del sistema penal basada en estimaciones de riesgo que la que hay para sustentar otros modelos que prioricen otros criterios. Por tanto, el éxito que este enfoque está teniendo en EE. UU. se debe esencialmente a otros factores: que las promesas de eficiencia en términos de coste-beneficio son bien recibidas en momentos de crisis económica; que el énfasis en la reducción de la reincidencia logra concitar acuerdos entre actores políticamente alejados<sup>73</sup>; el discurso tecnocrático y actuarial, que encaja bien en el modelo gerencial actualmente dominante en la política estadounidense; que, al estar centrada en la fase de imposición y de ejecución de las penas, puede implementarse sin cuestionar aspectos ideológicamente más espinosos como la reducción de las penas para delitos graves; o que su aplicación se haya planteado en un momento histórico en que las tasas de delincuencia están en niveles muy bajos.

Ello no impide reconocer que el *evidence-based sentencing* sí se basa, en parte, en conocimientos sólidamente contrastados. Diversos programas rehabilitadores han demostrado reiteradamente buenos resultados en la reducción del riesgo de reincidencia para determinados perfiles de delincuentes. Las herramientas de valoración del riesgo diferencian bastante bien entre grupos con mayor y menor riesgo relativo, e identifican con notable grado de acierto a los grupos de riesgo más bajo de reincidencia violenta. Todo ello constituye conocimiento muy útil para ayudar en la gestión de los programas de tratamiento y, en mi opinión, es un argumento sólido para ampliar el ámbito de las alternativas a la prisión y para destinar recursos a una verdadera infraestructura de acompañamiento de los exreclusos y los liberados condicionales en países que, como el nuestro, carecen de ella.

Pero esto no significa que exista base científica suficiente —ni tampoco, en mi opinión, mejores razones valorativas— para erigir la prevención especial articulada sobre la base del riesgo en la clave de bóveda del sistema penal. Cuando da este salto y lo justifica con el argumento simplista de que hay que adoptar las prácticas «que funcionan», el *evidence-based sentencing* incurre en algunos de los problemas que aquejan a las «políticas basadas en la evidencia» y que han sido descritos en el subepígrafe 3.3.2. de este trabajo. El EBS se presenta como una orientación basada en la «mejor evidencia» disponible, pero omite explicitar las múltiples carencias que (junto a sus indudables virtudes) aquejan a las valoraciones estructuradas de riesgo, por lo que ofrece una imagen en el mejor de los casos incompleta (y, en el peor, sesgada) del verdadero estado del conocimiento científico sobre la capacidad predictiva de las valoraciones de riesgo y sobre sus efectos reales (o la ausencia de efectos) en las tasas de delincuencia. Por otro lado, favorece (ya sea de manera deliberada o inadvertida) una despolitización del debate sobre la justicia penal, al presentar las penas basadas en el nivel de riesgo como una práctica «que funciona» y que, solo por eso, debería ser aceptada, cuando la pregunta crucial es: funciona ¿para qué? El *evidence-based*

*sentencing* asume que el para qué ha de ser la reducción de la reincidencia, pero ni ha sido capaz de demostrar hasta ahora que realmente lo consiga ni es en modo alguno obvio que ese deba ser el objetivo principal del sistema penal. Hay muchos otros fines posibles, y legítimos, como por ejemplo asegurar una aplicación no discriminatoria de las penas, evitar la imposición de castigos desproporcionados o, por supuesto, reducir las tasas de delincuencia, que no es lo mismo que reducir las tasas de reincidencia. El debate sobre cuáles de estos fines son preferibles o prioritarios incorpora necesariamente argumentos de tipo valorativo (normativo, político, ético). Y aunque puede (y debe!, si quiere ser racional) manejar argumentos relacionados con lo que la Criminología sabe sobre los efectos de las penas, no puede ser sustituido por ellos.

Por todo ello, considero que el *evidence-based sentencing*, al menos en la forma que ha adoptado en los últimos años este movimiento en los EE. UU., más que una práctica basada realmente en la evidencia (*evidence-based practice*) es un ejemplo de selección interesada de solo cierta evidencia científica para justificar determinadas políticas públicas que enmascara, tras supuestas razones técnicas, opciones que no pueden dejar de ser objeto de debate político (*policy-based evidence*). Además, al arrogarse el título de «basado en la evidencia» da a entender que otras políticas no estarían basadas en la ciencia o tendrían menos apoyo científico, lo que no es el caso. Por ello, sería preferible referirse a este enfoque con alguna otra de las expresiones que, como ya vimos al inicio del trabajo, describen más apropiadamente su contenido: *predictive sentencing* o *risk-based sentencing*, por ejemplo.

En definitiva, la utilización de evidencia empírica en Derecho penal es imprescindible si quiere llevarse a cabo una política criminal racional. Pero la relación entre política y ciencia es compleja y bidireccional, y en las ciencias sociales las fuentes de conocimiento no se limitan a las dominantes en otro tipo de ciencias. Si bien los datos cuantitativos son imprescindibles, proporcionan una imagen muy incompleta e imperfecta de la realidad. Conviene, por ello, evitar simplificaciones y reduccionismos y no aceptar automáticamente como «mejor» evidencia la que vaya acompañada de montañas de datos y complejos algoritmos. Porque en el debate sobre la automatización del uso masivo de datos para el control de delito (del que el uso de valoraciones de riesgo en la determinación y ejecución de las penas es solo una pequeña parte) está en juego en buena medida el Derecho penal que queremos para el siglo XXI.

## NOTAS

1. No ha ocurrido lo mismo en otros aspectos de la política criminal estadounidense, que sí se han mostrado mucho más receptivos al conocimiento proporcionado por la criminología (Tonry, 2010).
2. Datos tomados del Bureau of Justice Statistics, Key Statistics, «Estimated number of inmates held in local jails or under the jurisdiction of state or federal prisons and incarceration rate, 1980-2016». Disponible en: «[www.bjs.gov](http://www.bjs.gov)». [Consulta: 22/08/2020]. La tasa asciende a 1000/100.000 si se calcula respecto del total de la población adulta residente en los EE. UU.

3. National Research Council (2014). Puede verse un resumen de los principales indicadores hasta 2017 en The Sentencing Project, Criminal Justice Facts, «Fact sheet: Trends in U.S. Corrections». Disponible en: <<https://www.sentencingproject.org/criminal-justice-facts/>>. [Consulta: 22/08/2020.]
4. Es frecuente explicar el abandono de las sentencias indeterminadas a partir de finales de los 70 como un tránsito hacia un modelo inspirado en ideales retributivos. Aunque esta descripción sí encaja con las reformas ocurridas hasta mitad de la década de los ochenta, resulta difícil identificar en las producidas a partir de entonces la lógica fundamental propia de cualquier paradigma retributivo: la proporcionalidad con la gravedad el hecho cometido (Tonry, 2013b: 145 y 150).
5. Si bien ya no se han dictado nuevas leyes tan severas como las anteriores, tampoco se han derogado la mayoría de las existentes. Diversas reformas han limitado o suavizado su alcance en ciertos ámbitos, por ejemplo despenalizando algunos delitos relacionados con drogas y seguridad vial, eliminando penas mínimas obligatorias para infracciones leves de drogas, aumentando los programas alternativos a la prisión pre y postcondena, y extendiendo mecanismos que permiten adelantar el momento de excarcelación como las recompensas por buena conducta (Klingele, 2015: 538; Tonry, 2013b: 144 ss.).
6. *Probation* equivale más o menos a la suspensión de la pena de prisión (con sometimiento del condenado a vigilancia y al cumplimiento de una serie de condiciones durante el periodo estipulado), aunque también incluye supuestos de suspensión de la condena (sin pronunciamiento de culpabilidad).
7. En un listado actualizado a fecha de 2019, Stevenson y Doleac enumeran 28 estados que aplican alguna forma de *evidence-based sentencing*, y otros 7 en los que algún condado también lo utiliza (2019: 54-55).
8. La traducción al castellano sería probablemente «el sentenciar basado en la evidencia» o «sentenciar sobre la base de la evidencia» más que «sentencias basadas en la evidencia». Esta última opción generaría confusión porque, obviamente, toda sentencia está basada en la evidencia, en el sentido de la evidencia probatoria que se ha aportado en el juicio oral, pero el EBS no se refiere a la decisión sobre la inocencia o culpabilidad del acusado, sino a la fase posterior de elección y determinación de la pena adecuada una vez que el acusado ha sido considerado culpable (fase del proceso penal que en el sistema anglosajón está claramente diferenciada de la anterior), y además incluye decisiones relativas no solo a la imposición de la condena, sino también a su ejecución, especialmente las referidas a la posible suspensión y su régimen. En la medida en que en castellano el uso de infinitivos con valor de sustantivo es menos elástico que en inglés, en este trabajo he optado por mantener la expresión en el idioma original, pues creo que no ofrece dificultades de comprensión (especialmente tras las precisiones que se realizan en este epígrafe del trabajo), y cualquiera de las traducciones que he barajado me parecen un poco forzadas.
9. En palabras de uno de los autores que más ha impulsado la adopción de esta perspectiva, el EBS consiste en «[...] sentencing and corrections policies and programs based on the best research evidence of practices shown to be effective in reducing recidivism» (Warren, 2010: 153). Cfr., asimismo, la definición que ofrece en su página web el Center for Sentencing Initiatives: «A set of sentencing practices designed to promote public safety through risk reduction and management of probation-eligible offenders». Disponible en: <<https://www.ncsc.org/csi/evidence-based-sentencing>>. [Consulta; 23/08/2020.]
10. «Smart sentencing, also known as, «evidence-based sentencing», relies on the use of social scientific evidence, particularly about recidivism and correctional interventions that reduce criminal propensities, to determine a sentence that is likely to improve public safety» (Hannah-Moffat, 2013: 271).
11. «The principal use of EBS practices at the state trial court level is the use of actuarial risk —and needs— assessment information in state sentencing proceedings» (Warren, 2010: 156).

12. No confundir con la Smarter Sentencing Act, una propuesta de ley que ha sido discutida varias veces en el Congreso de los EE. UU. desde 2013 destinada fundamentalmente a suavizar las penas para delitos de tráfico de drogas no violentos, algunas de cuyas disposiciones han ido siendo aprobadas en años anteriores. La última versión está en discusión en el Senado a la fecha de cierre de este trabajo está disponible en: <<https://www.congress.gov/bill/116th-congress/senate-bill/2850/all-info>>. [Consulta: 23/08/2020.]

13. Como afirma Marcus: «The resulting critical issue in sentencing policy thus runs along a very different axis than the traditional divide between punitivists and advocates of reformation. Evidence-based «smart sentencing» posits that by rigorously scrutinizing data on what works or not on which offenders, we can allocate our correctional resources far more efficiently —measured by public safety— than if we continue to settle for «just deserts» with no accountability for outcomes» (Marcus, 2006: 57). En sentido muy parecido, cfr. el documento *Principles of Effective State Sentencing and Corrections Policy*, publicado por el National Conference of State Legislatures (2011: 4): «Many of these approaches leave behind outdated notions of being «soft» or «tough» on crime. Instead they look to be smart on crime to ensure that sentencing policies contribute to a favourable state return on public safety expenditures».

14. Entre otros, su encaje en las teorías convencionales sobre las finalidades de la pena; los conflictos que genera con determinadas garantías penales al erigir el riesgo de reincidencia en el criterio básico sobre el que articular la respuesta punitiva o hasta qué punto palia o exacerba otros problemas importantes ligados al castigo penal como la discriminación étnica, las desigualdades socioeconómicas, etc. Para una panorámica reciente y muy detallada de todas estas cuestiones, así como del concreto aspecto que se estudia en este trabajo, *vid.* las diversas aportaciones contenidas en la obra colectiva editada por de Keijser, Roberts y Ryberg (2019). En castellano, puede verse Martínez Garay (2019:149 ss.).

15. Lo que incluye no solo las herramientas de estimación de riesgo digamos «convencionales», sino también el uso de inteligencia artificial en sentido estricto, es decir, herramientas con un nivel de automatización completo o casi completo. Sobre los ámbitos de aplicación de la inteligencia artificial en el Derecho penal, cfr. Miró Llinares (2018: 87 ss. y 105 ss.) para el campo de la valoración de riesgo de reincidencia. Sobre algunos peligros derivados de esta utilización, y comentando el conocido caso Loomis, cfr. Martínez Garay (2018).

16. Sí lo aborda Benito Sánchez (2020: 27 ss.).

17. Cullen describe con claridad el firme rechazo al que se enfrentaban los partidarios de la rehabilitación en las décadas de los ochenta y noventa, y los esfuerzos que se hicieron para poder aportar «números» que convencieran a los escépticos de que sí había pruebas de que ciertas intervenciones funcionaban (2013: 335 ss.).

18. Expresión que se popularizó tras el conocido trabajo de Martinson (1974: 22-54).

19. Por ejemplo, el *Public Safety Policy Brief*, núm. 8 (mayo 2009), editado por el PEW Center on The States como parte de su Public Safety Performance Project, titulado «Arming the Courts with Research», afirma que «[...] we cannot afford to ignore the opportunity to reduce offender recidivism and resulting high crime rates through use of these cost-effective evidence-based practices», y en la página siguiente matiza: «The goal of recidivism reduction is to reduce crime, not just to rehabilitate offenders. [...] It is not «soft» on crime». Disponible en <[https://www.ncsc.org/\\_\\_data/assets/pdf\\_file/0018/25551/pew\\_armingthecourtwithresearch.pdf](https://www.ncsc.org/__data/assets/pdf_file/0018/25551/pew_armingthecourtwithresearch.pdf)>. [Consulta: 25/08/2020.]. La Guía publicada en 2011 por el National Center for State Courts sobre EBS enfatiza que el objetivo principal es el de «reducir y manejar el riesgo» que supone el delincuente para mejor proteger la seguridad pública. Las escasas referencias a la rehabilitación que aparecen en la guía la colocan al mismo nivel que otras finalidades legítimas de la pena como la inocuización o la intimidación selectiva (Casey, Warren y Elek, 2011: 1, 11 y 12).

20. El National Center for State Courts es de las pocas instituciones que expresamente afirmaba no perseguir ese objetivo. *Vid.* «NCSC Fact Sheet on Evidence-Based Sentencing», 2014. Disponible en: <<https://www.ncsc.org/csi/evidence-based-sentencing/overview>>. [Consulta: 25/08/2020.]. No obstante, un documento más reciente de la misma organización reconoce que las evaluaciones de riesgo sí pueden ser un factor que contribuya a evitar imponer penas de prisión a delincuentes de riesgo bajo (Casey, Elek y Warren, 2017: 3).
21. Lo expresan con claridad Monahan y Skeem cuando se preguntan cuál es la razón por la que tras 40 años de *just deserts* estemos asistiendo a un notable resurgimiento del riesgo como un componente esencial en la imposición y ejecución de las penas: «Money appears to be the principal answer» (2014: 158).
22. Lo expresa con absoluta claridad el *Public Safety Policy Brief* del PEW Center al que ya nos hemos referido *supra* (nota 20): «The goal of recidivism reduction is to reduce crime, not just to rehabilitate offenders. It includes both effective treatment services—programs proven to reduce reoffending—and swift and effective use of graduated sanctions. It highlights the importance of holding offenders more strictly accountable than we do now for compliance with courtorders and conditions of supervision. It is not «soft» on crime».
23. Para una panorámica sobre los antecedentes y el surgimiento de este movimiento, cfr. Parkhurst (2016: 14 ss.).
24. Una argumentación de este tipo puede verse en el documento «Principles of Evidence-Based Policymaking» elaborado por el grupo Evidence-Based Policymaking Collaborative (integrado por distintas fundaciones y organizaciones como Laura and John Arnold Foundation, Urban Institute y Brookings Institution): «In an era of intense partisanship and constrained public resources, evidence-based policymaking can help bridge the partisan political divide and support research-based debate about what outcomes we want to achieve, for whom, and at what cost. It encourages transparency and accountability [...]. Disponible en: <[https://www.urban.org/sites/default/files/publication/99739/principles\\_of\\_evidence-based\\_policymaking.pdf](https://www.urban.org/sites/default/files/publication/99739/principles_of_evidence-based_policymaking.pdf)>. [Consulta: 26/08/2020.]
25. En relación con el Derecho penal, se ha afirmado que «[...] la revolución basada-en-la-evidencia en medicina proporciona un modelo sobre cómo una profesión apegada a la tradición puede cambiar su enfoque para complementar y apoyar el tradicional ámbito de juicio discrecional con investigación científica sólida» (Garret, 2018: 1496).
26. El movimiento cuenta con notables apoyos, entre los que cabe destacar el que le prestó el Gobierno laborista británico de Tony Blair, que en la década de los noventa incorporó explícitamente en su programa el enfoque de las políticas basadas en la evidencia con el propósito declarado de «[...] eliminar concepciones ideológicas y partidistas del proceso de toma de decisiones políticas, asentando dichas decisiones en la evidencia basada en la investigación científica [...] sobre lo que funcionada y sobre cómo se podrían producir mejores resultados» (Benito Sánchez, 2020: 28). En la actualidad, existen numerosas agencias gubernamentales y no gubernamentales que persiguen incrementar el uso de evidencia científica en el diseño de las políticas públicas. Cfr. Parkhurst (2016: 16).
27. Puede verse un resumen de los orígenes de la EBM y una referencia a sus antecedentes en Berguer y Berguer (2003); y Junquera *et al.* (2003).
28. Como afirman Berguer y Berguer (2003: 277), «[...] un clínico que siga únicamente las diez mejores publicaciones de Medicina Interna, [tendría] que leer una media mensual de 200 artículos y 10 editoriales para considerarse científicamente “al día”».

29. Existen múltiples escalas de niveles de evidencia elaboradas por diversas instituciones nacionales e internacionales; puede verse una sistematización de varias de ellas en Manterola, Asenjo-Lobos y Otzen (2014). Junquera *et al.* (2003: 269) ofrecen un esquema mucho más sencillo.
30. «En general los investigadores que participan en un [ensayo clínico randomizado] no tienen control sobre su diseño o sobre el análisis de los datos en los que se basan las conclusiones. La industria farmacéutica generalmente controla la metodología, el análisis de datos y las conclusiones finales, ignorando, a veces resultados desfavorables» (Berguer y Berguer, 2003: 277).
31. Heneghan *et al.* (2017: 1). En similar sentido se pronuncian Junquera *et al.* (2003: 269), que afirman que «la mayoría» de los trabajos que se publican en medicina no realiza aportaciones significativas al conocimiento, o Berguer y Berguer (2003: 277), que subrayan que, según los defensores de la medicina basada en la evidencia, apenas el 3 % de los trabajos que se publican están elaborados con rigor metodológico.
32. Cfr. sobre todas estas críticas, y desde distintas perspectivas, Ioannidis (2016); Fava (2017), Berguer y Berguer (2003); Horwitz y Singer (2017); Knottnerus y Tugwell (2017); Guyatt (2017); Heneghan *et al.* (2017); y Greenhalgh, Howick y Maskrey (2014). Sobre este último aspecto, Ioannidis afirma que en algunos contextos estamos ya cerca, e incluso hemos sobrepasado, el punto a partir del cual la medicina está contribuyendo a disminuir, en lugar de aumentar, el bienestar en nuestra sociedad (2016: 85).
33. Ioannidis se muestra escéptico sobre la posibilidad de poder «tirar por la borda» a los piratas que han secuestrado la medicina basada en la evidencia (2016: 86; y 2017: 13). Quizá una muestra de lo complicado que puede resultar corregir el rumbo es el Manifiesto publicado en 2017 por los directores del Centre for Evidence Based Medicine de la Universidad de Oxford y los editores del *British Medical Journal*, texto que, si bien detalla de manera muy concreta y exhaustiva todos los problemas que aquejan a la medicina basada en la evidencia, resulta mucho más vago e inconcreto en el listado de soluciones que propone (Heneghan *et al.*, 2017: 2, donde comparan las tablas Box 1 y Box 2).
34. En la propia determinación de cuáles son los hechos, qué debe considerarse «evidencia» y qué no, influye el poder y la influencia relativas que en cada país, cultura, y momento histórico tienen diversas tradiciones epistemológicas (Strassheim y Kettunen, 2014: 260).
35. Por recordar solo algunos ejemplos conocidos: la Organización Mundial de la Salud consideró la homosexualidad como un trastorno mental hasta 1990; el DSM 5 ha incorporado, si bien (todavía) no como trastornos acreditados, sino dentro de una Sección titulada «Alteraciones que necesitan más estudio», el trastorno de duelo complejo persistente o el trastorno por consumo de cafeína. También hay ejemplos fuera del campo de las enfermedades mentales: la menopausia se considera en la actualidad, si bien no directamente una enfermedad, sí una situación que requiere la administración de medicamentos.
36. Hammersley (2005: 94 ss.). Sobre la simplificación asociada a los análisis del tipo «what works» y los peligros que entraña, cfr., en detalle, Parkhurst (2016: 18 ss.).
37. Greenhalgh y Russel (2009: 307); y Hammersley (2005: 89 ss.). Como afirma Klein, en el caso de las ciencias sociales, la investigación tiende a plantear tantas nuevas preguntas como cuestiones resuelve (2000: 65).
38. Un ejemplo en el ámbito de la política criminal es el Crime Reduction Program puesto en marcha por el gobierno laborista británico en el año 1999, que pretendía obtener evidencia científica sólida sobre las estrategias «que funcionan» en el control de la delincuencia y las que no, y trasladar ese conocimiento posteriormente al diseño de las políticas criminales. Sin embargo, este programa fracasó lamentablemente poco tiempo después y fue cancelado sin haber logrado casi ninguno de sus objetivos. Sobre las características de este programa y las dificultades que enfrentó, cfr. Maguire (2004); Hope (2004) y Hough (2004).

39. Estas limitaciones epistemológicas del enfoque «basado en la evidencia» cuando se aplica a los fenómenos sociales no son problemas episódicos que pueden aparecer en una investigación concreta, sino dificultades de principio. La evidencia empírica que proporcionan las investigaciones cuantitativas hechas según la jerarquía establecida en medicina no es sin más evidencia válida desde el punto de vista científico para las ciencias sociales en general, y para la política criminal en particular. Surge la duda de si al aplicar la metodología propia de una ciencia a un campo de conocimiento esencialmente distinto es verdaderamente «ciencia» lo que se hace y «evidencia» lo que se obtiene, o si, por el contrario, se estaría fabricando una «pseudociencia»

40. También, desde luego, en el conocimiento acumulado sobre programas de rehabilitación «que funcionan». No obstante, y como ya hemos advertido *supra*, la premisa central del EBS es que hay que ajustar la sentencia al nivel de riesgo de cada sujeto no solo para aplicar programas de rehabilitación, sino también de pura neutralización cuando estos últimos sean más efectivos. Tanto en un caso como en el otro, la condición necesaria es que sea posible estimar el riesgo con un grado suficiente de acierto: si esto no fuera viable, carecería de cualquier lógica diseñar las sanciones en atención a la prevención especial.

41. Sobre las diferentes técnicas de estimación de riesgo y sus características, cfr., por todos, Loinaz (2017).

42. Cfr., por todos, Ægisdóttir *et al.* (2006).

43. Fazel *et al.* (2016: 539 y 540): 20 % en la predicción de reincidencia violenta en un periodo de un año; 37 % si el periodo se extiende a dos años. En sentido similar Fazel *et al.* (2019: 4): 17 % para la reincidencia violenta a un año; 20 % a dos.

44. Si en lugar de la comisión de nuevos delitos violentos lo que se estima es el riesgo de comisión de cualquier tipo de delito, los valores predictivos positivos son más elevados: 52 % de media en el metaanálisis de Fazel *et al.* (2012: 10); entre el 40 y el 60 % en el estudio de Fazel *et al.* (2019: 4) (dependiendo de las características de la muestra y de en qué nivel de riesgo se sitúe el umbral de discriminación); o incluso 75 % en el estudio de Fazel *et al.* (2016: 540). El problema es que este valor es mucho menos relevante desde el punto de vista de la justicia penal, pues resulta muy discutible que la probabilidad de comisión de delitos no violentos y de escasa gravedad justifique (ni desde la racionalidad puramente económica ni desde la valorativa) regímenes diferentes de dureza de las penas y de intensidad de la supervisión.

45. Y aun esto es una simplificación muy notable del problema, pues no es en modo alguno indiferente —desde la óptica del sistema de justicia penal— que se cometa un nuevo delito de hurto o un asesinato, que se cometa solo un nuevo delito o muchos, que esa reincidencia se observe a los pocos meses de haber salido de prisión o muchos años después, etc. Si las valoraciones estructuradas de riesgo de violencia no pueden ofrecer al juez información diferenciada sobre estos extremos, es dudoso que puedan considerarse adecuadas al fin que se supone que deben servir (Slobgin, 2019).

46. ROC para Receiver Operating Characteristics.

47. Sobre el significado y la información que proporciona el área bajo la curva ROC cfr., con mayor detalle, Mossman (1994); y Martínez Garay y Montes Suay (2018).

48. Zedner expresa con gran claridad el núcleo del problema: está muy bien saber las probabilidades medias de reincidencia de diversos grupos de delincuentes y que el hecho de estar clasificado en un nivel de riesgo más elevado se corresponde realmente con una probabilidad incrementada de reincidir. Pero eso no elimina el problema de que muchas decisiones siguen siendo dicotómicas, por lo que, si se quiere utilizar la estimación de riesgo como criterio de decisión, habrá que fijar un umbral de discriminación y un 30 % o un 40 % de probabilidad de reincidencia puede seguir siendo insuficiente para justificar

según qué decisiones por los altos número de falsos positivos que se siguen produciendo (Zedner, 2019: 232 ss.). Para un ejemplo de cómo el área bajo la curva puede «enmascarar» tasas de error muy notables en las estimaciones de riesgo en función del punto de corte que se elija y de la prevalencia del fenómeno, cfr. Hester (2019: 220-226), relatando el caso real del diseño de una herramienta de valoración del riesgo en Pennsylvania.

49. Y, sin embargo, es fácil que el juez, que no es experto en estadística, malinterprete en alguno de estos dos sentidos la información que se le está proporcionando. Con ello, se sientan las bases para un preocupante malentendido.

50. Sobre algunos de los distintos indicadores existentes y su significado, cfr. Loinaz (2017).

51. Por ejemplo, se ha señalado que en este campo muchos conflictos de interés no son reconocidos (muchas herramientas son diseñadas y comercializadas por compañías privadas, para las que trabajan los investigadores que publican los resultados sobre su capacidad predictiva) y que hay también indicios de sesgos de publicación y de autoría (Douglas *et al.*, 2017: 135; y Fazel 2019: 202). Por otro lado, la complejidad de los indicadores estadísticos es tan elevada que los propios investigadores en ocasiones los utilizan de manera incorrecta (Singh, Desmarais y Van Dorn, 2013).

52. Un juicio todavía más severo puede encontrarse en Fazel, autor que, tras enumerar una serie de criterios de validez científica a los que deberían ajustarse las herramientas de valoración de riesgo de reincidencia o de violencia, afirma que muy pocas de las actualmente existentes cumplen más de uno o dos de ellos, y concluye que «[...] the most commonly used tools in criminal justice are not fit for purpose for prediction purposes» (Fazel, 2019: 207).

53. Cfr., por ejemplo, Yang, Wong y Coid (2010: 761), que afirman que, debido a sus moderados niveles de eficacia predictiva, los instrumentos de valoración del riesgo no deberían utilizarse como el único ni como el principal medio para tomar decisiones clínicas o judiciales que dependen de un alto nivel de exactitud predictiva, como la «preventive detention». Cfr., también, Fazel (2019:199 y 201).

54. Porque la rapidez con la que avanza la tecnología puede cambiar drásticamente el panorama en cualquier momento.

55. Es la idea que se suele expresar con la frase «garbage in, garbage out», es decir, «si metes basura, saldrá basura».

56. La distorsión que las conformidades pueden introducir en el sistema no es en absoluto menor: según un informe publicado en 2018 por la National Association of Criminal Defense Lawyers, el 97 % de los procesos en los EE. UU. terminan con una conformidad. Este porcentaje tan elevado se explica porque si el acusado rechaza la propuesta de pacto del fiscal y va a juicio, se enfrentará a una petición de condena muchísimo mayor, riesgo que muchos no se atreven a asumir (el texto del informe, titulado «The Trial Penalty: The Sixth Amendment Right to Trial on the Verge of Extinction and How to Save It» está disponible en: <<https://www.nacdl.org/Document/TrialPenaltySixthAmendmentRighttoTrialNearExtinct>>. [Consulta: 20/11/2020.]

57. Mentira que puede consistir tanto en no desvelar delitos realmente cometidos que no han llegado a conocimiento de las autoridades como en seguir declarándose culpable de aquellos por los que ha sido condenado, aunque no los haya cometido. Esto último no cabe descartarlo, dado que no aceptar la culpabilidad suele ser interpretado por el personal penitenciario y por los encargados del tratamiento como un signo de mal pronóstico, y ser un argumento para denegar permisos o progresiones de grado. El penado puede entender por ello que le conviene más no discutir una condena injusta a cambio de poder obtener antes la libertad.

58. Que no son, ni mucho menos, los únicos problemas. Enumerarlos todos excede con mucho el objeto de este trabajo, por lo que apuntaré solo algunas de las preguntas que ya están sobre la mesa en relación con las valoraciones de riesgo, y que se plantearán incluso con mayor urgencia según aumente el grado de automatización de las decisiones en este ámbito. ¿Pueden utilizarse en el proceso penal algoritmos a cuya fórmula no tiene acceso la defensa del acusado porque son propiedad de empresas privadas y están protegidos como secretos de empresa, como ocurrió en el caso *Loomis* (Martínez Garay, 2018)? ¿Dónde queda en estos casos la supuesta mayor transparencia, que se presenta como una de las ventajas de las valoraciones estructuradas de riesgo frente a los viejos juicios clínicos de peligrosidad? ¿Qué ocurre si las prácticas policiales y judiciales incluyen sesgos sistémicos, raciales o socioeconómicos (Starr, 2014; y Harcourt, 2015), que los algoritmos no hacen sino reproducir y perpetuar, pero ocultándolos tras la supuesta objetividad de las matemáticas? ¿Puede estar en estos casos la valoración de riesgo retroalimentándose en un círculo vicioso y creando ella misma las condiciones para que las estimaciones de riesgo se conviertan en realidad (Harcourt, 2007)? Sobre este complejo panorama, que va mucho más allá del objeto de este trabajo, cfr. Miró Llinares (2018); y Martínez Garay (2019).

59. Por ejemplo, Casey, Elec y Warren (2017: 4). En esta línea, Ostrom y Kauder (2013) valoran positivamente la introducción de EBS en el Estado de Virginia (a la que más abajo nos referiremos con mayor detalle), ofreciendo datos sobre el nivel de seguimiento de las recomendaciones por parte de los jueces, el aumento de la carga de trabajo de los agentes de supervisión o el grado de satisfacción que expresan diversos actores del sistema.

60. Brandariz ha señalado que esta es una característica propia de los modelos gerenciales de gestión de la criminalidad: que tienden a prestar más atención «[...] a lo que las instancias del sistema penal hacen que a los efectos y beneficios sociales que producen» (2016: 122-123). De esta manera, se «[...] facilita la evaluación positiva del funcionamiento del sistema, en la medida en que los criterios tienden a acomodarse a las tareas que efectivamente pueden desarrollarse, y las agencias disfrutan de una evidente capacidad de control sobre los indicadores de rendimiento; la consecuencia de ello, como es obvio, es que se dificulta la crítica y el análisis externos» (Brandariz, 2016: 124).

61. Otro experimento de aplicación, no propiamente de EBS sino más bien de *evidence-based corrections* a gran escala, es la Justice Reinvestment Initiative, sobre la que, por razones de espacio, no podemos referirnos en este trabajo más extensamente. Cfr., al respecto, Klingele (2015: 562 ss.); Garret (2018: 1503 ss.); Austin *et al.* (2013: *passim*), así como Sabol y Baumann (2020: *passim*).

62. Cfr. Ostrom y Kauder (2013: 161): «Viable strategies are needed that allow states to reserve scarce and expensive prison beds for the most dangerous offenders and use less costly punishment options for less serious offenders—all while maintaining public safety».

63. La primera tiene como argumento en contra el de que las valoraciones de riesgo que se proporcionan a los jueces no contienen valores absolutos (*v. gr.* «50 % de probabilidad de reincidir»), sino solo relativos, por lo que es difícil que los jueces puedan hacerse una idea de si el riesgo es mayor o menor del que, en ausencia de la valoración, ellos habrían pensado que hay (Stevenson y Doleac, 2019: 19).

64. El estudio solo analiza la evolución de las tasas de reincidencia en el grupo de delincuentes de bajo riesgo; en el de delincuentes sexuales no, por falta de datos disponibles, ya que es un grupo reducido y con condenas muy largas que aún se estaban ejecutando en el momento de realizarse el análisis.

65. Cfr. National Center for State Courts, *NCSC Fact Sheet. Evidence-Based Sentencing* (2014): «EBS does not replace judicial discretion. EBS provides additional information for the judge to consider in crafting an offender's sentence. Judges are free to use the information as they deem best in light of all the facts of the case and the parties' presentations in court to address all the purposes of sentencing».

Disponible en: <[https://www.ncsc.org/\\_\\_data/assets/pdf\\_file/0018/25290/ebs-fact-sheet-8-27-14.pdf](https://www.ncsc.org/__data/assets/pdf_file/0018/25290/ebs-fact-sheet-8-27-14.pdf)>. [Consulta: 25/09/2020.]

66. A mi juicio, Slobogin alcanza una conclusión similar a esta (2018: 587 y 593), si bien insiste en la defensa de lo que él denomina *risk-based sentencing* fundamentalmente porque considera que la alternativa (el enfoque retributivo o *desert-based sentencing*) plantea problemas de igual o mayor magnitud (Slobogin, 2019: 116 ss.).

67. Monahan y Skeem (2014: 161); y Monahan y Skeem (2016: 498, 506).

68. Cfr. Stevenson y Slobogin (2018:694): en la herramienta COMPAS, ampliamente utilizada por diversas jurisdicciones en EE. UU., la edad explica el 48 % de la variación en las puntuaciones de riesgo. Los autores advierten de que, debido a que el algoritmo con el que funciona COMPAS está protegido como secreto comercial de la empresa propietaria, no han podido acceder al detalle de su funcionamiento, y que, por ello, los cálculos son aproximados (*Ibid.*: 690 ss. y 695 ss.). La relación entre el peso de la edad y el de los antecedentes criminales en distintas herramientas de valoración del riesgo puede verse en la tabla contenida en ese mismo trabajo (*Ibid.*: 699).

69. Cfr. Tonry (2013b: 182), con ulteriores referencias.

70. No deja de ser llamativo, por otra parte, que tanto este problema del diseño retrospectivo como otras debilidades metodológicas de los estudios que defienden las políticas de incapacitación selectiva han sido puestos de manifiesto reiteradamente por la doctrina penal y criminológica, desde hace más de 30 años —cfr., tan solo, von Hirsch (1985: 105-127); y Auerhahn (1999: *passim*), ambos con ulteriores referencias bibliográficas), y que, a pesar de ello, sigan publicándose trabajos que defienden una política criminal parecida sin preocuparse por aportar evidencia de que dichos problemas estén superados. Tonry utiliza la expresión «vino viejo en viejos odres» en el título de uno de sus últimos trabajos publicados sobre la materia (Tonry, 2019). Como afirma Auerhahn: «Given that we have every reason to believe a small subset of criminal offenders contribute disproportionately to the total volume of crime in a society, a strategy that promises to locate and incapacitate this group is almost irresistible in its elegance. The seductive simplicity of selective incapacitation leads otherwise conscientious researchers to conclude that it works, despite the total lack of evidence to support such a conclusion. It is such a marvelous idea; it simply has to work! The problem is, unfortunately, that it doesn't» (1999: 727).

71. Algo muy parecido cabría decir de otro tipo de factores como algunas enfermedades mentales. Monahan y Skeem (2016: 504) ponen el ejemplo de los traumas de los veteranos de guerra.

72. Un claro ejemplo de ello es, precisamente, lo que ocurrió con el famoso trabajo de Martinson (1974). La enorme trascendencia que tuvo se debió a que apareció en un momento en que la opinión pública y los actores políticos habían dejado (por diversas razones) de creer en la rehabilitación: el clima ya era propicio para abandonar ese ideal, y el trabajo de Martinson vino a poner el último clavo en el ataúd de la rehabilitación (Cullen, 2013: 329) al añadir el argumento de la falta de evidencia científica. El hecho de que poco después se publicaran trabajos igualmente rigurosos en sentido contrario (algunos revisaron los mismos estudios que Martinson había analizado), demostrando que muchos programas de tratamiento sí reducían la reincidencia e incluso provocando que el propio Martinson corrigiera su conclusión, ya no consiguió ninguna resonancia en un ambiente que se había vuelto decididamente hostil (Pratt, Gau y Franklin, 2011: 85).

73. Si, con respecto al EBS, algunos comparten el objetivo de ofrecer sanciones alternativas a la prisión para un número mayor de personas, desde el otro extremo del espectro ideológico lo que se defiende es el objetivo de mantener a los delincuentes más peligrosos encerrados durante más tiempo.

## BIBLIOGRAFÍA

- ANDREWS, Don Arthur y James BONTA (2010): «Rehabilitating criminal justice policy and practice», *Psychology, Public Policy, and Law*, 16(1), 39-55.
- AUERHAHN, Kathleen (1999): «Selective incapacitation and the problem of prediction», *Criminology*, 37(4), 703-734.
- AUSTIN, James, Eric CADORA, Todd R. CLEAR, Kara DANSKY, Judith GREENE, Vanita GUP-TA, Marc MAUER, Nicole PORTER, Susan TUCKER y Malcolm C. YOUNG (2013): «Ending Mass Incarceration Charting a New Justice Reinvestment» [en línea] <<https://www.sentencingproject.org/wp-content/uploads/2015/12/Ending-Mass-Incarceration-Charting-a-New-Justice-Reinvestment.pdf>>. [Consulta: 27/09/2020.]
- ÆGISDÓTTIR, Stefania *et al.* (2006): «The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction», *The Counseling Psychologist*, 34(3), 341-382.
- BENITO SÁNCHEZ, Demelsa (2020): *Evidencia empírica y populismo punitivo. El diseño de la política criminal*, Barcelona: Bosch.
- BERGER, Alberto y R. BERGUER (2003): «Medicina basada en evidencia (MBE). Contras», *Revista Española de Cirugía Oral y Maxilofacial*, 25, 273-279.
- BRANDARIZ GARCÍA, José Ángel (2016): *El modelo gerencial-actuarial de penalidad: Eficiencia, riesgo y sistema penal*, Madrid: Dykinson.
- CASEY, Pamela, Jennifer ELEK y Roger WARREN (2011): *Using Offender Risk and Needs Assessment Information at Sentencing. Guidance for Courts from a National Working Group*, National Center for State Courts [en línea] <[https://www.ncsc.org/\\_\\_data/assets/pdf\\_file/0019/25174/rna-guide-final.pdf](https://www.ncsc.org/__data/assets/pdf_file/0019/25174/rna-guide-final.pdf)>. [Consulta: 25/09/2020.]
- (2017): «Use of risk and needs assessment information in state sentencing proceedings» [en línea] <<https://www.ncsc.org/csi/evidence-based-sentencing/overview>>. [Consulta: 25/08/2020.]
- CULLEN, Francis T. (2013): «Rehabilitation: Beyond Nothing Works», *Crime and Justice*, 42(1), 299-376.
- DOUGLAS, Thomas, Jonathan PUGH, Illina SINGH, Julián SAVULESCU y Seena FAZEL (2017): «Risk assessment tools in criminal justice and forensic psychiatry: The need for better data», *European Psychiatry* 42, 134-137.
- FAVA, Giovanni A. (2017): «Evidence-based medicine was bound to fail: a report to Alvan Feinstein», *Journal of Clinical Epidemiology*, 84, 3-7.
- FAZEL, Seena (2019): «The scientific validity of current approaches to violence and criminal risk assessments», en J. de Keijser, J. Roberts y J. Ryberg (eds.), *Predictive sentencing. Normative and empirical perspectives*, Oxford: Hart Publishing, 197-211.
- FAZEL, Seena, Achim WOLF, María D. L. A. VAZQUEZ-MONTES y Thomas R. FANSHawe (2019): «Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec)», *Scientific Reports* 9, 841, DOI: <https://doi.org/10.1038/s41598-018-37539-x>.
- FAZEL, Seena, Zheng CHANG, Thomas FANSHAWE, Niklas LÄNGSTRÖM, Paul LICHTENSTEIN, Henrik LARSSON y Susan MALLETT (2016): «Prediction of violent reoff ending on release from prison: derivation and external validation of a scalable tool», *Lancet Psychiatry*, 3, 535-543.
- FAZEL, Seena, Jay P. SINGH, Helen DOLL y Martín GRANN (2012): «Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24827 people: systematic review and meta-analysis», *British Medical Journal*, 345: e4692.
- GARRETT, Brandon L. (2018): «Evidence-Informed Criminal Justice», *George Washington Law Review*, 86(101), 1490-1524.
- GREENHALGH, Trisha, Jeremy HOWICK y Neal MASKREY (2014): «Evidence based medicine: a movement in crisis?», *British Medical Journal* 2014; 348: g3725, DOI: 10.1136/bmj.g3725.

- GREENHALGH, Trisha y Jill RUSSELL (2009): «Evidence-Based Policymaking: a critique», *Perspectives in Biology and Medicine*, 52(2), 304-318.
- GUYATT, Gordon (2017): «EBM has not only called out the problems but offered solutions», *Journal of Clinical Epidemiology*, 84, 8-10.
- HAGGERTY, Kevin D. (2004): «Displaced expertise: Three constraints on the policy relevance of criminological thought», *Theoretical Criminology*, 8(2), 211-231.
- HANNAH-MOFATT, Kelly (2013): «Actuarial Sentencing: An ‘Unsettled’ Proposition», *Justice Quarterly*, 30 (2), 270-296.
- HANSON, R. Karl (2017): «Assessing the Calibration of Actuarial Risk Scales. A Primer on the E/O Index», *Criminal Justice and Behavior*, 44(1), 26-39. DOI: 10.1177/0093854816683956.
- HAMMERSLEY, Martyn (2005): «Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers’ case for research-based policy making and practice», *Evidence & Policy*, 1(1), 85-100.
- HARCOURT, Bernard E. (2007): *Against prediction: profiling, policing, and punishing in an actuarial age*, Chicago: University of Chicago Press.
- (2015): «Risk as a Proxy for Race: The Dangers of Risk Assessment», *Federal Sentencing Reporter*, 27(4), 237-243.
- HENEGHAN, Carl, Kamal R MAHTANI, Ben GOLDACRE, Fiona GODLEE, Helen MACDONALD y Duncan JARVIES (2017): «Evidence based medicine manifesto for better healthcare. A response to systematic bias, wastage, error, and fraud in research underpinning patient care», *British Medical Journal*, 357: j297, DOI: 10.1136/bmj.j2973.
- HESTER, Rhys (2019): «Risk assessment at sentencing. The Pennsylvania experience», en J. de Keijser, J. Roberts y J. Ryberg (eds.), *Predictive sentencing. Normative and empirical perspectives*, Oxford: Hart Publishing, 213-237.
- HOPE, Tim (2004), «Pretend it works. Evidence and governance in the evaluation of the Reducing Burglary Initiative», *Criminal Justice*, 4(3), 287-308.
- HORWITZ, Ralph I. y Burton H. SINGER (2017): «Why evidence-based medicine failed in patient care and medicine-based evidence will succeed», *Journal of Clinical Epidemiology*, 84, 14-17.
- HOUGH, Mike (2004): «Modernization, scientific rationalism and the Crime Reduction Programme», *Criminal Justice*, 4(3), 239-253.
- IOANNIDIS, John P.A. (2016): «Evidence-based medicine has been hijacked: a report to David Sackett», *Journal of Clinical Epidemiology* 73 (2016), 82-86.
- (2017): «Hijacked evidence-based medicine: stay the course and throw the pirates overboard», *Journal of Clinical Epidemiology* 84, 11-13.
- JUNQUERA, Luis Manuel, J. BALADRÓN, José María ALBERTOS y Sonsoles OLAY (2003): «Medicina basada en la evidencia (MBE). Ventajas» *Revista Española de Cirugía Oral y Maxilofacial*, 25, 265-272.
- KLEIN, Rudolf (2000): «From evidence-based medicine to evidence-based policy?», *Journal of Health Services Research and Policy*, 5(2), 65-66.
- KLINGELE, Cecelia (2015): «The Promises and Perils of Evidence-Based Corrections», *Notre Dame Law Review*, 91 (2), 537-584.
- KNOTTNERUS, J. André y Peter TUGWELL (2017): «Evidence-based medicine: achievements and prospects», *Journal of Clinical Epidemiology*, 84, 1-2.
- LOINAZ, Ismael (2017): *Manual de evaluación del riesgo de violencia. Metodología y ámbitos de aplicación*, Madrid: Pirámide.
- LÓPEZ-OSSORIO Juan José, José Luis GONZÁLEZ-ÁLVAREZ y Antonio Andrés PUEYO (2016): «Eficacia predictiva de la valoración policial del riesgo de la violencia de género», *Psychosocial Intervention*, 25, 1-7.
- MAGUIRE, Mike (2004): «The Crime Reduction Programme in England and Wales: Reflections on the vision and the reality», *Criminal Justice*, 4(3), 213-237.

- MANTEROLA Carlos, Claudia ASENJO-LOBOS y Tamara OTZEN (2014): «Jerarquización de la evidencia. Niveles de evidencia y grados de recomendación de uso actual», *Revista Chilena de Infectología*, 31(6), 705-718.
- MARCUS, Michael H. (2006): «Smart Sentencing: Public Safety, Public Trust and Confidence Through Evidence-Based Dispositions», National Center for State Courts, 2006 [en línea] <<https://ncsc.contentdm.oclc.org/digital/collection/criminal/id/161/>>. [Consulta: 23/08/2020.]
- MARTÍNEZ GARAY, Lucía (2018): «Peligrosidad, algoritmos y *due process*: El caso *State vs. Loomis*», *Revista de Derecho penal y Criminología*, 20, 485-502.
- (2019): «La relación entre culpabilidad y peligrosidad», en M. Maraver y L. Pozuelo (coords.), *La culpabilidad*, Montevideo-Buenos Aires: Editorial BdF, 115-199.
- MARTÍNEZ GARAY, Lucía y Francisco MONTES SUAY (2018): «El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias», *Indret. Revista para el análisis del Derecho*, 2, 1-47.
- MARTINSON, Robert (1974): «What works? Questions and answers about prison reform», *The Public Interest*, 35, 22-54.
- MIRÓ LLINARES, Fernando (2017): «La función de la pena ante “el paso empírico” del Derecho penal», *Revista General de Derecho Penal*, 27.
- (2018): «Inteligencia artificial y Justicia Penal: Más allá de los resultados lesivos causados por robots», *Revista de Derecho Penal y Criminología*, 20, 87-130.
- MONAHAN, John (1981): *Predicting violent behaviour. An assessment of clinical techniques*, London: Sage Publications.
- MONAHAN, John y Jennifer L. SKEEM (2014): «Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning», *Federal Sentencing Reporter*, 26(3), 158-166.
- (2016): «Risk Assessment in Criminal Sentencing», *Annual Review of Clinical Psychology*, 12, 489-513.
- MOSSMAN, Douglas (1994): «Assessing predictions of violence: Being accurate about accuracy», *Journal of Consulting and Clinical Psychology*, 62(4), 783-792.
- (2006): «Critique of pure risk assessment o Kant meets Tarasoff», *University of Cincinnati Law Review*, 75, 523-609.
- National Conference of State Legislatures (2011): *Principles of Effective State Sentencing and Corrections Policy. A Report of the NCSL Sentencing and Corrections Work Group* [en línea] <<https://www.ncsl.org/documents/cj/pew/WGprinciplesreport.pdf>>. [Consulta: 23/08/2020.]
- National Research Council (2014): *The Growth of Incarceration in the United States: Exploring Causes and Consequences*, Washington, DC: The National Academies Press [en línea] <<https://doi.org/10.17226/18613>>.
- OSTROM, Brian J. y Neal B. KAUDER (2013): «The Evolution of Offender Risk Assessment in Virginia», *Federal Sentencing Reporter*, 25(3), 161-167.
- PARKHURST, Justin (2016): *The Politics of Evidence. From evidence-based policy to the good governance of evidence*, London: Routledge [en línea] <<https://www.taylorfrancis.com/books/9781315675008>>. [Consulta: 02/11/2020.]
- PRATT, Travis C., Jacinta M. GAU y Travis W. FRANKLIN (2011): *Key Ideas in Criminology and Criminal Justice*, Thousand Oaks: Sage.
- RICHARDSON, W. Scott (2017): «The practice of evidence-based medicine involves the care of whole persons», *Journal of Clinical Epidemiology*, 84, 18-21.
- ROSSEGGER, Astrid, Jérôme ENDRASS, Juliane GERTH y Jay P. SINGH (2014): «Replicating the Violence Risk Appraisal Guide: A Total Forensic Cohort Study», *PLoS ONE* 9 (3): e91845. DOI:10.1371/journal.pone.0091845.
- ROSSEGGER, Astrid, Juliane GERTH, Katharina SEEWALD, Frank URBANIOK/Jay P. SINGH y Jérôme ENDRASS (2013). «Current obstacles in replicating risk assessment findings: a systematic review of commonly used actuarial instruments», *Behavioural Sciences and the Law*, 31, 154-164.

- SABOL, William J. y Miranda L. BAUMANN (2020): «Justice Reinvestment: Vision and Practice», *Annual Review of Criminology*, 3, 317-339.
- SACKETT, David L., William M. ROSENBERG, J. A. GRAY, R. Brian HAYNES y W. Scott RICHARDSON (1996): «Evidence based medicine: what it is and what it isn't», *British Medical Journal*, 1996, January, 13; 312 (7023), 71-72. DOI: 10.1136/bmj.312.7023.71.
- SALTELLI, Andrea y Mario GIAMPIETRO (2017): «What is wrong with evidence based policy, and how can it be improved?», *Futures*, 91, 62-71.
- SHEPHERD, Stephan M. y Danny SULLIVAN (2017): «Covert and Implicit Influences on the Interpretation of Violence Risk Instruments», *Psychiatry, Psychology and Law*, 24(2), 292-301.
- SÁNCHEZ VILANOVA, María (2017): «Responsabilidad de los delincuentes juveniles a la luz de la neurociencia», *Revista de Derecho y Genoma Humano. Genética, Biotecnología y Medicina Avanzada*, 47, 199-218.
- SINGH, Jay P. (2013): «Predictive validity performance indicators in violence risk assessment: a methodological primer», *Behavioural Sciences and the Law*, 31, 8-22.
- SINGH, Jay P., Seena FAZEL, Ralitsa GUEORGUIEVA y Alec BUCHANAN (2014): «Rates of violence in patients classified as high risk by structured risk assessment instruments», *British Journal of Psychiatry*, 204(3), 180-187. DOI: 10.1192/bj.p.113.131938.
- SINGH, Jay P., Sarah L. DESMARAIS y Richard A. VAN DORN (2013): «Measurement of Predictive Validity in Violence Risk Assessment Studies: A Second-Order Systematic Review», *Behavioral Sciences and the Law*, 31, 55-73.
- SLOBOGIN, Christopher (2018): «Principles of Risk Assessment: Sentencing and Policing», *Ohio State Journal of Criminal Law*, 15, *Vanderbilt Law Research Paper* núm. 18-09.
- (2019): «A Defense of Modern Risk-Based Sentencing», en J. de Keijser, J. Roberts y J. Ryberg (eds.), *Predictive sentencing. Normative and empirical perspectives*, Oxford: Hart Publishing, 107-125.
- STARR, Sonja B. (2014): «Evidence-Based Sentencing and the Scientific Rationalization of Discrimination», *Stanford Law Review* 66(4), 803-872.
- STEVENSON, Megan T. y Jennifer L. DOLEAC (2019): *Algorithmic Risk Assessment in the Hands of Humans*, IZA-Institute of Labor Economics Discussion Paper Series núm. 12853 [en línea] <<https://www.iza.org/publications/dp/12853/algorithmic-risk-assessment-in-the-hands-of-humans>>. [Consulta: 26/09/2020.]
- STEVENSON, Megan T. y Christopher SLOBOGIN (2018): «Algorithmic Risk Assessments and the Double-Edged Sword of Youth», *Washington University Law Review*, 96(3), 681-706.
- STRASSHEIM, Holger y Pekka KETTUNEN (2014): «When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms», *Evidence & Policy*, 10(2), 259-277.
- SZMUKLER, George, Brian EVERITT y Morven LEESE (2012): «Risk assessment and receiver operating characteristic curves», *Psychological Medicine*, 42, 895-898.
- TONRY, Michael (2019): «Sentencing and Prediction. Old wine in old bottles», en J. de Keijser, J. Roberts y J. Ryberg (eds.), *Predictive sentencing. Normative and empirical perspectives*, Oxford: Hart Publishing, 269-298.
- TONRY, Michael (2010): «Public Criminology and Evidence-Based Policy», *Criminology and Public Policy*, 9(4), November, 783-798.
- (2013a): «Evidence, Ideology, and Politics in the Making of American Criminal Justice Policy», *Crime and Justice: A Review of Research*, 42(1), 1-18.
- (2013b): «Sentencing in America, 1975-2025», *Crime and Justice: A Review of Research*, 42(1), 141-198.
- VELÁSQUEZ VALENZUELA, Javier (2014): «El origen del paradigma de riesgo», *Política Criminal*, 9 (17), julio, 58-117.
- VILJOEN, Jodi L., Dana M. COCHRANE y Melissa R. JONNSON (2018): «Do Risk Assessment Tools Help Manage and Reduce Risk of Violence and Reoffending? A Systematic Review», *Law and Human Behavior*, 42(3), 181-214.

- VON HIRSCH, Andrew (1985): *Past or future crimes. Deservedness and dangerousness in the sentencing of criminals*, New Jersey: Rutgers University Press.
- WARREN, Roger K. (2010): «Evidence-Based Sentencing: Are We Up to the Task?», *Federal Sentencing Reporter*, 23(2), 153-158.
- YANG, Min, Stephen C. P. WONG y Jeremy COID (2010): «The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools», *Psychological Bulletin*, 136(5), 740-767.
- ZEDNER, Lucia (2019): «Erring on the side of safety: risk assessment, expert knowledge and the criminal court», en G. R. Sullivan y I. Dennis (eds.), *Seeking Security. Pre-empting the commission of criminal harms*, Oxford: Hart Publishing, 217-241.

**Fecha de recepción: 5 de octubre de 2010.**

**Fecha de aceptación: 2 de diciembre de 2020.**