



Consejo General
del Poder Judicial

NÚMERO
3
AÑO 2020

CDF

CUADERNOS DIGITALES
DE FORMACIÓN

**Lógica científica y valoración probatoria.
Especial referencia a las pruebas biológicas**

Derecho penal y riesgo: la prueba del riesgo de reincidencia

Lucía Martínez Garay

**Profesora de Derecho Penal
Universidad de Valencia**

Derecho penal y riesgo: la prueba del riesgo de reincidencia (1)

Lucía Martínez Garay

Profesora de Derecho Penal
Universidad de Valencia

Resumen

Este trabajo estudia cómo de sólida puede ser la prueba del riesgo de reincidencia hecha con herramientas estructuradas de valoración del riesgo cuando la ley exige tener en cuenta el pronóstico de reincidencia a la hora de tomar decisiones restrictivas de derechos del acusado o condenado. Se utiliza como hilo conductor del trabajo el conocido aforismo estadístico según el cual "todos los modelos son falsos, pero algunos son útiles" para examinar si las estimaciones cuantitativas de riesgo de reincidencia o de violencia, que siempre son "falsas" (inexactas) en alguna medida, son lo suficientemente buenas como para utilizarlas en derecho penal, pregunta que, a su vez, depende de cuáles sean los objetivos que con el derecho penal se quieran conseguir.

Palabras clave

Reincidencia, Peligrosidad, Valoración de la prueba, Política criminal

ÍNDICE:

- I. Paso de la noción de peligrosidad a la de riesgo
- II. "Todos los modelos probabilísticos son falsos, pero algunos son útiles" (George E. P. Box)
 1. Ningún modelo estadístico refleja de manera completamente fiel la realidad
 2. En cualquier estimación de probabilidad hay siempre errores
 - A) Algunos estudios recientes en España
 - B) Datos a nivel internacional
 - C) Una primera conclusión
 3. ¿Son los modelos lo suficientemente buenos para lo que pretendemos?

III. "Evidencias empíricas" y derecho penal

IV. Conclusiones

Bibliografía

Tanto en la fase de instrucción como al dictar sentencia y, sobre todo, en la fase de ejecución de la pena, hay bastantes decisiones que tienen que adoptar los jueces y otros operadores jurídicos en las que, según la regulación vigente, hay que tomar en consideración el pronóstico de comportamiento futuro del sujeto: desde la suspensión de la pena, pasando por las medidas cautelares, la imposición o la modificación de medidas de seguridad, la resolución de recursos sobre clasificación penitenciaria y permisos, la libertad condicional, o, en un futuro, la revisión de la pena de prisión permanente. La cuestión que es objeto del presente trabajo es la siguiente: ¿cómo de sólida puede llegar a ser la prueba del riesgo de reincidencia?

I. PASO DE LA NOCIÓN DE PELIGROSIDAD A LA DE RIESGO

Hace unas décadas los psicólogos y los psiquiatras opinaban mayoritariamente que no era posible efectuar pronósticos científicamente rigurosos sobre el comportamiento futuro violento de las personas. En un informe presentado en 1982 ante la Corte Suprema de Estados Unidos en un caso de pena capital, la Asociación Americana de Psiquiatría afirmó que "la falta de fiabilidad de las predicciones psiquiátricas de peligrosidad a largo plazo es hoy por hoy un hecho demostrado dentro de la profesión" (American Psychiatric Association, 1982) (2).

El convencimiento de que no era científicamente posible efectuar pronósticos fiables de peligrosidad estaba muy ligado a los resultados de algunos estudios publicados en los años setenta, en los que se había hecho un seguimiento de grandes grupos de personas que habían permanecido confinadas sobre la base de su peligrosidad, pero que tuvieron que ser puestos en libertad debido a que el proceso para su internamiento no había respetado todas las garantías. En uno de los casos, el caso Baxstrom, 967 pacientes fueron transferidos desde dos hospitales psiquiátricos penitenciarios en el estado de Nueva York (Matteawan y Dannemora) a dieciocho instalaciones civiles. Cuatro años más tarde, menos del 1 % había vuelto a prisión y el 2 %, a los psiquiátricos penitenciarios (solo veintiséis volvieron a ser ingresados en psiquiátricos

penitenciarios en algún momento entre 1966 y 1970). En general, en los primeros cuatro años tras la derivación, solo el 20 % había cometido agresiones en los hospitales o fuera de ellos o tenía antecedentes penales (3). En el caso Dixon, alrededor de quinientos enfermos mentales considerados peligrosos fueron liberados de Farview, un hospital para enfermos mentales en el estado de Pensilvania, igualmente por razones procesales; solo el 14 % de los sujetos evidenció algún comportamiento violento posterior tras un seguimiento de cuatro años (4).

En ambos casos, solo un pequeño porcentaje de los individuos presuntamente peligrosos (menos del 20 %) llevó a cabo conductas que corroboraran el pronóstico de peligrosidad; el resto, alrededor del 80 %, fueron falsos positivos. De ello se extrajeron consecuencias directas para el derecho: el profesor de Derecho y Psiquiatría Bernard L. Diamond afirmaba en 1974 que ni los psiquiatras ni otros científicos de la conducta eran capaces de predecir el comportamiento violento con un nivel de fiabilidad suficiente como para justificar la restricción de la libertad de las personas sobre la base de su potencial peligrosidad y que, por tanto, los tribunales debían dejar de recabar la opinión de estos expertos sobre dicha cuestión (Diamond, 1974: 452).

Sin embargo, es importante subrayar que en aquel momento histórico el escepticismo se proyectaba especialmente sobre la capacidad de los peritos clínicos para hacer juicios fiables de peligrosidad. Es decir, lo que se cuestionaba no era tanto la posibilidad en sí de efectuar estimaciones de peligrosidad, sino que los médicos, psiquiatras y psicólogos estuviesen capacitados para emitir sobre ello juicios basados en los conocimientos científicos propios de su disciplina. El informe de la Asociación Americana de Psiquiatría al que nos hemos referido *supra*, en el que desautorizaba las opiniones que dos peritos habían emitido sobre la peligrosidad del condenado a muerte, refleja esta diferencia con claridad. Tras subrayar la falta de fiabilidad de las predicciones psiquiátricas de peligrosidad a largo plazo, la Asociación Americana de Psiquiatría añadía: "El hecho de que los psiquiatras sean incapaces de predecir el comportamiento futuro violento no significa que estas predicciones no se puedan hacer nunca. [...] Lo que sostenemos, por el contrario, es que la predicción a largo plazo de la peligrosidad es esencialmente una determinación no especializada que no debería estar fundamentada en los diagnósticos u opiniones de expertos médicos, sino sobre la base de información prospectiva estadística o actuarial que es sustancialmente de naturaleza no médica" (American Psychiatric Association, 1982: 5).

Pues bien, la situación desde entonces ha evolucionado precisamente en esa dirección. Ha cambiado tanto la metodología con la que se hacen las estimaciones de riesgo de reincidencia como también la manera en que se informa sobre sus resultados.

Por lo que hace a lo primero, en criminología la idea de "peligrosidad" ha sido sustituida por el enfoque de la "valoración del riesgo" (*violence risk assessment*) (5), según el cual

de lo que se trata no es de averiguar si el individuo posee o no la cualidad subjetiva de peligroso (como si se le estuviera diagnosticando una patología), sino de valorar un conjunto de factores, personales —pero también ambientales— y cambiantes en el tiempo, que favorecen en mayor o menor grado la comisión de nuevos delitos y que permiten efectuar pronósticos sobre la reincidencia futura en términos de probabilidad: muy alta, alta, moderada, baja, muy baja, etc. (Steadman, 2000; Andrés Pueyo y Redondo, 2007; Andrés Pueyo y Echeburúa, 2010; Andrés Pueyo, 2013; Loinaz, 2017: 73 y ss.). Es decir, no se efectuaría un juicio sobre si el sujeto es o no es de una determinada manera, sino sobre la probabilidad de que acontezca determinada clase de hechos en el futuro. Además, el cambio en la conceptualización de aquello que se valora —de la peligrosidad al riesgo— ha ido acompañado de un cambio en la forma de medir dicho riesgo. El tradicional método clínico no estructurado, que se considera poco fiable, sesgado y carente de suficiente base empírica, ha ido siendo sustituido por métodos estructurados de estimación del riesgo, ya se trate de métodos estrictamente actuariales o bien de instrumentos de juicio clínico estructurado (6). Ambos tipos de herramientas de predicción del riesgo están basados en la observación empírica de grupos de sujetos y en la cuantificación y combinación estadística de los factores de riesgo (y, en algunos casos, de factores protectores) que concurren en ellos y que demuestran estar significativamente asociados a la aparición de conducta violenta y/o delictiva (7).

Los nuevos métodos de estimación del riesgo de reincidencia comportan una nueva manera de informar sobre sus resultados. Cuando la peligrosidad se valoraba según el método clínico tradicional, el experto emitía un informe pericial en el que describía la situación del sujeto, explicaba la metodología que había seguido para alcanzar su diagnóstico (que podía ser muy variada: entrevistas personales con el sujeto; revisión de su historial médico o psiquiátrico; administración de alguna prueba psicométrica de personalidad, ansiedad, inteligencia, etc.; entrevistas con familiares o personal sanitario que lo hubiera tratado con anterioridad; etc.) y argumentaba por escrito u oralmente ante el tribunal cómo había alcanzado su conclusión. En los nuevos métodos estructurados, sin embargo, aunque por supuesto puede hacerse también un informe escrito u oral en el que se expliquen tanto el procedimiento seguido como la conclusión, el protagonismo lo tienen los indicadores estadísticos que informan sobre la capacidad predictiva de la herramienta aplicada.

Con el cambio de enfoque y de metodología, la investigación sobre la valoración del riesgo de violencia ha experimentado un crecimiento espectacular. Pero tan significativo o más que el aumento del interés científico sobre esta cuestión es el notable cambio de opinión que ha tenido lugar en cuanto a la valoración de los resultados de toda esta investigación. En la criminología actual, aquel pesimismo al que nos hemos referido

líneas arriba ha dado paso a un moderado (pero firme) optimismo sobre la capacidad predictiva de estos instrumentos y sobre su utilidad en el ámbito forense (Andrés Pueyo y Redondo, 2007:169; Andrés Pueyo y Echeburúa, 2010: 408). Se considera que las estimaciones son más precisas (producen un menor número de errores) y, además, más objetivas y rigurosas, carentes de sesgos y basadas en la evidencia empírica. Steadman, uno de los autores que más se significó por sus estudios críticos sobre el estado de las predicciones de peligrosidad hace cuarenta años, afirmaba, sin embargo, en el año 2000 que "incluso teniendo en cuenta las diversas limitaciones del estado actual del conocimiento, existe espacio [hoy] para un optimismo que habría resultado inapropiado en 1970" (Steadman, 2000: 270). La opinión que hoy día sostiene la Asociación Americana de Psicología es que "los enfoques 'estructurados', basados en la ciencia y en los datos empíricos, han demostrado ser más fiables que los enfoques clínicos no estructurados y pueden estimar válidamente la peligrosidad futura en casos apropiados" (American Psychological Association, 2011).

¿Está fundamentado este optimismo? ¿Son realmente más objetivas y precisas las nuevas estimaciones de riesgo? ¿Producen menos errores que los antiguos pronósticos de peligrosidad?

Ante todo, me parece importante destacar que los jueces lo tienen ahora más difícil para responder a estas preguntas por sí mismos, porque el complejo y sofisticado lenguaje matemático en que se expresan los resultados de los estudios sobre las modernas herramientas estructuradas de valoración del riesgo de reincidencia es bastante más complicado de comprender para el profano que los antiguos informes de peligrosidad. Por ejemplo, en un trabajo publicado en 2016 sobre la eficacia predictiva de la valoración policial del riesgo de la violencia de género en España se puede leer lo siguiente:

"E[!] análisis mostró los siguientes resultados: en cuanto a la bondad de ajuste del modelo, se obtuvo un valor en el test de Hosmer y Lemeshow, $X^2 = 0.000$; gl. = 1; $p = 0.003$ y unos valores de $-2LL = 147.084$; la R^2 de Nagelkerke de 0.093 y una R^2 de Cox y Snell de 0.030. Los valores de ajuste del modelo ($\beta = 1.885$, $SE = 0.634$, $Wald = 8.824$ y una $Exp(B) = 6.585$ con un IC 95 % de 1.899-22.835. Estos resultados avalan la utilidad del VPR en la predicción de la reincidencia a 3 meses, ya que en este análisis se ha obtenido una *odds ratio* de 6.585 que es relevante en este tipo de instrumentos. [...]

Posteriormente se procedió a calcular el AUC (fig. 2) para obtener el indicador de la capacidad predictiva. En este cálculo se obtuvo un valor de $AUC = 0.714$ (IC 95 % = 0.607-0.821), que corresponde a un valor aceptable con tendencia a ser calificado de bueno según los expertos del área (Singh, 2013). [...]

La sensibilidad o identificación correcta del riesgo de violencia cuando existe la reincidencia fue del 85 %, y la capacidad del instrumento para descartar el riesgo cuando no se dio reincidencia o especificidad, fue del 53.7 %" (López, González y Andrés-Pueyo, 2016: 5 y 6).

De todo el fragmento transcrito sobre la eficacia predictiva del VPR, lo único que el juez entiende sin problemas es la conclusión ("estos resultados avalan la utilidad del VPR en la predicción de la reincidencia a 3 meses"). Sin embargo, salvo que disponga de conocimientos de estadística para entender qué significa la *odds ratio*, el valor de la AUC y todos los demás indicadores estadísticos que en él se mencionan, el juez no puede valorar críticamente esa conclusión. Es decir, no puede formarse por sí mismo una opinión acerca de si el procedimiento a través del cual se ha alcanzado esa conclusión le parece razonable, si los parámetros que se han calculado son los más relevantes o si es esa conclusión y no otra la que realmente se deduce de todos los valores numéricos allí enumerados. El lenguaje matemático, las tablas, los gráficos, las fórmulas y la multitud de cifras aportan, qué duda cabe, una apariencia de solidez científica y permiten presentar las estimaciones del riesgo de reincidencia como similares a las que podrían hacerse sobre cualquier otro tipo de riesgo como el de recesión económica para los próximos seis meses en la eurozona o la morbilidad asociada a un nuevo virus epidémico (8). Sin embargo, ¿son realmente análogas las valoraciones del riesgo de reincidencia a estas otras estimaciones? ¿Cuánto hemos mejorado, respecto de la situación de hace cuarenta años, con estas estimaciones cuantitativas?

II. "TODOS LOS MODELOS PROBABILÍSTICOS SON FALSOS, PERO ALGUNOS SON ÚTILES" (GEORGE E. P. BOX)

El título de este apartado es una conocida cita de George E. P. Box, un reputado estadístico británico fallecido en 2013 (9). Los trabajos en los que Box expuso esta idea no tenían nada que ver con la criminalidad ni con el sistema de justicia penal. Con todo, si las modernas estimaciones estructuradas de riesgo que se efectúan en relación con la criminalidad realmente son similares a las estimaciones que se hacen en cualquier otro ámbito del conocimiento, no parece inadecuado aplicarles esta cita y las interesantes implicaciones que en ella subyacen, ya que no supone más que aplicar criterios similares a los que un especialista en estadística podría usar para enjuiciar cualquier modelo probabilístico empleado en el estudio de cualquier tipo de materias.

Vayamos, entonces, por partes y preguntémonos, en primer lugar, qué significa que todos los modelos sean falsos.

1. Ningún modelo estadístico refleja de manera completamente fiel la realidad

En primer lugar, significa que cualquier modelo matemático o estadístico es solo una aproximación a la realidad, lo cual a su vez implica dos cosas.

Por un lado, implica que muchas veces el modelo no se elabora con la totalidad de los datos, sino solo con una muestra, que confiamos que represente de manera adecuada las características de la población que estamos estudiando. Cuando se hace una encuesta sobre las preferencias políticas de los ciudadanos no se pregunta a todos los potenciales votantes del país; cuando se estima el número de personas que padecen depresión en la población adulta no se pasan pruebas psicométricas a todos los individuos. La forma de seleccionar la muestra condiciona de manera muy importante la validez de las conclusiones que luego se extraigan y se pretendan aplicar a la totalidad de la población.

Por otro lado, precisamente porque tienen que operar con números y algoritmos, todos los modelos estandarizan y simplifican la información que recogen para poder convertirla en cifras o categorías que permitan hacer cálculos. Por ejemplo, un modelo que pretenda estimar el nivel de formación de los estudiantes universitarios podrá recoger la puntuación numérica de las notas que han obtenido, los años que han tardado en graduarse o las veces que se han tenido que examinar de una misma asignatura para aprobarla, pero todo ello es una información solo aproximada acerca del grado en que realmente cada estudiante domina o no la materia. Un estudiante brillante, pero muy tímido, puede haber tenido bajas notas si en muchas asignaturas ha tenido que enfrentarse a exámenes orales, que se le dan especialmente mal; una universidad privada deseosa de atraer clientes puede aplicar estándares de calificación muy laxos y, así, inflar artificialmente los expedientes de sus graduados; alumnos que hayan aprendido de memoria el Código Penal y el Civil, porque eso era lo que se les exigía en los exámenes para obtener matrícula de honor, pueden ser luego incapaces de aplicar las normas que han memorizado a la solución de problemas jurídicos reales; y así sucesivamente.

Tenemos, por tanto, que cualquier modelo estadístico es siempre incompleto y, por ello, "falso" en alguna medida, sea cual sea el ámbito de la realidad que estemos estudiando. Ahora bien, cuando de lo que estamos hablando es de delincuencia, aparecen problemas muy importantes en la selección y definición de la información con la que trabaja el modelo. Para empezar, la delincuencia no se puede medir directamente, como la presión sanguínea, el nivel de humedad en el aire o la densidad del hielo en los glaciares. Se puede estimar a través de cifras oficiales (de detenciones practicadas, denuncias presentadas, condenas dictadas), pero es sabido que los datos oficiales esconden siempre una cifra negra de criminalidad no denunciada o no descubierta, cifra que, según la tipología delictiva, el nivel de eficacia del sistema de justicia y el nivel de confianza de los ciudadanos en las instituciones, puede ser en ocasiones bastante

elevada. Por otro lado, la actividad policial y la investigación judicial no son neutras: hay colectivos de personas (determinadas minorías étnicas, colectivos de inmigrantes sin papeles...) y zonas en las ciudades (barrios con índices altos de trapicheo de drogas, socioeconómicamente desfavorecidos, etc.) que son objeto de un escrutinio policial mucho más intenso que otros, de manera que la probabilidad de ser detenido (y con ello, de que se inicie un proceso y que finalmente termine con una sentencia condenatoria) varía enormemente en función de estos factores.

Si, debido a los problemas que generan los datos oficiales, intentamos sustituirlos o complementarlos con la información que se puede obtener directamente de los acusados o condenados o de las víctimas, podemos ciertamente obtener datos valiosos, pero tampoco hay garantía de que la información se ajuste a la realidad. Los sujetos pueden mentir deliberadamente, exagerar o disimular por querer causar determinada impresión o para evitar consecuencias que se temen negativas (no se puede olvidar que estamos hablando de datos sobre delitos cometidos o experimentados), o pueden simplemente no mencionar determinados episodios por no darles importancia, por haberlos olvidado o por no entender correctamente lo que se les está pidiendo, etc.

En definitiva, hay que contar con que, cuando hablamos de delincuencia, los datos que recogemos pueden no estar reflejando la realidad en aspectos muy relevantes. Así, si los datos que entran en el modelo son incompletos, sesgados, inexactos o inveraces, la predicción que se elabore sobre dicha base no podrá ser mejor. Por muy sofisticado que sea el algoritmo que después se aplique, no puede corregir los déficits que se arrastran desde la recogida de la información.

2. En cualquier estimación de probabilidad hay siempre errores

Decir que todos los modelos probabilísticos son falsos implica también que las estimaciones que con ellos se realizan no son nunca cien por cien exactas, sino que incluyen un determinado porcentaje de errores. Son, precisamente, estimaciones de probabilidad, luego no es completamente seguro que ocurra lo que se ha predicho.

En relación con los pronósticos de peligrosidad, fue un determinado tipo de errores, los falsos positivos, lo que contribuyó decisivamente a desacreditar en los años setenta y ochenta del siglo pasado los pronósticos clínicos de peligrosidad. Debemos preguntarnos, por tanto, qué tipo y qué cantidad de errores cometemos ahora cuando estimamos la probabilidad de reincidencia con las modernas herramientas estructuradas de valoración del riesgo.

Cuando se intenta contestar a esta pregunta aparece de inmediato una primera dificultad, y es que la sofisticación estadística a la que ya nos hemos referido ha multiplicado las

formas en que puede ofrecerse información sobre el rendimiento de las herramientas de valoración del riesgo. En efecto, los estudios sobre valoraciones de riesgo ya no se limitan a informar sobre cuántas personas de las que se han considerado de alto riesgo vuelven a delinquir o cuántas de las que se han valorado como de bajo riesgo no lo hacen; es más, en ocasiones ni siquiera ofrecen esta información, porque la capacidad predictiva de estos instrumentos se puede expresar con muchos indicadores diferentes, cada uno de los cuales mide una dimensión distinta de esa capacidad predictiva (Singh, 2013; Loinaz, 2017: 87 y ss.; Muñoz y López, 2016): algunos expresan cómo de bien detecta el instrumento la reincidencia y otros, cómo de bien la predice; algunos indican cómo de bien se estima el riesgo alto y otros, cómo de bien se estima el riesgo bajo; algunos son medidas de riesgo relativo y otros, de riesgo absoluto; etc. (10). Además, puesto que cada indicador mide una dimensión diferente de la capacidad predictiva, cada uno puede adoptar valores muy distintos para una misma herramienta de estimación del riesgo. 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. Es decir, la capacidad predictiva de un instrumento de valoración del riesgo puede ser calificada a veces 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.

Esa situación no es necesariamente mala, ni mucho menos. Al contrario, la metodología cuantitativa y los múltiples análisis estadísticos que sobre los datos pueden hacerse ofrecen información sobre aspectos que antiguamente no podían estudiarse con el método clínico, de tipo más "artesanal". Además, la multiplicación de estudios sobre valoración del riesgo de reincidencia ha permitido acumular y sistematizar una gran cantidad de información de la que no se disponía hace cuarenta o cincuenta años. Pero, obviamente, también tiene inconvenientes. Ya hemos mencionado *supra* que la elevada complejidad estadística dificulta a los operadores jurídicos entender el significado de la información que se les transmite sobre las valoraciones del riesgo. Es más, el hecho de que el funcionamiento de una herramienta de valoración del riesgo pueda describirse con muchos indicadores diferentes hace que la información que se proporciona sobre su rendimiento pueda ser fácilmente malinterpretada, especialmente por quien no conozca con precisión el significado de dichos parámetros, en función de qué datos se le faciliten y cómo se le presenten. En todo caso, no es posible profundizar aquí en los problemas que esta "impenetrabilidad" estadística puede comportar (11); lo que interesa en este trabajo es comparar la situación actual con la existente hace unas décadas para ver hasta qué punto los problemas que entonces más preocupaban se han podido resolver o no. Esos problemas eran, sobre todo, las altas tasas de falsos positivos. Hay que comprobar, por tanto, cuáles son las tasas de falsos positivos, y de otro tipo de errores, que arrojan

las estimaciones de riesgo hechas con las modernas herramientas estructuradas de valoración.

A) Algunos estudios recientes en España

Hace unos años se aplicó en Cataluña un instrumento de evaluación del riesgo de violencia de juicio clínico estructurado (el *Sexual Violence Risk-20*, conocido como SVR-20) a un grupo de 163 agresores sexuales que habían quedado en libertad tras cumplir condena (12). Con este instrumento se evaluó a cada uno de ellos y se le asignó una probabilidad de reincidencia, para después comprobar cuántos de ellos habían reincidido en un período de seguimiento promedio de cuatro años. Los autores del estudio pretendían, con ello, ir explorando las posibilidades de aplicar estos instrumentos —tan consolidados ya en el mundo anglosajón— a población española, y contrastar si es posible obtener con su uso predicciones correctas de peligrosidad (13). De los 45 sujetos que el SVR consideró de riesgo alto, el 37,7 % volvió a delinquir. En el otro 62 % de los casos, por tanto, el comportamiento real posterior de los sujetos no corroboró la predicción; fueron falsos positivos. Los valores son muy distintos, sin embargo, si atendemos al grupo de sujetos de riesgo bajo: de 118 personas así calificadas solo 7 delinquieron en el período de seguimiento.

En 2015 se publicó un estudio sobre la tasa de reincidencia penitenciaria en Cataluña (Capdevila *et al.*, 2015), que analiza la reincidencia penitenciaria de todos los liberados de las prisiones catalanas en 2010, con seguimiento hasta el 31 de diciembre de 2013. A 648 sujetos que habían sido condenados por delitos violentos se les aplicó un instrumento de valoración del riesgo desarrollado por el Departament de Justícia de la Generalitat catalana y la Universidad de Barcelona, el RisCanvi, y se hizo un seguimiento de 3,5 años de media. De los 301 sujetos pronosticados como probables reincidentes (130 de riesgo alto más 171 de riesgo moderado) (14), solo 54 reincidieron, es decir, el 17,94 %. Sin embargo, nuevamente encontramos aquí unos porcentajes de acierto mucho más elevados si atendemos al grupo de sujetos considerados de riesgo bajo: respecto de este colectivo (347 personas), 331 no volvieron a cometer un delito violento, lo que significa que se acertó en el 95,4 % de los casos (15).

B) Datos a nivel internacional

Los resultados de estos estudios están en la línea de otros muchos llevados a cabo a nivel internacional. El indicador estadístico que mide el porcentaje de sujetos que efectivamente reinciden respecto del conjunto de los que se habían valorado como de riesgo alto se denomina *valor predictivo positivo*. Dicho valor suele estar por debajo del

50 % en todos los casos y, con frecuencia, muy por debajo, especialmente cuando la delincuencia que se intenta predecir es de tipo violento o sexual. A su vez, el indicador que mide el porcentaje de sujetos que no reinciden respecto del conjunto de los que se habían valorado como de riesgo bajo se denomina *valor predictivo negativo*. Este suele ser muy elevado, cercano o incluso superior al 90 %, según los resultados de los estudios internacionales (16).

Aquí es necesario recordar una cuestión a la que ya nos hemos referido antes: la capacidad predictiva de las herramientas de valoración del riesgo se puede medir de muchas formas diferentes, y para cada herramienta los distintos indicadores pueden arrojar valores muy dispares. Una herramienta puede producir, a la vez, un número muy alto de falsos positivos y uno muy pequeño de falsos negativos. Así, si la información que se da sobre su rendimiento incluye solo este último indicador y no el primero, se ofrece la impresión de que su funcionamiento es mejor (globalmente) de lo que realmente es.

Uno de los indicadores utilizados con más frecuencia para evaluar el rendimiento de las herramientas estructuradas de valoración del riesgo es la llamada *área bajo la curva ROC* (17). Se trata de un indicador de riesgo relativo: indica cuál es 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, es un indicador que informa sobre cómo de bien diferencia la herramienta entre los sujetos de mayor y de menor riesgo, pero no dice cuáles son las probabilidades de riesgo asociadas a cada nivel. Por ejemplo, en una herramienta de valoración, 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 del 75 % (que convencionalmente se considera un valor elevado); mientras que 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 de 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.

La investigación internacional se ha centrado en los últimos años en los indicadores de riesgo relativo, como el área bajo la curva, pero se ha investigado mucho menos sobre cómo de válidas son las estimaciones de riesgo absoluto. Es decir, si una herramienta de valoración del riesgo dice que en la muestra inicial de personas con las que se elaboró la herramienta el grupo de alto riesgo tuvo una tasa de reincidencia del 40 %, ¿cómo de probable es que, aplicando esa herramienta a otros grupos de personas, encontremos tasas de reincidencia parecidas? Sobre este extremo la investigación internacional es

mucho más escasa (Rossegger *et al.*, 2014; Singh, 2013; Helmus *et al.*, 2012; Hanson, 2017).

El hecho de que la consistencia de las tasas de reincidencia observada en diferentes muestras sea una cuestión aún por estudiar, que la información sobre el rendimiento de las herramientas de valoración del riesgo se suela proporcionar con indicadores de riesgo relativo (y no absoluto) y que con frecuencia no se proporcione información sobre los valores predictivos positivos ni sobre las tasas de reincidencia observada en cada nivel de riesgo genera, en conjunto, una visión optimista sobre el funcionamiento de estas herramientas que corresponde, probablemente, con la realidad, pero solo con una parte de ella (18). La capacidad de estos instrumentos en lo tocante a la identificación de los grupos de riesgo alto sigue teniendo, en gran medida, el problema de las altas tasas de falsos positivos que se identificó hace cuarenta años: "Típicamente, más de la mitad de los individuos considerados por las herramientas como de riesgo alto están incorrectamente clasificados - no van a seguir reincidiendo" (Douglas *et al.*, 2017: 135).

C) Una primera conclusión

Llegados a este punto, cabe extraer ya una primera conclusión: ha cambiado la forma en que se estima la probabilidad de reincidencia y también la forma en que se comunica esa información, pero el problema de las altas tasas de falsos positivos se mantiene, en términos bastante similares a como estaba hace cuarenta años. Lo que resulta llamativo es que esta situación no parece provocar a día de hoy, ni mucho menos, la preocupación que generó hace cuarenta o cincuenta años, ni tampoco se extraen de ello las consecuencias para el derecho penal que en aquel momento se propusieron. La cuestión es: ¿por qué?

A mi juicio, hay dos factores principales que explican esta situación. En primer lugar, que no vemos el problema. Como ya he señalado varias veces en las páginas precedentes, la información que nos da la investigación criminológica dominante en materia de valoraciones del riesgo de reincidencia es muy compleja, con múltiples indicadores estadísticos, tablas, gráficos, fórmulas y cifras difíciles de entender para el profano. Y, aunque el problema de los falsos positivos podría evidenciarse sin excesiva dificultad, simplemente destacando en los estudios los valores tan modestos que alcanza el valor predictivo positivo o informando de los porcentajes reales de reincidencia asociados a cada categoría de riesgo, ocurre que estos datos o bien no se ofrecen cuando se evalúa el rendimiento de las herramientas estructuradas de valoración de riesgos o bien se dejan en un segundo plano porque se pone el énfasis en otro tipo de indicadores, como la *odds ratio*, la sensibilidad, la especificidad o, sobre todo, el área bajo la curva.

En segundo lugar, además de no percibir el problema de los falsos positivos con la claridad de hace cuarenta años porque resulta más difícil de entrever entre una miríada de cifras e indicadores matemáticos, otra razón decisiva para que no se tematice ya como un problema grave es que, simplemente, ya no nos produce tanta preocupación como antaño. Al contrario, en la actualidad tanto la opinión pública como el legislador (¿y, quizá, los operadores jurídicos?) están sensibilizados sobre todo por el problema contrario: el de los falsos negativos, es decir, los sujetos que se valoran como de bajo riesgo (y por ello obtienen algún tipo de ventaja, ya sea la suspensión de la pena, la concesión de la libertad condicional, permisos penitenciarios, etc.) y, sin embargo, vuelven a cometer un delito.

Con esta reflexión sobre cuáles son los problemas que más nos preocupan o, mejor, que más nos deberían preocupar, llegamos al tercero de los aspectos implicados en la cita de Box que estamos usando como hilo conductor de este trabajo: "todos los modelos son falsos, pero algunos son útiles". La pregunta relevante no es, por tanto, si el modelo es verdadero, porque nunca lo es del todo, sino si es lo suficientemente útil, bueno, para el uso que le queremos dar. Puesto que en toda predicción hay errores, lo relevante es decidir en qué sentido preferimos equivocarnos: cuántos y qué tipo de errores preferimos cometer y cuáles no estamos dispuestos a asumir, en función de cuáles sean los objetivos que pretendemos conseguir.

3. ¿Son los modelos lo suficientemente buenos para lo que pretendemos?

Para responder a esta pregunta, ante todo, hay que saber cuáles son los fines que pretendemos alcanzar cuando hacemos valoraciones de riesgo en derecho penal. Una vez tengamos claro qué es lo que queremos conseguir, podremos valorar si los modelos, con sus imprecisiones y sus errores, son o no lo suficientemente buenos.

Un primer nivel de respuesta a la pregunta planteada sería remitirse a las situaciones en las cuales la legislación vigente exige que se realicen dichas estimaciones. De esta manera, podríamos contestar diciendo que hacemos valoraciones de riesgo para poder resolver sobre la suspensión de la pena, sobre la imposición de medidas de seguridad, sobre la concesión de la libertad condicional, y así sucesivamente. De hecho, este es uno de los argumentos en los que se basan los partidarios de introducir los métodos estructurados de estimación del riesgo en la justicia penal: ya que la ley exige hacer estimaciones, hagámoslas entonces con la mejor y más moderna metodología disponible. Sin embargo, esta respuesta es insuficiente, porque las decisiones para las que es necesario realizar pronósticos dependen de la configuración contingente de la legislación en un momento y lugar dados y cambian con el tiempo y en función de cuál sea la política criminal que se estime preferible (19). En consecuencia, la pregunta no es

solo para qué cosas se exigen los pronósticos según la regulación vigente hoy, sino para qué cosas tiene sentido que se exijan y para cuáles no, es decir, si dichos supuestos deben ampliarse, restringirse o modificarse. La cuestión por resolver es, por tanto, si queremos que ciertas decisiones —y, en su caso, cuáles— dependan esencialmente de estimaciones de riesgo o no.

Con ello entramos en un nivel de discusión diferente, que es el relativo a los fines que el derecho penal ha de perseguir en general. Porque hacer depender la gravedad, la duración o la forma de ejecución de determinadas consecuencias jurídicas de un pronóstico de reincidencia significa asumir que la prevención especial es el eje alrededor del cual deben girar dichas consecuencias. Ahora bien, siendo dicho fin un fin legítimo y, además, un fin que en su vertiente de prevención especial positiva tiene expreso reconocimiento en nuestra [Constitución \(art. 25.2 CE\)](#), no es el único posible —ni siquiera el más importante— que el derecho penal puede perseguir, como por otro lado ha subrayado ya en numerosas ocasiones el Tribunal Constitucional (20). De hecho, en la discusión en torno a los fines de la pena, la prevención especial —y, sobre todo, la prevención especial negativa— se ha enfrentado desde siempre a una objeción de peso: que entra en conflicto con el principio del hecho. El derecho penal de una democracia liberal respetuosa con los principios de legalidad y de culpabilidad solo debe castigar hechos ilícitos y lesivos para bienes jurídicos, no formas de ser o tipos de persona. Permitir que la gravedad de la sanción varíe mucho en función de lo peligroso (o malvado) que nos parezca el autor es establecer diferencias que no atienden a la gravedad del hecho, sino a las características individuales del sujeto, y en esa medida significa alejarse del más garantista derecho penal del hecho y acercarse al autoritario e ilegítimo derecho penal de autor.

Ahora bien, era más fácil oponerse al derecho penal de autor (en especial a las penas y medidas de corte inocular basadas en la peligrosidad) cuando, además de los principios liberales, podía invocarse también la debilidad científica y la arbitrariedad de los pronósticos de peligrosidad, porque de esta manera podía aducirse que el derecho penal de autor no solo era valorativamente indeseable, sino también inviable desde el punto de vista técnico. Por el contrario, si fuera cierto que las estimaciones del riesgo de reincidencia se pueden hacer de forma científicamente rigurosa y con un elevado grado de acierto, se perdería uno de los dos argumentos, lo que aparentemente debilita la postura defendida.

Sin embargo, el otro argumento subsiste y es el decisivo. Es decir, el hecho de que los métodos de estimación del riesgo de reincidencia funcionen mejor o peor no afecta al problema —que es de tipo valorativo— de que establecer diferencias entre unos y otros sujetos en atención a su mayor o menor riesgo casa mal con valores esenciales de un derecho penal liberal: el principio del hecho, el de culpabilidad, el

de legalidad y seguridad jurídica y el de proporcionalidad. Ocurre algo parecido con otra de las prohibiciones derivadas del principio del hecho: la de castigar por los meros pensamientos. Esto no puede hacerse porque supone adelantar en exceso la intervención penal a un estadio en el que el eventual peligro para los bienes jurídicos es inexistente o excesivamente lejano, aparte de que invade la intimidad de la persona y la libertad ideológica y de pensamiento; pero también porque técnicamente no se puede saber con certeza lo que una persona desea o está pensando, a menos que lo exteriorice con palabras o, sobre todo, con hechos. Esta segunda objeción, que es de tipo técnico, dejaría de ser un problema si mediante sofisticadas técnicas de escáner cerebral o similares pudiéramos llegar a conocer algún día los pensamientos de los demás, pero, obviamente, poder conocerlos no convertiría en legítimo castigarlos penalmente, porque la vulneración de las garantías penales se seguiría produciendo igual.

En otras palabras, acecha aquí el riesgo de la falacia naturalista de derivar del ser el deber ser: no porque sea posible (si es que lo es) estimar con un elevado grado de acierto el riesgo futuro de reincidencia debemos diseñar el sistema penal en función de dichas estimaciones. Conviene estar alerta ante el peligro de que el optimismo en relación con el rendimiento de las estimaciones estructuradas de riesgo acabe modificando inadvertidamente nuestras percepciones sobre el castigo legítimo (21). Los objetivos que deben perseguir las sanciones penales no se deducen de los modelos estadísticos con que se estima el riesgo de reincidencia ni de que estos funcionen mejor o peor. Dichos objetivos deben establecerse de forma autónoma, en función de cuáles sean las finalidades legítimas y los límites del derecho penal en una democracia liberal respetuosa con los derechos fundamentales. Si están claros dichos fines y dichos límites, podremos comprobar, a continuación, hasta qué punto las estimaciones estructuradas del riesgo de reincidencia se ajustan a ellos.

De todo ello deriva, como primera consecuencia, que la prevención especial no debe erigirse en la finalidad única ni principal de las penas, ni siquiera aun en el caso de que las estimaciones del riesgo de reincidencia tuvieran unos niveles de acierto muy superiores a los que hoy en día evidencian, porque un sistema configurado de esta manera vulneraría los principios limitadores básicos del *ius puniendi*, especialmente el de culpabilidad, pero, a mi juicio, también los de igualdad, lesividad, proporcionalidad y dignidad de la persona. La prevención especial sí puede, no obstante, jugar un papel dentro de dichos límites, más si cabe en un país como el nuestro, donde la propia norma suprema establece que la reeducación y la reinserción social son las finalidades a las que deben estar orientadas las penas privativas de libertad y las medidas de seguridad. De esta manera, llegados a este punto, podemos reformular la pregunta ya de forma más concreta: ¿son suficientemente buenos los modelos de estimación del riesgo como

para diferenciar la respuesta penal en función del nivel de riesgo del sujeto, dentro de lo que permiten los principios de legalidad, culpabilidad y proporcionalidad?

Pues bien, a mi juicio, estos modelos podrían considerarse bastante buenos cuando se trata de favorecer alternativas a la prisión y flexibilizar el régimen penitenciario para las personas valoradas como de bajo riesgo, porque, como hemos visto *supra*, los valores predictivos negativos suelen ser muy elevados, por encima incluso del 90 %. Es decir, respecto del grupo de sujetos calificados como de bajo riesgo por una de estas herramientas estructuradas, generalmente hay una probabilidad muy alta — cercana o incluso superior al 90 %— de que dicho pronóstico se vea corroborado por el comportamiento real de los sujetos. En estos casos la valoración del riesgo me parece un argumento sólido desde el punto de vista de la calidad de la evidencia aportada, porque los datos demuestran que el nivel de acierto es elevado; pero, lo que es más importante: es también un buen argumento porque encaja con los fines propios del ordenamiento jurídico penal, que no pretende solo evitar la reincidencia a cualquier precio, sino también minimizar la violencia ejercida por el aparato estatal, y que además, según el [art. 25.2 CE](#), debe favorecer las posibilidades de reinserción social. La asunción de un pequeño número de falsos negativos es un coste que, desde esta perspectiva valorativa, la sociedad debe asumir (22).

Por el contrario, considero que los modelos de estimación del riesgo de reincidencia no son lo suficientemente buenos cuando se trata de fundamentar agravaciones de la pena o de su régimen de ejecución para las personas catalogadas como de alto riesgo, porque los datos demuestran que la probabilidad de que dicha estimación acabe corroborada por la realidad es baja (lo que los estudios evidencian es que solo menos de la mitad reincidirán; generalmente, muchos menos). La calidad de la evidencia aportada es, por tanto, mucho más débil que en el supuesto anterior. Además, y lo que es más importante, tomar decisiones aflictivas para el sujeto sobre la base de una evidencia tan débil no encaja con los fines que persigue el derecho penal y los principios que lo deben inspirar y limitar. Como acabamos de recordar en el párrafo anterior, el sistema de justicia penal no pretende solo evitar la reincidencia a cualquier precio, sino también minimizar la violencia ejercida por el aparato estatal. Ello supone imponer restricciones de tipo normativo a algunas prácticas que quizá permitirían una lucha más eficaz contra la criminalidad (23). La prioridad en nuestro sistema de justicia penal no es asegurar que quien resulta sospechoso sea en todo caso castigado ni que quien quizá supone un peligro no vuelva a hacer daño a nadie; la prioridad (derivada de los derechos reconocidos en el [art. 24.2 CE](#), así como del principio de legalidad y de la orientación de las penas a la reinserción social establecida en el [art. 25.2](#)) es asegurar que quien no lo merece no recibe un castigo indebido o excesivo. Por tanto, si esto es así, aducir como fundamento para tomar una decisión restrictiva de los derechos del condenado o acusado la existencia

de un pronóstico de riesgo alto no es un buen argumento, pues sabemos (nos lo dicen los estudios sobre la capacidad predictiva de las herramientas de riesgo de reincidencia) que para más de la mitad de los sujetos clasificados de esta manera el pronóstico no se va a cumplir, es decir, son falsos positivos. Sacrificarlos en aras de un posible aumento de la seguridad en algún grado es, desde el punto de vista de los objetivos que queremos conseguir con el sistema penal, un sacrificio excesivo.

La trascendencia de lo que hasta aquí he desarrollado como principios generales varía mucho en función del tipo de situación en la que haya que hacer pronósticos de reincidencia, pues no todas las decisiones judiciales (o penitenciarias) afectan con la misma intensidad a las garantías básicas de legalidad/seguridad jurídica, culpabilidad y proporcionalidad. Por ejemplo, establecer regímenes que restrinjan los derechos del sujeto más allá del margen permitido por la pena adecuada a la gravedad del hecho y de la culpabilidad sobre la base de pronósticos de riesgo elevado —como, por ejemplo, sumar a la pena medidas de seguridad adicionales— creo que es inaceptable, porque directamente las vulnera. En este sentido, me parece particularmente problemático el régimen de revisión de la pena de prisión permanente, que, según lo establecido en el [art. 92.1.c CP](#), depende básicamente del pronóstico de reincidencia, pues, cuando dicho pronóstico arroje un resultado de riesgo alto, la evidencia empírica que lo sustenta es excesivamente endeble como para justificar la permanencia en prisión por tiempo indefinido.

Sin embargo, cuando la ley permite (u obliga a) utilizar la probabilidad de reincidencia como un criterio a la hora de ajustar las penas al caso concreto —dentro del marco previamente establecido en función de la magnitud del injusto culpable (24)—o a la hora de tomar decisiones sobre el modo de ejecución de una pena que ya se ha impuesto en atención a la gravedad del injusto y de la culpabilidad, las decisiones que se tomen basadas en juicios de pronóstico pueden ser legítimas si están adecuadamente fundamentadas. Si la decisión es en beneficio del sujeto, la existencia de una valoración del riesgo con resultado de riesgo bajo es un argumento potente en el que fundamentarla, por las razones que acabamos de ver. Si la decisión, por el contrario, es en perjuicio del reo, entonces la existencia de una valoración de riesgo alto no es, por sí sola, un buen argumento (por los altos niveles de falsos positivos), de manera que, si el juez quiere apurar el marco penal hasta el máximo (denegar la suspensión de la pena, la libertad condicional o la progresión de grado), deberá razonar su decisión aportando otras razones (25).

III. "EVIDENCIAS EMPÍRICAS" Y DERECHO PENAL

Antes de terminar, quisiera añadir algunas consideraciones sobre el tema más genérico de la denominada *evidencia empírica o prácticas basadas en la evidencia* (*evidence-based practices*) y el derecho penal, que quizá pueden contribuir a contextualizar todo lo que hasta aquí he expuesto sobre la prueba del riesgo de reincidencia.

En relación con las valoraciones del riesgo de reincidencia, es cada vez más frecuente escuchar el argumento de que el derecho no puede ignorar la evidencia científica, de que esta debe ser tomada en cuenta para conseguir una administración de justicia "mejor". En el contexto en el que se realizan estas consideraciones, "mejor" normalmente significa más objetiva, con mejor fundamento, menos sesgada por prejuicios personales (26). Suele afirmarse que, ya que las normas obligan muchas veces a hacer pronósticos, mejor hacerlos con un método riguroso que sometidos a la subjetividad, los sesgos y quién sabe si la arbitrariedad del juez o el funcionario de turno. Ya que se tienen que hacer —sería el argumento—, hagámoslos de la mejor manera posible. Y lo cierto es que, en este ámbito de la valoración del riesgo, dichos argumentos están teniendo una acogida bastante entusiasta por parte de los poderes públicos: en el ámbito anglosajón las valoraciones del riesgo están muy extendidas, especialmente en la ejecución penitenciaria, pero también, cada vez más, en las distintas fases del proceso penal (incluidas la adopción de medidas cautelares y la utilización de los niveles de riesgo para graduar la gravedad de la pena en la sentencia) (27); y, en nuestro contexto más cercano, la comunidad autónoma de Cataluña ha implementado una herramienta estructurada de valoración del riesgo en la gestión de todo el sistema penitenciario (el RisCanvi).

Sin embargo, existen otras evidencias empíricas relacionadas con la delincuencia y su castigo que no parece que quienes tienen en su mano diseñar las políticas criminales estén igual de dispuestos a incorporar. Por ejemplo, se sabe desde hace tiempo que el aumento en la severidad de los castigos no se correlaciona, en general, con un descenso de la criminalidad. Tampoco se ha podido demostrar que la pena de muerte reduzca la tasa de homicidios, mientras que sí hay evidencia científica —con pruebas de ADN, pero no solo— de que no pocos condenados a muerte son inocentes. La psicología ha demostrado que las penas de prisión superiores a quince años producen deterioros graves e irreversibles de las capacidades cognitivas de los delincuentes. También hay evidencia de que la política de guerra sin cuartel contra las drogas, con la persecución criminal de todos los actos relacionados con el tráfico a pequeña escala, no ha logrado reducir el tráfico de estas sustancias ni sus terribles consecuencias para la salud. No obstante, ninguna de estas evidencias ha conseguido aún erradicar la pena de muerte en los países occidentales, revertir la tendencia del incremento de la severidad de las penas ni cambiar la política criminal sobre drogas (28).

Esto indica que el hecho de que algunas evidencias empíricas sí se utilicen como argumentos para introducir cambios en el funcionamiento de los sistemas de justicia penal (y otras no) no depende principalmente de que unas sean más sólidas que las otras ni de que estén más y mejor respaldadas por datos. Depende, sobre todo, de que estén en sintonía con los objetivos políticos y de control social —expresos o implícitos— que en cada caso se persiguen con la política criminal, y las valoraciones estructuradas del riesgo de reincidencia encajan bien con dos tendencias que caracterizan la evolución del sistema penal contemporáneo.

Por un lado, encaja con la tendencia a la anticipación de la intervención penal: no esperar a que el daño se haya producido para castigarlo, sino intentar prevenirlo impidiendo las conductas que eventualmente lo llegarían a producir. Una de las varias manifestaciones en que se concreta esta tendencia general (29) es la atención prestada al nivel de riesgo de cada delincuente para adaptar a dicho riesgo la sanción penal. Así, para que esto pueda percibirse como una práctica racional y coherente, es preciso subrayar que el riesgo de reincidencia o de violencia de cada individuo puede estimarse de manera fiable con ayuda de metodologías científicas rigurosas. El énfasis en la capacidad predictiva de las valoraciones estructuradas del riesgo de reincidencia ayuda a hacer más aceptable una política criminal orientada a la prevención especial, negativa si es necesario.

Por el otro lado, insistir en que el conocimiento sobre la criminalidad que proporcionan estas herramientas se debe aplicar en el sistema de justicia penal está también en consonancia, según creo, con la política neoliberal dominante de colocar el delito en el centro de cualquier agenda política, priorizándolo sobre cualquier otro problema social (30), y de pretender calmar ansiedades e inseguridades de los ciudadanos debidas a cambios sociales estructurales muy profundos (paro, cambio climático, desmantelamiento del estado del bienestar, etc.) a través de una sobreactuación penal que, supuestamente, aumentará nuestros niveles de seguridad frente al delito (populismo punitivo).

La funcionalidad de las modernas estimaciones estructuradas de riesgo para este tipo de política deriva de que la información que proporcionan se pretende utilizar solo o principalmente en el seno del proceso penal y en la ejecución penitenciaria, pero no se reclama extraer consecuencias en otros ámbitos, lo cual encierra una paradoja. Para advertirla conviene recordar que el moderno enfoque de la valoración del riesgo se presenta como superador del viejo paradigma de la peligrosidad, entre otras cosas, porque ya no clasifica a sujetos en peligrosos o no peligrosos; ha dejado de considerar la peligrosidad como una característica subjetiva —patológica— del individuo y, en lugar de ello, estima probabilidades objetivas de ocurrencia de un evento futuro (un nuevo delito). En coherencia con este punto de partida, se subraya que dicha probabilidad depende, en parte, de factores personales (edad, sexo, patología mental, personalidad,

etc.), pero también, en muy buena medida, de factores contextuales o situacionales. Por ello, muchas herramientas de valoración incorporan factores de este tipo (por ejemplo, estar empleado o en paro, disponer o no de apoyo familiar y social, disponer o no de recursos económicos, el nivel educativo alcanzado, que haya o no antecedentes delictivos en la familia de origen, la pertenencia a un grupo social de riesgo, etc.), porque han comprobado que se correlacionan de manera estadísticamente significativa con un incremento de la probabilidad de reincidencia. En otras palabras, las valoraciones estructuradas de riesgo parecen evidenciar —es evidencia empírica— una correlación bastante fuerte entre marginación social y económica, delincuencia y prisión.

La paradoja a la que me refiero radica en que, si la investigación criminológica está evidenciando que hay factores contextuales que se correlacionan muy significativamente con el riesgo de reincidencia y los poderes públicos quieren utilizar esa información para reducir los niveles de delincuencia, entonces esa información no debería servir solo para tomar decisiones sobre los individuos en el seno del proceso penal o de la ejecución penitenciaria, sino también para actuar sobre los factores contextuales de riesgo que están asociados a la criminalidad (la evidencia empírica así lo demuestra). La información que proporcionan las herramientas de valoración del riesgo sería un muy buen argumento para implementar toda una serie de políticas sociales que, al reducir la marginalidad y la falta de oportunidades, ayudarían a reducir los niveles de delincuencia: un esfuerzo de las administraciones públicas por mejorar los sistemas de educación públicos para asegurar una verdadera igualdad de oportunidades para todos, las políticas de vivienda, los servicios sociales, la sanidad pública (especialmente en salud mental), una política fiscal progresiva y redistributiva para poder financiarlo, etc. Si nos parece que las correlaciones entre factores contextuales de riesgo y aumento de la probabilidad de reincidencia que detectan las herramientas estructuradas de valoración son una evidencia empírica fiable y válida, entonces no tiene sentido limitar el alcance de esta información al seno del proceso penal o de la ejecución penitenciaria.

Los jueces o los funcionarios de prisiones o de servicios de gestión de penas y medidas alternativas solo muy limitadamente tienen la posibilidad de incidir sobre los factores sociales. Si el nivel de riesgo de la persona es elevado, se le denegará la libertad condicional, y si es bajo, se le concederá, y el hecho de que ese nivel de riesgo esté condicionado por factores contextuales —falta de familia que la acoja, de oportunidades de trabajo, de lugar donde vivir y de medios con que mantenerse— no puede conducir, dentro del proceso penal, a una intervención que altere esos factores (el juez no puede proporcionar al sujeto educación, casa ni trabajo), sino solo a una decisión que afectará únicamente al sujeto individual. Así, la evaluación de riesgo, que supuestamente era una estimación de la probabilidad de ocurrencia de un evento futuro (un nuevo delito) atendiendo no solo a las características personales del individuo, sino al contexto en que

este se encuentra, termina desembocando en una decisión que solo amplía o restringe los derechos del individuo afectado y que se desentiende de actuar sobre los factores contextuales que también influyen de manera decisiva en ese riesgo. A mi modo de ver, de esta manera, las estimaciones estructuradas de riesgo terminan quedándose bastante cerca de la antigua clasificación de sujetos en peligrosos y no peligrosos.

No pretendo afirmar con esto que dentro del sistema penal-penitenciario sea imposible actuar sobre los factores de riesgo, tanto individuales como sociales; por el contrario, se puede hacer —y, además, creo que es conveniente hacerlo—, pero en la medida en que sea posible, que es una medida limitada. Además, por supuesto, todo lo que estoy diciendo depende del grado en que el sistema de justicia penal y de ejecución de penas existente en cada lugar y momento dado disponga de la infraestructura y los medios suficientes para adaptarse a esos factores contextuales. Si, por ejemplo, hay un modelo potente de acompañamiento a los liberados condicionales que les proporcione asistencia y ayuda en el tránsito a la libertad, asesoramiento en la búsqueda de empleo, de vivienda o de ayudas públicas, etc., todo ello contribuirá a disminuir el riesgo asociado a esos factores y permitirá, a la vez, adoptar decisiones favorables al reo, aun en presencia de un riesgo inicialmente medio o incluso elevado, porque cabe confiar en que el propio sistema está implementando medidas con las que contrarrestar ese riesgo. Por lo que hace a nuestro país, en cualquier caso, dicha infraestructura es simplemente inexistente. Por tanto, en ausencia de estos medios de gestión del riesgo, la estimación del riesgo se asemeja bastante a un mero pronóstico de peligrosidad (31).

Sin embargo, incluso aunque los medios para la supervisión y el acompañamiento a los condenados a penas y medidas alternativas fueran mucho mejor de lo que son, no deja de ser cierto que la información que proporcionan las herramientas de valoración del riesgo no tiene por qué limitarse al ámbito de la prevención terciaria. La evidencia empírica que proporcionan es un muy buen argumento para fomentar medidas de prevención primaria destinadas a reducir la reincidencia a través de políticas no penales, porque los modelos estructurados de estimación del riesgo sí pueden considerarse lo suficientemente buenos —en los términos de Box— para este tipo de finalidad: las altas tasas de falsos negativos no son un obstáculo cuando de lo que se trata no es de restringir derechos a nadie, sino de implementar políticas que amplían o mejoran oportunidades.

Lo mismo ocurre con toda una serie de consideraciones normativas que en este trabajo no ha habido espacio para comentar, pero que suponen obstáculos si el riesgo se utiliza como criterio en decisiones penales o penitenciarias y dejan de serlo cuando se trata de otro tipo de contextos. Por ejemplo: ¿es justo denegar la suspensión de la pena o la libertad condicional por razones como estar desempleado, ser soltero, ser joven, tener hermanos en prisión o vivir en un determinado barrio, si muchas de estas cosas son

situaciones que el sujeto no puede cambiar o de las que no se le puede considerar responsable, aunque estadísticamente se correlacionen con un nivel de riesgo elevado? Sin embargo, políticas que mejoren las condiciones de vida y los servicios básicos en barrios desfavorecidos y políticas de fomento del empleo juvenil de ninguna manera pueden considerarse injustas, y contribuyen a disminuir algunos de los factores de riesgo de la delincuencia.

En definitiva, no digo nada que no sea de sobra conocido cuando afirmo que centrar el discurso de la prevención del delito únicamente en las posibilidades de actuación del sistema penal es como centrar el discurso de la lucha contra los incendios forestales en la cantidad de efectivos y de medios de que deben disponer los consorcios provinciales de bomberos: con ser esto, sin duda alguna, un factor muy importante, de poco servirá si no hay una política forestal seria de limpieza y mantenimiento del monte, una regulación sensata de los usos agrícolas y recreativos, una reforestación con las variedades más adecuadas al clima y la orografía, una prohibición de destinar suelos quemados a promociones urbanísticas, etc. Insistir en que se usen los resultados de la investigación en materia de riesgo de reincidencia principalmente para adoptar decisiones sobre el tipo, la gravedad y la forma de ejecución de las condenas en el seno del proceso penal y penitenciario es poner el foco solo en el sistema de control penal (en el peor de los casos, en la inocuidad; y, en el mejor, en la prevención terciaria) y renunciar a implementar políticas no penales que, según la evidencia empírica que proporcionan estas mismas herramientas estructuradas de valoración del riesgo, muy probablemente redundarían en una menor tasa de delincuencia global, y sin provocar los problemas de colisión con algunos principios penales que genera una política penal centrada en la prevención especial.

IV. CONCLUSIONES

Cualquier política criminal que pretenda influir de algún modo en los tipos y en los niveles de criminalidad necesita, qué duda cabe, conocer la realidad de la delincuencia. Si hay situaciones, características personales o contextos sociales que incrementan las posibilidades de que ocurran delitos, los poderes públicos deben disponer de esa información, y para ello las herramientas de valoración del riesgo pueden ser útiles. Ahora bien, su utilización en el proceso penal y en la ejecución de las penas debe tener en cuenta las siguientes cautelas:

1. La información que dan estas herramientas no refleja toda la realidad. Es solo información cuantitativa, muy relevante —desde luego— para conocer cualquier fenómeno social, pero no completa. Los datos son parte de la realidad, pero no son toda

la realidad. Además, precisamente en relación con el delito, las dificultades de recogida de datos fiables que reflejen los niveles reales de criminalidad son particularmente complejas.

2. A pesar de esa primera cautela, la información que proporcionan las valoraciones estructuradas de riesgo es muy interesante y merece ser tomada en cuenta, pero siempre que se comprenda correctamente su significado, que se valore su utilidad real para los fines que se pretenden alcanzar con ella y que esos fines sean legítimos. A este respecto, hay al menos cuatro aspectos que considerar:

- i. La metodología estadística hace más difícil que antes entender el significado de las cifras que aportan los estudios sobre el riesgo de reincidencia. Antes de utilizar las valoraciones estructuradas de riesgo en el proceso penal, los operadores jurídicos deberían tener muy claro qué significa cada indicador.
- ii. El mejor o peor funcionamiento de los métodos de valoración del riesgo de reincidencia es un dato que se debe tener en cuenta a la hora de diseñar la política criminal, pero no es un argumento suficiente para dar mayor protagonismo a la prevención especial en el sistema de justicia penal. Insistir en el carácter científico de las valoraciones del riesgo oscurece esta cuestión, que es valorativa y que es la esencial desde el punto de vista del derecho y de los derechos fundamentales.
- iii. En todo caso, lo que el estado actual de los conocimientos sobre el tema permite concluir es que, si bien la capacidad de discriminación entre delincuentes de mayor o menor riesgo parece relativamente buena, es insuficiente la investigación empírica sobre la calibración de las herramientas estructuradas de valoración del riesgo. Además, aunque en ciertos aspectos las estimaciones de riesgo hechas con la moderna metodología estructurada alcancen niveles altos de validez predictiva, en otros aspectos no es así, y en particular no lo es cuando se trata de identificar a los delincuentes con mayor riesgo de cometer delitos graves. En este grupo, los niveles de falsos positivos siguen siendo muy altos, en ocasiones no muy diferentes de los que se criticaron por inasumibles cuando las predicciones de peligrosidad se hacían con el método clínico tradicional. Este problema no es suficientemente destacado, a mi juicio, por la investigación criminológica actual, y puede generar un optimismo exagerado sobre el rendimiento de las estimaciones estructuradas de riesgo.
- iv. En consecuencia, puede ser admisible —desde los principios que inspiran un derecho penal respetuoso con los derechos fundamentales

— utilizar valoraciones de riesgo bajo para fundamentar decisiones que favorecen al sujeto, pero no lo es utilizar valoraciones de riesgo alto como argumento exclusivo para tomar decisiones que le perjudican. Esta afirmación ha de matizarse, a su vez, en función de las características de cada decisión, el grado en que limite derechos fundamentales, el tiempo que vaya a durar dicha restricción, las medidas alternativas de control de dicho riesgo de que disponga el sistema, etc.

3. El proceso penal no es el único ámbito en el que puede tener sentido aplicar la información sobre factores de riesgo y factores protectores que proporcionan los modelos estructurados de valoración del riesgo de reincidencia. En la medida en que muchos de los factores que, según estos modelos, se correlacionan significativamente con la reincidencia son contextuales, dicha información (basada en evidencia empírica) podría igualmente utilizarse para: (i) justificar la inversión de recursos en una infraestructura potente de acompañamiento y asistencia a las personas sometidas a penas y medidas alternativas a la prisión, y (ii) diseñar políticas (no penales) que contribuyeran a la prevención primaria de la delincuencia. Renunciar a reclamar estas aplicaciones y optar por (o conformarse con) que las valoraciones de riesgo se usen solo para tomar decisiones judiciales o penitenciarias individuales en el reducido marco de los recursos disponibles permite a los poderes públicos, a mi modo de ver, legitimar con la etiqueta de "científica" la política criminal securitaria que padecemos.

Bibliografía

1. ALONSO RIMO, Alberto. ¿Impunidad general de los actos preparatorios?: La expansión de los delitos de preparación. En: *Indret: Revista para el Análisis del Derecho*. 2017, n.º 4. ISSN-e 1698-739X. [Consultado el 1 de febrero de 2020]. Disponible en: <https://indret.com/wp-content/uploads/2017/12/1337.pdf>.
2. AMERICAN PSYCHIATRIC ASSOCIATION. Barefoot v. Estelle: Brief Amicus Curiae. 1982. [Consultado el 15 de febrero de 2020]. Disponible en: <https://www.psychiatry.org/psychiatrists/search-directories-databases/library-and-archive/amicus-briefs>.
3. AMERICAN PSYCHOLOGICAL ASSOCIATION. Coble v. Texas: Brief Amicus Curiae. 2011. [Consultado el 15 de febrero de 2020]. Disponible en: <https://www.apa.org/about/offices/ogc/amicus/index-issues>.
4. ANDRÉS PUEYO, Antonio. Peligrosidad criminal: análisis crítico de un concepto polisémico. En: *Neurociencias y derecho penal: nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad*. Demetrio Crespo, Eduardo (dir.). Madrid: Edisofer, 2013, pp. 483-503. ISBN 978-84-15276-15-9.

5. ANDRÉS PUEYO, Antonio y ECHEBURÚA ODRIOZOLA, Enrique. Valoración del riesgo de violencia: instrumentos disponibles e indicadores de aplicación. En: *Psicothema*. 2010, vol. 22, n.º 3, pp. 403-409. ISSN 0214-9915.
6. ANDRÉS PUEYO, Antonio y REDONDO ILLESCAS, Santiago. Predicción de la violencia: entre la peligrosidad y la valoración del riesgo de violencia. En: *Papeles del Psicólogo*. 2007, vol. 28, n.º 3, pp. 157-173. ISSN 0214-7823.
7. ARBACH-LUCIONI, Karin, et al. La práctica de la evaluación del riesgo de violencia en España. En: *Revista de la Facultad de Medicina*. 2015, vol. 63, n.º 3, pp. 357-366. ISSN 0120-0011.
8. AUERHAHN, Kathleen. *Selective incapacitation and public policy. Evaluating California's imprisonment crisis*. Nueva York: State University of New York Press, 2003. ISBN 978-0791457986.
9. BOX, George E. P. Science and Statistics. En: *Journal of the American Statistical Association*. 1976, vol. 71, n.º 356, pp. 791-799.
10. BOX, George E. P, LUCEÑO VÁZQUEZ, Alberto y PANIAGUA-QUIÑONES, María del Carmen. *Statistical Control by Monitoring and Adjustment*. 2.ª ed. Hoboken-New Jersey: John Wiley & Sons, 2009. ISBN 978-0-470-14832-7.
11. CAPDEVILA CAPDEVILA, Manel, et al. *Tasa de reincidencia penitenciaria 2014*. Barcelona: Centre d'Estudis Jurídics i Formació Especialitzada, Generalitat de Catalunya, 2015. [Consultado el 1 de febrero de 2020]. Disponible en: http://cejfe.gencat.cat/web/.content/home/recerca/cataleg/crono/2015/taxa_reincidencia_2014/tasa_reincidencia_2014_cast.pdf.
12. CERVELLÓ DONDERIS, Vicenta. Peligrosidad criminal y pronóstico de comportamiento futuro en la suspensión de la ejecución de la pena. En: *La Ley Penal*. 2014, n.º 106, p. 4. ISSN 1697-5758.
13. DIAMOND, Bernard L. The psychiatric prediction of dangerousness. En: *University of Pennsylvania Law Review*. 1974, vol. 123, pp. 439-452.
14. DOUGLAS, T, et al. Risk assessment tools in criminal justice and forensic psychiatry: The need for better data. En: *European Psychiatry*. 2017, n.º 42, pp. 134-137. ISSN 0924-9338.
15. FAZEL, S, et al. Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24827 people: systematic review and meta-analysis. En: *British Medical Journal*. 2012, 345:e4692.
16. HANSON, R. Karl. Assessing the Calibration of Actuarial Risk Scales. A Primer on the E/O Index. En: *Criminal Justice and Behavior*. 2017, vol. 44, n.º 1, pp. 26-39. DOI: 10.1177/0093854816683956.
17. HARCOURT, Bernard E. *Against prediction: profiling, policing, and punishing in an actuarial age*. Chicago: University of Chicago Press, 2007. ISBN 9780226316147.

18. HEILBRUN, Kirk. *Evaluation for risk of violence in adults*. Nueva York: Oxford University Press, 2009. ISBN 978-0195369816.
19. HELMUS, Leslie, et al. Absolute recidivism rates predicted by Static-99r and Static-2002r sex offender risk assessment tools vary across samples. A meta-analysis. En: *Criminal Justice and Behavior*. 2012, 39(9), pp. 1148-1171. DOI: 10.1177/0093854812443648.
20. KINZIG, Jörg. *Die Sicherungsverwahrung auf dem Prüfstand. Ergebnisse einer theoretischen und empirischen Bestandsaufnahme des Zustandes einer Maßregel*. Friburgo: Edition Iuscrim, 1996. ISBN 978-3861130185.
21. KRAUSS, Daniel A y SCURICH, Nicholas. Risk Assessment in the Law: Legal Admissibility, Scientific Validity, and Some Disparities between Research and Practice. En: *Behavioral Sciences and the Law*. 2013, n.º 31, pp. 215-229. DOI: 10.1002/bsl.2065.
22. LOINAZ CALVO, Ismael. *Manual de evaluación del riesgo de violencia. Metodología y ámbitos de aplicación*. Madrid: Pirámide, 2017. ISBN 978-84-368-3708-7.
23. LÓPEZ OSSORIO, Juan José, GONZÁLEZ ÁLVAREZ, José Luis y ANDRÉS PUEYO, Antonio. Eficacia predictiva de la valoración policial del riesgo de la violencia de género. En: *Psychosocial Intervention*. 2016, vol. 25, n.º 1, pp. 1-7. ISSN 2173-4712.
24. MARTÍNEZ GARAY, Lucía. La incertidumbre de los pronósticos de peligrosidad: consecuencias para la dogmática de las medidas de seguridad. En: *InDret. Revista para el Análisis del Derecho*. 2014, n.º 2. ISSN-e 1698-739X. [Consultado el 1 de febrero de 2020]. Disponible en: <https://indret.com/wp-content/themes/indret/pdf/1043.pdf>.
25. MARTÍNEZ GARAY, Lucía. Errores conceptuales en la estimación de riesgo de reincidencia. La importancia de diferenciar sensibilidad y valor predictivo, y estimaciones de riesgo absolutas y relativas. En: *Revista Española de Investigación Criminológica*. 2016, n.º 14. ISSN-e 1696-9219. [Consultado el 1 de febrero de 2020]. Disponible en: <https://reic.criminologia.net/index.php/journal/article/view/97/94>.
26. MARTÍNEZ GARAY, Lucía. Peligrosidad, algoritmos y "due process": el caso State vs. Loomis. En: *Revista de Derecho Penal y Criminología*. 2018, n.º 20, pp. 485-502. ISSN 1132-9955.
27. MARTÍNEZ GARAY, Lucía. La relación entre culpabilidad y peligrosidad. En: *La culpabilidad*. Maraver, Mario y Pozuelo, Laura (coord.). Montevideo-Buenos Aires: Editorial BdeF, 2019, pp. 115-199. ISBN 9789974745766.
28. MARTÍNEZ GARAY, Lucía y MONTES SUAY, Francisco. El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias. En: *InDret. Revista para el Análisis del Derecho*. 2018, n.º 2. ISSN-e 1698-739X.

- [Consultado el 1 de febrero de 2020]. Disponible en: <https://indret.com/wp-content/uploads/2019/01/1375.pdf>.
29. MONAHAN, John y SKEEM, Jennifer L. Risk Assessment in Criminal Sentencing. En: *Annual Review of Clinical Psychology*. 2016, 12(1), pp. 489-513. DOI: 10.1146/annurev-clinpsy-021815-092945.
 30. MUÑOZ VICENTE, José Manuel y LÓPEZ OSSORIO, Juan José. Valoración psicológica del riesgo de violencia: alcance y limitaciones para su uso en el contexto forense. En: *Anuario de Psicología Jurídica*. 2016, n.º 26, pp. 130-140. ISSN 1133-0740.
 31. PÉREZ RAMÍREZ, Meritxell, et al. Predicción de riesgo de reincidencia en agresores sexuales. En: *Psicothema*. 2008, vol. 20, n.º 2, pp. 205-210. ISSN 0214-9915.
 32. REDONDO ILLESCAS, Santiago, PÉREZ RAMÍREZ, Meritxell y MARTÍNEZ GARCÍA, Marian. El riesgo de reincidencia en agresores sexuales: investigación básica y valoración mediante el SVR-20. En: *Papeles del Psicólogo*. 2007, vol. 28, n.º 3, pp. 187-195. ISSN 0214-7823.
 33. ROSSEGER, Astrid, et al. Replicating the Violence Risk Appraisal Guide: A Total Forensic Cohort Study. En: *PLoS ONE*. 2014, 9 (3), e91845. DOI:10.1371/journal.pone.0091845.
 34. SCURICH, Nicholas. Structured Risk Assessment and Legal Decision-Making. En: *Advances in Psychology and Law*. Miller, Monica K y Bornstein, Brian H (edit.). Cham: Springer International Publishing Switzerland, 2016, pp. 159-183. ISBN 978-3319294056.
 35. SIMON, Jonathan. *Gobernar a través del delito*. Barcelona: Gedisa, 2012. ISBN 978-84-9784-305-8.
 36. SINGH, Jay P. Predictive validity performance indicators in violence risk assessment: a methodological primer. En: *Behavioural Sciences and the Law*. 2013, 31(1), pp. 8-22. DOI: 10.1002/bsl.2052.
 37. STEADMAN, Henry J. Implications from the Baxstrom Experience. En: *Journal of the American Academy of Psychiatry and the Law*. 1973, vol. 1 (3), pp. 189-196.
 38. STEADMAN, Henry J. From Dangerousness to Risk Assessment of Community Violence: taking stock at the turn of the century. En: *Journal of the American Academy of Psychiatry and the Law*. 2000, vol. 28 (3), pp. 265-271.
 39. TONRY, Michael. Evidence, Ideology, and Politics in the Making of American Criminal Justice Policy. En: *Crime and Justice*. 2013, vol. 42, n.º 1, pp. 1-18. DOI: 10.1086/671382.

Notas

- (1) Este trabajo se ha realizado en el marco de los proyectos de I+D+i "Derecho penal de la peligrosidad: tutela y garantía de los derechos fundamentales" (DER2017-86336-R, financiado por MINECO, AEI y FEDER-UE) y "Derecho penal y comportamiento humano" (RTI2018-097838-B-100, MICINN, AEI y FEDER-UE). Agradezco, como siempre, al profesor Francisco Montes Suay, catedrático de Estadística e Investigación Operativa de la Universidad de Valencia, su paciencia para revisar las partes más técnicas de este trabajo.
- (2) Se trataba de un *amicus brief* presentado en el caso Barefoot contra Estelle (1983); sobre las circunstancias de dicho caso, *cf.* Loinaz (2017: 65 y ss.) y Martínez (2014: 44 y ss.).
- (3) Caso Baxstrom contra Herold, 383 U.S. 107 (1966). Sobre este caso, *cf.* Steadman (1973).
- (4) Dixon contra Attorney General, 313 F.Supp. 653, Middle District of Pennsylvania (1970). Sobre el caso Dixon, *cf.* Kinzig (1996: 87 y ss.), así como Auerhahn (2003: 82), quien, a su vez, cita el trabajo original de Thornberry y Jacoby sobre el caso Dixon (*The Criminally Insane: A Community Follow-up of Mentally Ill Offenders*, 1979), trabajo este último que no he podido consultar directamente.
- (5) Aunque el concepto de peligrosidad se mantiene en muchos casos en la legislación vigente; *cf.* los artículos 6, 36.3, 83.3, 83.4, 90.5, 91, 92.3 y 97 (entre otros) del Código Penal español. Las leyes penales también utilizan otras expresiones, como *pronóstico de reinserción social* (p. ej., arts. 36.1 o 92.1 CP), en relación con las cuales se ha discutido si significan o no lo mismo que "peligrosidad" o incluso si en todos los casos en los cuales el Código Penal habla de "peligrosidad" lo hace en el mismo sentido (*cf.* Cervelló, 2014).
- (6) Conviene advertir que esta evolución se ha producido principalmente en el ámbito de la investigación criminológica; en la práctica forense es aún el método clínico el que continúan utilizando la mayor parte de los profesionales cuando emiten informes sobre la valoración de casos concretos, al menos en España (Arbach-Lucioni *et al.*, 2015: 359; Andrés-Pueyo y Echeburúa, 2010: 404 y ss.). También en EE. UU. parece que los tribunales siguen admitiendo sin problemas el juicio clínico como prueba sobre la peligrosidad, e incluso con mayor facilidad que periciales basadas en la aplicación de herramientas estructuradas de valoración del riesgo (Krauss y Scurich, 2013; Scurich, 2016).
- (7) La principal diferencia entre los instrumentos puramente actuariales y los de juicio clínico estructurado es que los primeros proporcionan una estimación de la peligrosidad automática, calculada con un algoritmo, a partir de la puntuación que el sujeto haya obtenido en los diferentes factores de riesgo que incluya el instrumento. Las herramientas de juicio clínico estructurado, por el contrario, son guías que indican cuáles son los factores que deben ser tenidos en cuenta en la valoración del nivel de riesgo y cómo deben ser apreciados, pero permiten que el profesional pondere su peso relativo con mayor libertad e incluso que añada otros factores que le parezcan decisivos en el caso concreto, aunque la herramienta no los contemple (*cf.*, sobre ello, con más detalles, Heilbrun, 2009: 3 y ss.; Loinaz, 2017: 201 y ss.).
- (8) *Cfr.* Andrés Pueyo y Redondo (2007: 169): "[...] las predicciones de riesgo de violencia, realizadas por procedimientos rigurosos como los descritos, son perfectamente comparables en su calidad y posibilidades a las de predicciones hechas en otras áreas profesionales reconocidas y útiles como la predicción meteorológica, las predicciones en ingeniería civil, las predicciones económicas o sociológicas".
- (9) "All models are approximations. Assumptions, whether implied or clearly stated, are never exactly true. All models are wrong, but some models are useful. So the question you need to ask is not "Is the model true?" (it never is) but "Is the model good enough for this particular application?" (Box, Luceño y Paniagua-Quiñones, 2009: 61). En un trabajo previo el autor

- había expresado la misma idea con otra matización: "Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad" (Box, 1976).
- (10) Sobre algunos de los distintos indicadores existentes y su significado *cfr.* Loinaz (2017).
 - (11) Sobre esta cuestión, que es excesivamente compleja para poder ofrecer aquí más detalles, *cfr.* ampliamente Martínez y Montes (2018).
 - (12) *Cfr.* Redondo, Pérez y Martínez (2007: 187-195), donde se ofrece una visión general de los resultados, y especialmente Pérez *et al.* (2008: 205-210), donde se ofrecen todos los datos detallados del estudio.
 - (13) Los autores afirman que se trata del "primer estudio empírico en España que investiga la eficacia predictiva de un instrumento de estimación de riesgo de violencia sexual, en concreto del *Sexual Violence Risk-20*" (Pérez *et al.*, 2008: 210).
 - (14) Esta afirmación implica una toma de postura previa sobre la valoración del colectivo al que correspondió un riesgo "moderado" de reincidencia. Quien aplica un instrumento de valoración del riesgo ha de decidir si considera los casos de riesgo moderado como supuestos de reincidencia probable o no probable, a efectos de juzgar después si el pronóstico se confirmó o no con la comisión de nuevos delitos (o la ausencia de ellos). En este caso, me limito a seguir la premisa asumida por los autores del estudio, que en la p. 151 asimilan los casos de riesgo moderado a los de riesgo alto, entendiendo que la posterior comisión de un delito violento confirma en ambos supuestos la predicción de riesgo de violencia. Si se calcula el porcentaje de sujetos que volvió a cometer un delito violento solo respecto del colectivo que había sido clasificado como de riesgo alto, el porcentaje asciende al 26,15 %.
 - (15) Para una crítica más detallada de este estudio, *cfr.* Martínez (2016).
 - (16) *Cfr.* Fazel *et al.* (2012: 10) y Douglas *et al.* (2017).
 - (17) ROC por *Receiver Operating Characteristics*; sobre su significado *cfr.* Martínez y Montes (2018).
 - (18) *Cfr.*, con ulteriores referencias, Martínez (2016) y Martínez y Montes (2018).
 - (19) Cosa que se comprueba fácilmente con solo atender a la evolución reciente de nuestro derecho penal: a los casos que siempre habían dependido de los pronósticos (decisiones en materia de ejecución penal y régimen de las medidas de seguridad para inimputables o semiimputables, básicamente) se han añadido varios nuevos supuestos: inscripción de identificadores de ADN del [art. 129 bis CP](#), medida de libertad vigilada para imputables o revisión de la pena de prisión permanente.
 - (20) *Cfr.*, por ejemplo, por todas, la [STC 160/2012, de 20 de septiembre](#).
 - (21) Lo ha advertido con claridad Harcourt (2007).
 - (22) Al decir esto soy muy consciente de que en la actualidad las preferencias de la ciudadanía y, desde luego, de las autoridades parecen ir en buena medida en la línea contraria: la de la "tolerancia cero" (contra cada vez más tipos de delincuencia: el terrorismo, su apología, los "delitos de odio", la violencia de género, los abusos sexuales, etc.). Pero en el texto estoy exponiendo no la tendencia que hoy día caracteriza a las campañas políticas, sino el marco que, a mi juicio, debería caracterizar a una política criminal ilustrada, inspirada especialmente en los valores de los [arts. 1.1, 10, 24 y 25 CE](#).
 - (23) *Rectius*, contra un cierto tipo de criminalidad. Porque relajar o eliminar las garantías normativas que limitan a los poderes públicos en la lucha contra la delincuencia favorece que el Estado abuse de sus poderes, lo que acaba desembocando en el aumento de otro

- tipo de criminalidad: detenciones ilegales, vulneraciones de la intimidad y del secreto de las comunicaciones, torturas, condenas sin pruebas, malos tratos, etc.
- (24) Siempre que dicho margen no sea excesivo, porque si lo fuera estarían en cuestión los principios limitadores del *ius puniendi* a que nos hemos referido *supra*.
- (25) A las distinciones esbozadas en el texto habría que añadir, aún, muchas más matizaciones: tampoco está sujeta a los mismos principios la decisión de concretar el tipo y la gravedad de la pena en la sentencia que la de tomar decisiones sobre la forma en que se continúa ejecutando la pena una vez impuesta, y aun en este último ámbito son muy diferentes las decisiones relativas a las formas de ejecución que no varían la situación del sujeto —por ejemplo, ofrecerle uno u otro tratamiento dentro de prisión— que las que suponen adelantar la puesta en libertad, etc. Con todo, por razones de espacio, no es posible profundizar aquí en dichos extremos.
- (26) Sobre todo ello *cfr.*, con ulteriores referencias, Martínez (2019).
- (27) *Cfr.* Martínez (2018).
- (28) *Cfr.* Tonry (2013).
- (29) Sobre la anticipación en general *cfr.*, p. ej., la intervención de González Cussac en esta misma publicación. Sobre la incriminación creciente de actos y delitos de preparación, *cfr.* Alonso (2017).
- (30) De "gobernar a través del delito" habla Simon (2012), quien subraya la centralidad del delito no solo para las clases sociales y colectivos que sufren especialmente la vigilancia policial y la dureza de las sanciones penales (colectivos caracterizados por la marginalidad socioeconómica), sino también para el resto de la población (clases medias y acomodadas), cuyas rutinas sociales en cualquier contexto (vivienda, trabajo, educación, etc.) están cada vez más condicionadas por el miedo al delito y la búsqueda de la seguridad.
- (31) Así lo advierten, por ejemplo, Monahan y Skeem (2016: 161-163), al afirmar que es una pérdida de tiempo dedicarse a evaluar necesidades criminógenas de los reos, que podrían utilizarse en estrategias de gestión del riesgo orientadas a su reducción, si el sistema no dispone de los medios para implementar los programas necesarios ni está dispuesto a instaurarlos. En estos casos, la reducción del riesgo es inútil, pero la predicción del riesgo sigue siendo viable y económica.

Información sobre el artículo

Título del artículo: "Derecho penal y riesgo: la prueba del riesgo de reincidencia"

Autor: Lucía Martínez Garay

Incluido en el número monográfico sobre *Lógica científica y valoración probatoria. Especial referencia a las pruebas biológicas* de Cuadernos Digitales de Formación 3 - 2020 (Directores: Salvador Camarena Grau y José Luis Ramírez Ortiz)

DOI:

Editor: Consejo General del Poder Judicial (Madrid)

Fecha de publicación: 2020

Copyright 2020, Consejo General del Poder Judicial

License:

Notas

39 referencias bibliográficas