

Errores conceptuales en la estimación de riesgo de reincidencia.

La importancia de diferenciar sensibilidad y valor predictivo, y
estimaciones de riesgo absolutas y relativas

Conceptual errors in violence risk assessments.

The importance of differentiating sensitivity and predictive value,
and absolute and relative risk estimates

Recibido: 19 noviembre 2015/Publicado: 16 septiembre 2016

Lucía Martínez Garay¹
Universidad de Valencia

RESUMEN

En un estudio recientemente publicado sobre *Tasa de reincidencia Penitenciaria 2014* en Cataluña se ofrecen, entre otra información de gran interés, los resultados de aplicar el RisCanvi como herramienta de valoración del riesgo de violencia a un conjunto de internos, y la comparación entre el riesgo de reincidencia violenta predicho y la reincidencia violenta efectivamente observada. En dicho estudio se valora muy positivamente la capacidad predictiva del RisCanvi; sin embargo, en este trabajo sostengo que debe hacerse una interpretación diferente de los datos que conduce a conclusiones menos optimistas. Ello se debe a que es necesario diferenciar con claridad entre sensibilidad y valor predictivo, y entre estimaciones de riesgo relativas y absolutas, lo que aconseja utilizar tanto el RisCanvi como otros instrumentos de valoración del riesgo con mucha prudencia en determinados contextos, especialmente

¹ La correspondencia debe dirigirse a: Lucía Martínez Garay. Departamento de Derecho Penal, Universidad de Valencia. Avda. Tarongers, s/n. 46022 Valencia (España). lucia.martinez@uv.es

cuando de estas valoraciones dependan decisiones sobre la libertad u otros derechos fundamentales de los internos.

Palabras clave: Valoración del riesgo de violencia, riesgo relativo y riesgo absoluto de reincidencia, valor predictivo, sensibilidad, AUC, RisCanvi.

ABSTRACT

A study recently published about prison recidivism rate in Catalonia 2014 offers, among other interesting information, the results of applying RisCanvi as a violence risk assessment tool to a sample of inmates, and the comparison between the risk of violent recidivism predicted and the violent recidivism actually observed. The study makes a very positive evaluation of the RisCanvi's predictive ability, but this paper maintains that data in that study should be interpreted differently, which leads to less optimistic conclusions. It is necessary to differentiate carefully between sensitivity and predictive value, and between relative and absolute predictive accuracy; and this in turn recommends a very cautious utilization of this and other predictive tools in specific contexts, especially when decisions about each inmate's liberty or fundamental rights depend upon the risk assessment.

Keywords: Violence risk assessment, relative and absolute risk of recidivism, predictive value, sensitivity, AUC, RisCanvi

1. Introducción

El interés por la valoración del riesgo de violencia, y en concreto por la valoración del riesgo de reincidencia en la comisión de delitos violentos, ha aumentado notablemente en las últimas décadas en muchos países. La regulación legal de numerosas instituciones jurídicas (libertad condicional, suspensión de la pena, aplicación de medidas de seguridad, etc.) exige al juez tomar decisiones sobre la libertad del sujeto en función de estimaciones sobre cómo dicho sujeto va a comportarse en el futuro, y especialmente sobre si existe riesgo de reiteración delictiva. La necesidad de dar respuesta a estas demandas legales ha favorecido que desde mediados de los años 80 se haya intensificado la investigación sobre los factores asociados a un mayor riesgo de violencia y se hayan multiplicado los instrumentos diseñados para evaluar dicho riesgo (especialmente los instrumentos actuariales y los de juicio clínico estructurado) al mismo tiempo que el tradicional concepto de peligrosidad iba siendo sustituido por el de valoración del riesgo de violencia (*violence risk assessment*) (Andrés-Pueyo y Redondo,

2007; Skeem y Monahan, 2011; Steadman, 2000; para la evolución de la valoración del riesgo en el ámbito psiquiátrico cfr. Buchanan, Binder, Norko y Swartz, 2012).

En España la recepción de esta corriente se ha acelerado en los últimos años y no sólo se han adaptado a nuestro idioma y contexto diversos instrumentos diseñados en otros países, sino que también se han creado herramientas nuevas (Andrés-Pueyo y Echeburúa, 2010). Entre estas últimas destaca el RisCanvi, que es una herramienta original de valoración del riesgo de conducta violenta desarrollada por el Grupo de Estudios Avanzados en Violencia de la Universidad de Barcelona y los Servicios Penitenciarios de la Generalitat de Cataluña para gestionar el riesgo de violencia en las prisiones catalanas. Según la información que proporciona la propia Generalitat de Cataluña, el RisCanvi no es únicamente un protocolo para la valoración del riesgo, sino que forma parte de un proyecto más amplio en el que la segunda vertiente es la gestión del riesgo: una vez evaluado el riesgo que presenta el sujeto hay que poner en marcha los programas o intervenciones más adecuados para influir en los factores de riesgo presentes en cada sujeto susceptibles de ser modificados o controlados, todo ello con el objetivo de mejorar la reinserción social de los penados². El RisCanvi se aplica desde 2010 a los internos en las prisiones catalanas, ya sea en su versión de *screening* o en la completa, lo que está permitiendo obtener gran cantidad de información sobre las características criminales, personales y biográficas, sociofamiliares, clínicas y de personalidad de los internos³.

Desde el punto de vista del Derecho penal y penitenciario – y con independencia de muchas otras cuestiones problemáticas que suscita el paradigma de la prevención en el que se inserta esta preocupación por la peligrosidad de los delincuentes (cfr., muy críticamente, Harcourt, 2007 o Albrecht, 2004) – una de las cuestiones que más interesa acerca de estos instrumentos de valoración del riesgo de violencia es la capacidad predictiva que tienen, esto es, con qué grado de acierto efectúan los pronósticos sobre la conducta futura de las personas. En otras palabras, qué probabilidad hay de equivocarse

² Cfr. la información que proporciona el Departament de Justícia de la Generalitat de Catalunya sobre el programa RisCanvi en el temario de las oposiciones para la Administración en el ámbito de ejecución de penas para el acceso a los cuerpos del grupo A, subgrupos A1 i A2, en concreto el Tema 13 (Resolució JUS/1119/2010, de 8 de abril, publicada en el Diari Oficial de la Generalitat de Catalunya nº 5609, de 16.04.2010, http://justicia.gencat.cat/ca/serveis/ofertes_de_treball/serveis_penitenciaris_i_rehabilitacio/convocatories_d_oposicions/temari_ambit_funcional_execucio_penal/)

³ Cfr. más información sobre este instrumento en Andrés-Pueyo (2013), Soler (2013).

cuando los operadores jurídicos toman decisiones basadas en estos pronósticos, ya sea por haber considerado de alto riesgo a un sujeto que en realidad luego no vuelve a delinquir (falsos positivos), o por haber considerado de riesgo bajo a una persona que sin embargo más tarde sí comete nuevos delitos violentos (falsos negativos). La preocupación por ambos tipos de errores ha acompañado desde siempre a cualquier clase de pronóstico que fuera necesario efectuar sobre la conducta criminal, preocupación basada, de un lado, en la alarma que generan siempre los falsos negativos cuando vuelven a delinquir (especialmente cuando se trata de un hecho violento), pero debida también, por otra parte, a la sospecha de que probablemente las predicciones de peligrosidad sobreestiman sistemáticamente el peligro de reincidencia. Esta sospecha se vio confirmada en los años 70 por los estudios de seguimiento que se realizaron sobre algunas cohortes de convictos liberados considerados peligrosos (los más conocidos fueron los casos Baxstrom y Dixon en Estados Unidos), que evidenciaron tasas muy altas de falsos positivos, lo que condujo a un generalizado escepticismo sobre la capacidad de acierto de los pronósticos clínicos de peligrosidad (Monahan, 1981). No obstante, la situación ha cambiado bastante desde entonces y en la actualidad el auge de los modernos instrumentos actuariales y de juicio clínico estructurado parece haber generado un moderado aunque firme optimismo sobre la capacidad predictiva de estos instrumentos (Buchanan et al., 2012; Singh y Petrila, 2013; Skeem y Monahan, 2011; Steadman, 2000). En palabras de Andrés-Pueyo y Redondo (2007, p. 169), “las predicciones de riesgo de violencia, realizadas por procedimientos rigurosos [...], 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”.

Con todo, al mismo tiempo diversos estudios demuestran que si bien la mayor parte de los modernos instrumentos de predicción ofrece valores elevados de sensibilidad y especificidad, así como unos niveles relativamente bajos de falsos negativos, su principal debilidad radica en el bajo valor predictivo positivo que alcanzan, es decir, en que sobreestiman la peligrosidad y generan por tanto muchos falsos positivos. Recientes meta-análisis que comparan los valores obtenidos en diferentes estudios de validación de los instrumentos de valoración del riesgo de violencia más usados internacionalmente así lo atestiguan, pues el valor predictivo

positivo suele estar bastante por debajo del 50% y además fluctúa enormemente de una muestra a otra (Fazel, Singh, Doll y Grann, 2012; Helmus, Hanson, Thornton, Babchishin y Harris, 2012; Singh, Fazel, Gueorguieva y Buchanan, 2014; Singh, Grann y Fazel, 2011). Trabajos publicados en los últimos años en Alemania en relación con la custodia de seguridad sobre supuestos reales de puesta en libertad de delincuentes tenidos por muy peligrosos arrojan tasas de falsos positivos de alrededor del 70% (Alex, 2013; Kinzig, 2010; Müller y Stolpmann, 2015; Rusche, 2004). Y los escasos datos de que disponemos en España sobre esta cuestión también parecen evidenciar una importante sobreestimación de la peligrosidad en los pronósticos (Martínez Garay, 2014).

En este contexto se ha publicado en 2015 el estudio sobre *Tasa de reincidencia penitenciaria 2014* de Cataluña, coordinado por Capdevila Capdevila (Capdevila Capdevila et al., 2015), que continúa la serie de datos ya iniciada años atrás sobre reincidencia penitenciaria en esta Comunidad Autónoma. El documento analiza la reincidencia penitenciaria de todos los liberados de las prisiones catalanas en 2010 (con seguimiento hasta el 31.12.2013) y ofrece muchos e interesantes datos sobre esta materia que además seguramente – y como apuntan sus autores, en la p. 9 – en buena medida podrían ser extrapolados al resto de España, habida cuenta además de que lamentablemente no existen muchos estudios sobre reincidencia en relación con todo el territorio del Estado. Aunque el objeto de este artículo no es en modo alguno analizar todo el estudio sino únicamente – como enseguida se verá – una parte muy específica del mismo, quiero aprovechar la ocasión para destacar la gran cantidad y calidad de la información que contiene. Es sabido que en España la política criminal de los últimos lustros se ha caracterizado por prescindir de cualquier análisis riguroso de la realidad criminológica y penitenciaria y perseguir únicamente finalidades simbólicas y de propaganda, cuando no se ha convertido directamente en arma arrojadiza en la contienda electoral (Díez Ripollés, 2012); además, los datos sobre criminalidad que ofrecen las instancias oficiales y especialmente el Ministerio del Interior reúnen tantas deficiencias que se ha llegado a afirmar que carecen de utilidad científica y que más que una imagen real de la delincuencia en nuestro país constituyen un instrumento de propaganda del Ministerio del Interior (Díez Ripollés et al., 2010). Un estudio como el de Capdevila Capdevila et al. (2015) constituye, insisto, un trabajo muy valioso que

pone a disposición del público mucha información de gran interés, y que ha de ser especialmente bienvenido en un contexto como el que venimos padeciendo desde hace años.

Pero volviendo al objeto de este trabajo, el estudio del que venimos hablando incluye respecto de los publicados en años anteriores la novedad de poner en relación la reincidencia penitenciaria efectivamente observada con la valoración del riesgo que se había hecho a un grupo de internos mediante el RisCanvi. Así, en el apartado 3.8 del estudio (Capdevila Capdevila et al., 2015, pp. 123 y ss.) se describe el perfil específico de la población evaluada con el RisCanvi completo. Se trata de un total de 410 internos que no han sido seleccionados aleatoriamente sino en función de que hubieran cometido delitos violentos o que hubieran manifestado durante su internamiento penitenciario “actitudes y comportamientos violentos o que potencialmente podrían serlo”; de modo que en este grupo se encuentran todos los sujetos que habían cometido delitos contra las personas, contra la libertad sexual, y contra la propiedad que fueran robos con violencia y/o intimidación (Capdevila Capdevila et al., 2015, pp. 124, 128). Por otro lado, en el apartado 4.3.1 (pp. 145 y ss.) se compara, dentro de la población evaluada con el RisCanvi, qué factores presentan diferencias significativas entre los sujetos que han reincidido y los que no (Gráfico 25 y Tabla 37), y por último en la Tabla 38 (p. 151) se compara la previsión de riesgo de reincidencia violenta que se hizo en su momento para estos sujetos con los datos empíricos sobre reincidencia penitenciaria violenta efectiva.

Los datos contenidos en dicha Tabla 38 son los que van a constituir el objeto de análisis crítico en este trabajo, porque la valoración que a partir de ellos hacen Capdevila Capdevila et al. (2015) sobre la capacidad predictiva del RisCanvi es muy positiva. Sin embargo, al presentar esos datos de la Tabla 38 Capdevila Capdevila et al. (2015) no diferencian suficientemente, en mi opinión, entre sensibilidad y valor predictivo, como a continuación trataré de demostrar. Y ello es relevante porque si se explicita el valor predictivo y se compara con los valores de sensibilidad y especificidad la capacidad predictiva del RisCanvi aparece como bastante más modesta, lo cual por otro lado concuerda con resultados de otros trabajos realizados a nivel internacional sobre diversas herramientas de predicción del riesgo. La valoración que a la luz de todo ello deba merecer la capacidad predictiva tanto del RisCanvi como de los demás instrumentos estructurados de estimación del riesgo de violencia depende a su vez, a mi

juicio, en buena medida de que se distinga con claridad entre estimaciones de riesgo absoluto y de riesgo relativo, distinción que debe tenerse muy presente cuando se quieran utilizar este tipo de herramientas en la toma de determinadas decisiones en el ámbito penitenciario y judicial.

2. Los datos sobre eficacia predictiva del RisCanvi en el estudio *Tasa de Reincidencia Penitenciaria 2014 en Cataluña*

A continuación se reproduce la Tabla 38 (aunque a efectos internos de este trabajo la numeramos como Tabla 1) con los datos totales y porcentuales sobre reincidentes y no reincidentes en función de la previsión de riesgo de violencia hecha con el RisCanvi tal y como aparece en el trabajo de Capdevila Capdevila et al. (2015, p. 151). Las celdas coloreadas aparecen también así en el estudio original:

Tabla 1

Comparativa entre la previsión de riesgo de reincidencia violenta (RisCanvi) y la reincidencia violenta efectiva

Predicción de riesgo de reincidencia violenta (RisCanvi)		Comisión de una reincidencia violenta (datos empíricos)					
		Sí		No		Total	
		N	%	N	%	N	%
Alto	% de fila	34	26.2	96	73.8	130	100.0
	% de columna		48.6		16.6		20.1
Moderado	% de fila	20	11.7	151	88.3	171	100.0
	% de columna		28.6		26.1		26.4
Bajo	% de fila	16	4.6	331	95.4	347	100.0
	% de columna		22.9		57.3		53.5
Total	% de fila	70	10.8	578	89.2	648⁴	100.0
	% de columna		100.0		100.0		100.0

Fuente: Reproducción literal de la Tabla 38 contenida en Capdevila Capdevila et al. (2015, p. 151).

⁴ El total de sujetos evaluados es de 648, mientras que en páginas anteriores del estudio los autores habían afirmado que el total de sujetos evaluados con la versión completa del RisCanvi era de 410 (p. 124). La diferencia seguramente se debe a que en esta tabla están incluidos no sólo los que fueron evaluados con la versión completa de la herramienta, sino también los que fueron evaluados con el RisCanvi *screening*, que según la información que proporciona el propio estudio en la Tabla 20 de la p. 125 fueron 245. 410 más 245 son 655; la diferencia de 7 personas hasta 648 quizá se deba a algún ajuste menor, que si se indica en el estudio no he sabido localizar.

La variable “reincidencia violenta” está descrita en el trabajo como “Tiene alguna reincidencia con delito violento (todas, no solo la primera)” (Capdevila Capdevila et al., 2015, p. 89), aunque en relación con la tabla que ahora estamos comentando el estudio advierte expresamente que se va a considerar reincidencia violenta haber tenido alguna, haya sido la primera reincidencia o no. Lo que no se describe como variable específica en el estudio es lo que se ha de entender como “delito violento”, por lo que cabe suponer que se ha considerado tal lo mismo que en el estudio constituyó el presupuesto para evaluar a los internos con el RisCanvi: haber cometido delitos contra las personas, contra la libertad sexual, y contra la propiedad que fueran robos con violencia y/o intimidación (Capdevila Capdevila et al., 2015, p. 128).

Por otro lado, y según se indica en el estudio, la muestra está constituida por las personas que terminaron su condena de prisión en 2010 y a las que se ha realizado un seguimiento hasta el 31 de diciembre de 2013 con el fin de registrar si se han producido reingresos en prisión. El periodo de seguimiento para determinar la reincidencia penitenciaria es por tanto de 3,5 años como media, con un mínimo de 3 años y un máximo de 4 años de seguimiento (Capdevila Capdevila et al., 2015, p. 77)

2.1. La interpretación de los datos en el estudio

La interpretación que Capdevila Capdevila et al. (2015) efectúan de estos datos es, literalmente, la siguiente (2015):

Aparecen destacados en verde los casos en los que se había previsto un riesgo de reincidencia violenta (alto o medio) y que efectivamente han cometido una reincidencia violenta. El acierto en el pronóstico es del 77,2%.

También están en verde los casos en los que los profesionales habían previsto un riesgo bajo de reincidencia violenta y efectivamente no la ha habido. Suponen el 95,4%.

Por otra parte, se han remarcado con color naranja los falsos negativos: es el caso de los internos para los que no se había identificado un riesgo de reincidencia violenta y que, en cambio, han acabado cometiendo al menos una. Representan el 4,6% del conjunto de internos con una evaluación de riesgo bajo.

Finalmente, en color rojo destacan los falsos positivos: es decir, el caso de los internos para los que se había previsto un riesgo alto o moderado de reincidencia violenta, pero que realmente no han reincidido. Representan el 42,7% de los no reincidentes. Por lo tanto, la probabilidad de acierto en la estimación hecha (medida OR) con la herramienta RisCanvi es de 4,52 (sumando alto + moderado). En cuanto a la sensibilidad de la herramienta –es decir, la capacidad para prever como casos de riesgo alto o moderado sujetos que efectivamente reincidirán–, es

del 77,15%. Finalmente, la especificidad o capacidad de la herramienta para dar como casos de bajo riesgo sujetos que efectivamente no reincidirán es del 57,26%. (pp. 151-152).

En las conclusiones del estudio se valoran de forma muy positiva estos datos: “La herramienta se ha comportado de modo excelente en la predicción de riesgo sobre los sujetos que efectivamente reincidirán (77,15%) y aceptable para clasificar como casos de bajo riesgo sujetos que efectivamente no reincidirán (57,26%)” (Capdevila Capdevila et al., 2015, p. 327).

2.2. Crítica

Como puede observarse en el texto que acabamos de transcribir, los autores del citado estudio valoran muy positivamente los resultados de la aplicación del RisCanvi porque “el acierto en el pronóstico” (p. 151) o la “predicción de riesgo” (p. 237) ha alcanzado un valor del 77.2%⁵. Sin embargo, estas expresiones son ambiguas, pues con ellas nos podemos estar refiriendo a dos cosas distintas: por un lado, al porcentaje de casos correctamente identificados como reincidentes por el RisCanvi de entre el total de sujetos que finalmente reincidieron; pero también, por otro lado, al porcentaje de casos que finalmente reincidieron de entre el total de sujetos a los que el RisCanvi había asignado previamente un riesgo alto o moderado de reincidir. Lo primero se conoce como sensibilidad de un instrumento de valoración del riesgo; lo segundo, como valor predictivo o valor pronóstico. Conviene insistir en la diferencia entre ambas categorías: la sensibilidad es el porcentaje de casos correctamente detectados por el instrumento en relación con los casos que reinciden; el valor predictivo positivo, por el contrario, es el porcentaje de casos reincidentes de entre los detectados como tales por la prueba. Expresado de forma un poco más coloquial, la sensibilidad diría: sabiendo cuántos han delinquido, vamos a ver cuántos de esos habíamos sido capaces de identificar con nuestra herramienta de predicción. El valor predictivo, por el contrario, diría: sabiendo cuántos dijimos que serían peligrosos, vamos a ver cuántos de ellos realmente han delinquido después.

⁵ En la p. 151 se afirma que es el 77.2% y en las páginas 152 y 237 que es el 77.15%, pero se trata en ambos casos de la misma cifra, redondeada a un solo decimal o a dos.

Tanto la sensibilidad como el valor predictivo son probabilidades condicionadas. En Probabilidad se habla de probabilidad condicionada del suceso A (ser reincidente) a otro suceso B (haber obtenido una puntuación indicativa de riesgo alto en un instrumento de estimación del riesgo de reincidencia) y se la representa por $P(A|B)$, que significa la probabilidad de que ocurra el suceso A sabiendo que ha ocurrido el suceso B. En un estudio sobre eficacia predictiva de una herramienta de estimación del riesgo de reincidencia la sensibilidad nos daría la probabilidad $P(B|A)$, mientras que el valor predictivo nos daría la probabilidad $P(A|B)$. Es evidente que no es lo mismo $P(A|B)$ que $P(B|A)$, pero ocurre sin embargo con demasiada frecuencia que ambas probabilidades se confunden, confusión que se conoce en estadística forense como "falacia de la condicional transpuesta" o "falacia del fiscal" (Aitken y Taroni, 2004). La tesis que sostengo en este trabajo es que, sin haber caído directamente en la falacia del fiscal, la forma en que se presentan los datos de eficacia predictiva del RisCanvi en el estudio sobre tasa de reincidencia penitenciaria en Cataluña de Capdevila Capdevila et al. (2015) propicia que los lectores puedan incurrir, involuntariamente, en ella. Veamos por qué.

Cuando Capdevila Capdevila et al. (2015) en su estudio afirman que "la predicción de riesgo" efectuada con el RisCanvi es excelente parece en principio que aluden a la sensibilidad, pues en la p. 152 utilizan expresamente este término, al afirmar que "la sensibilidad de la herramienta –es decir, la capacidad para prever como casos de riesgo alto o moderado sujetos que efectivamente reincidirán–, es del 77.15%". Esta cifra indica el porcentaje que representan, entre las 70 personas reincidentes, las 34 y 20 a las que se predijo, respectivamente, un riesgo alto y moderado de reincidir – $P(B|A)$ –.

El término 'valor predictivo', por el contrario, no aparece expresamente mencionado en el trabajo de Capdevila Capdevila et al. (2015), lo que podría hacernos pensar que no se ha tenido en cuenta al valorar la bondad del RisCanvi como herramienta de predicción del riesgo. Sin embargo, lo cierto es que el valor predictivo sí se ha calculado, aunque los autores no utilicen expresamente esta denominación, y hay una referencia al mismo en la p. 151, cuando se afirma que "También están en verde los casos en los que los profesionales habían previsto un riesgo bajo de reincidencia violenta y efectivamente no la ha habido. Suponen el 95.4%." Este porcentaje es lo que suponen 331 personas que no volvieron a cometer un delito violento respecto de las 347

a las que se había considerado de riesgo bajo, y constituye el valor predictivo negativo – $P(\text{noA}|\text{noB})$ –.

Pero la presentación conjunta de ambos porcentajes (77.2% y 95.4%), que aparecen sin solución de continuidad en líneas sucesivas en la misma página del trabajo, unidas además por el adverbio “también”, bajo la denominación genérica de “acierto en el pronóstico” y sin advertir que en un caso se está hablando de la sensibilidad y en el otro del valor predictivo negativo es, en mi opinión, equívoca, porque genera la impresión de que se habla en ambos casos de valores predictivos, cuando no es así.

Además de la sensibilidad y del valor predictivo negativo (aunque en este caso ya hemos visto que sin denominarlo expresamente así) los autores del estudio que venimos comentando ofrecen también el dato de la especificidad, con buen criterio a mi juicio pues es un valor que complementa la información proporcionada por la sensibilidad, ya que si ésta última indica el porcentaje de sujetos clasificados como de alto riesgo entre los que efectivamente reincidieron ($P(B|A)$), la especificidad indica el porcentaje de sujetos clasificados como de bajo riesgo de entre el total de los que no reincidieron ($P(\text{noB}|\text{noA})$). Dicho porcentaje, que está correctamente calculado en el trabajo, es del 57.26% (Capdevila Capdevila et al., 2015, p. 152), porque del total de 578 sujetos que no reincidieron en el periodo estudiado, 331 habían sido previamente valorados como de riesgo bajo.

Sin embargo, en el estudio no se opera de la misma manera en relación con el valor pronóstico, pues a pesar de que se ha calculado el valor predictivo negativo (aun sin denominarlo así) no se dice cuál es el valor predictivo positivo. Y lo cierto es que esta información se echa en falta, ya que es una magnitud de importancia esencial habida cuenta del tipo de riesgo que se estima con el RisCanvi y el contexto – penitenciario y forense – en el que se aplica. En efecto, si el objetivo del RisCanvi es “proporciona[r] una predicción del riesgo de violencia futura” (Capdevila Capdevila et. al., 2015, p. 124), y si la información que proporciona se utiliza como criterio para asignar a cada individuo un nivel de supervisión acorde a su nivel de riesgo (Andrés Pueyo, 2013, p. 69), entonces una pregunta muy pertinente para poder valorar su funcionamiento y utilidad es saber qué porcentaje de los sujetos que el RisCanvi calificó como de riesgo alto o moderado efectivamente tuvieron después una reincidencia violenta ($P(A|B)$). A partir de los datos de la Tabla 1 es sencillo obtener su valor: de los

301 sujetos pronosticados como probables reincidentes (riesgo alto, 130 + riesgo moderado, 171)⁶, sólo 54 reincidentieron, es decir, un escaso 17.94%.

Todas estas cifras quizá se entiendan mejor si se añaden a la tabla que hemos estado comentando, que quedaría así:

Tabla 2

Comparativa entre la previsión de riesgo de reincidencia violenta (RisCanvi) y la reincidencia violenta efectiva

Predicción de riesgo de reincidencia violenta (RisCanvi)	Comisión de una reincidencia violenta (datos empíricos)			
	Sí	No	Total	
Alto	34	96	130	Valor predictivo positivo: 17.94%
Moderado	20	151	171	
Bajo	16	331	347	Valor predictivo negativo: 95.4%
Total	70	578	648	
	Sensibilidad: 77.15%	Especificidad: 57.26%		

Fuente: Elaborada a partir de la Tabla 38 contenida en Capdevila Capdevila et al. (2015, p. 151), añadiendo la última fila y la última columna y variando los colores originales de las celdas. En rojo están los falsos negativos (16) y los falsos positivos (247).

Obsérvese que de todos los porcentajes calculados el valor predictivo positivo es precisamente el que alcanza una cifra no sólo más modesta, sino además llamativamente alejada de las magnitudes que alcanzan la sensibilidad y el valor predictivo negativo, y sin embargo es el único parámetro del que no se informa al lector. Esta omisión puede contribuir decisivamente, en mi opinión, a que el lector malinterprete lo que los autores presentan como "acierto en el pronóstico", y entienda que se trata de $P(A|B)$ en lugar de $P(B|A)$. Riesgo de mala interpretación que me atrevería a decir que crece si el lector es un jurista o un operador jurídico, pues este tipo de público, generalmente mucho menos

⁶ 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.

formado en estadística de lo que lo está un psicólogo o un criminólogo, tiende a pensar que el grado de "acierto en el pronóstico" o la capacidad de "predicción de la reincidencia violenta" es lo que en el lenguaje cotidiano se entiende por tal: el porcentaje de casos en que la realidad confirma el pronóstico, o el porcentaje de sujetos que efectivamente reinciden de entre aquéllos a los que se les ha pronosticado un riesgo elevado de hacerlo. Es decir, el valor predictivo, y no la sensibilidad.

En definitiva, cuando en el estudio *Tasa de reincidencia penitenciaria 2014* sus autores concluyen que “La herramienta se ha comportado de modo excelente en la predicción de riesgo sobre los sujetos que efectivamente reincidirán (77,15%) y aceptable para clasificar como casos de bajo riesgo sujetos que efectivamente no reincidirán (57,26%)” (Capdevila Capdevila et al., 2015, p. 327) están exponiendo los porcentajes de sensibilidad y especificidad, respectivamente, que se obtuvieron en la muestra a la que se aplicó el RisCanvi, porcentajes que están correctamente calculados. No obstante, modestamente opino que la forma en que en este estudio se presentan e interpretan los resultados tiene un elevado riesgo de inducir a confusión al lector. Aunque son correctos en términos matemáticos los parámetros que se calculan⁷, generan una impresión excesivamente optimista sobre la capacidad predictiva del RisCanvi por dos razones. En primer lugar, porque informan sobre la sensibilidad del instrumento (77.2%) pero no sobre el valor predictivo positivo del mismo, parámetro muy relevante en este contexto y que sin embargo arroja un resultado muy inferior, y más desalentador (17.94%). Y en segundo lugar porque en lugar de utilizar la terminología más inequívoca de sensibilidad/especificidad y valor predictivo se utiliza una expresión ambigua, “acierto en el pronóstico”, para referirse aparentemente tanto a la sensibilidad como al valor predictivo, lo que favorece que el lector pueda tomar por el valor predictivo el porcentaje que se le ofrece (y que en realidad es la sensibilidad).

3. Discusión

3.1. El valor predictivo del RisCanvi en el contexto internacional

⁷ A excepción del porcentaje de falsos negativos: estos, que representan el porcentaje de personas a las que el instrumento había asignado un riesgo bajo de reincidir (16) de entre el total de las que finalmente reincidieron (70), no son el 4.6% como se afirma en la p. 151 de Capdevila Capdevila et al. (2015), sino el 22.85%.

Las matizaciones que en el apartado anterior he efectuado a la interpretación que hacen los autores del estudio *Tasa de reincidencia penitenciaria 2014* sobre la capacidad predictiva del RisCanvi en absoluto significan por sí solas que éste sea un mal instrumento de valoración del riesgo de reincidencia. Por el contrario, los valores de sensibilidad, valor predictivo positivo y valor predictivo negativo que obtuvo el RisCanvi en este caso son similares a los que evidencian otros instrumentos de predicción de la delincuencia violenta muy conocidos y utilizados internacionalmente. Para comprobarlo basta acudir a algunos meta-análisis publicados en los últimos años que ofrecen los valores medios o medianas de estos parámetros que se han constatado en la administración de diversos instrumentos actuariales de valoración del riesgo de violencia a numerosas muestras de sujetos, como por ejemplo el de Fazel et al. (2012):

Tabla 3

Sensibilidad, especificidad y valor predictivo de diversos instrumentos actuariales de predicción en 68 estudios de validación sobre 73 muestras que abarcan 24.827 sujetos

	Delincuencia violenta (HCR-20; SARA, SAVRY y VRAG)	Delincuencia sexual (SORAG, Static-99 y SVR-20)	Delincuencia en general (LSI-R y PCL-R)
Sensibilidad (IC 95%)	0.92 (0.88 – 0.94)	0.88 (0.83 – 0.92)	0.41 (0.28 – 0.56)
Especificidad (IC 95%)	0.36 (0.28 – 0.44)	0.34 (0.20 – 0.51)	0.80 (0.67 – 0.8)
Valor predictivo positivo (mediana - intervalo intercuartílico-)	0.41 (0.27 – 0.60)	0.23 (0.09 – 0.41)	0.52 (0.32 – 0.59)
Valor predictivo negativo (mediana - intervalo intercuartílico-)	0.91 (0.81 – 0.95)	0.93 (0.82 – 0.98)	0.76 (0.61 – 0.84)

Fuente: Fazel et al. (2012) (valores extraídos de la tabla contenida en la p. 10).

Como puede observarse, los instrumentos analizados en este estudio ofrecen en general tasas altas de sensibilidad: al menos los diseñados para predecir el riesgo de delincuencia violenta y delincuencia sexual consiguen detectar correctamente a alrededor del 90% de los delincuentes que efectivamente llevarán a cabo estas conductas. Pero estos valores elevados de sensibilidad contrastan con otros mucho más

bajos en cuanto al valor predictivo, tal y como ocurre también en los resultados del RisCanvi que estamos analizando en este trabajo. Obsérvese cómo en los casos en que se aplicaron los instrumentos actuariales de predicción que recoge la Tabla 3 el valor predictivo positivo es en casi todos ellos inferior al 50%, llegando incluso, en los instrumentos de valoración del riesgo de reincidencia sexual, a sólo el 23%. Esto es, utilizando dichos instrumentos, del total de sujetos que se predijo que sí cometerían delitos sexuales, el 77% no lo hizo. Únicamente en los instrumentos de valoración del riesgo que miden la probabilidad de comisión de delitos en general (en este estudio, la PCL-R y el LSI-R) el valor es superior al 50%, pero únicamente por dos décimas⁸.

Por otro lado, es importante también tener en cuenta que la Tabla 3 ofrece valores medios o medianas sometidos a márgenes de error o dispersión que constan entre paréntesis. Es decir, en realidad los distintos estudios de predicción hechos con los instrumentos de valoración del riesgo que ahí se analizan han obtenido distintos valores predictivos positivos (y negativos, pero nos interesan sobre todo los primeros, que son los que han acertado en la predicción de la existencia de peligrosidad), la mitad de los cuales oscilaban entre las dos cifras indicadas en el paréntesis, y siendo el número que precede al paréntesis la mediana de dichos datos. En consecuencia, incluso en los casos en que la mediana del valor predictivo positivo está cercana al 50% (tercera columna) no hay que perder de vista que con ese mismo instrumento de predicción se han obtenido en algunos de los estudios de validación revisados valores inferiores o superiores, la mayor parte de los cuales se sitúan entre el mínimo y el máximo indicados en el paréntesis, cuyo límite inferior está en el 32%. Y eso en el mejor de los casos, porque en lo que se refiere a los instrumentos de predicción del riesgo de delincuencia sexual (en la segunda columna) el límite inferior del intervalo intercuartílico para la mediana del valor predictivo positivo se sitúa en el 9%. En definitiva, lo que estos datos nos están diciendo es que en la inmensa mayoría de las muestras en las que se estimó el riesgo de violencia o de reincidencia utilizando los instrumentos estudiados en este meta-análisis, las predicciones se equivocaron más veces de las que acertaron, y además

⁸ Y además no sería a estos a los que cabría asimilar los resultados arrojados por la aplicación del RisCanvi a la muestra analizada en el estudio de Capdevila Capdevila et al. (2015), porque en ese trabajo el RisCanvi se aplicó a sujetos condenados por la comisión de delitos violentos, y con el objetivo de medir precisamente la reincidencia violenta, por lo que los valores más adecuados para comparar estos resultados del RisCanvi son los de las dos primeras columnas de la Tabla 3.

que las tasas reales de reincidencia (o de reincidencia violenta) que evidenció cada muestra fueron muy diferentes entre sí.

Frente a estos valores predictivos positivos tan bajos llama la atención que los valores predictivos negativos son mucho más elevados: del orden del 90% en los instrumentos de valoración del riesgo de delincuencia violenta y de delincuencia sexual, y además con unos límites mínimos y máximos en el intervalo intercuartílico que tampoco se alejan mucho de esas cifras. Esto significa que a la hora de predecir quienes *no* volverán a cometer delitos los instrumentos de valoración del riesgo aciertan mucho más, al igual que ocurre con el RisCanvi en el estudio de Capdevila Capdevila et al. (2015), donde el valor predictivo negativo fue del 95.4%.

En definitiva, si por “acierto en el pronóstico” entendemos el valor predictivo, el evidenciado por el Riscanvi en la muestra de sujetos a la que se evaluó en el estudio que estamos comentando de Capdevila Capdevila et al. (2015) está en la línea de muchos otros instrumentos de valoración del riesgo de violencia, muy conocidos y consolidados internacionalmente⁹, aunque alcance solo un modesto 17.94%.

Pero entonces, ¿cómo se explica el optimismo que según hemos dicho al principio se ha extendido en la actualidad respecto a la capacidad predictiva de este tipo de herramientas? ¿Cómo puede afirmarse, por un lado, que los modernos instrumentos estructurados de valoración del riesgo proporcionan estimaciones mejores que el azar y al menos moderadamente satisfactorias, si al mismo tiempo los valores predictivos positivos no sólo son enormemente volátiles sino además casi siempre inferiores al 50%? Si no me equivoco, esta aparente contradicción está al menos en parte relacionada con la forma de medir lo que denominamos “precisión” de las predicciones, pues según cómo se defina este parámetro una misma predicción podrá considerarse mejor o peor (Mossman, 1994). El próximo apartado aborda muy sucintamente esta cuestión¹⁰.

3.2. Estimaciones de riesgo relativo y de riesgo absoluto

⁹ Otro meta-análisis que revisa una gran cantidad de estudios y que ofrece valores similares a los aquí resumidos en la Tabla 3 es el de Singh et al., 2011 (especialmente los de la Tabla 4 en la p. 507).

¹⁰ Que constituye evidentemente sólo un problema de los muchos que plantea la cuestión de la capacidad predictiva de los instrumentos estructurados de valoración del riesgo. Sobre este amplio y complejo tema puede verse una panorámica del estado actual de la cuestión, así como ulteriores referencias, en Singh y Pettila (2013) y en los diferentes trabajos reunidos en el nº 31 de la revista *Behavioral Sciences and the Law*, un *special issue* dedicado monográficamente a esta cuestión.

La forma tradicional de medir el grado de acierto en las predicciones de peligrosidad era calcular el porcentaje de casos en los que la comisión de un nuevo acto violento confirmaba la peligrosidad del sujeto, respecto del total de personas a las que se había juzgado como peligrosas inicialmente. Esto fue lo que midieron los estudios de seguimiento hechos sobre los casos Baxstrom y Dixon y ese porcentaje resultó muy bajo, de ahí el escepticismo que se generalizó en los años setenta y ochenta del siglo pasado sobre la capacidad de los peritos clínicos para hacer predicciones de peligrosidad. Y esto es lo que no ha cambiado sustancialmente hasta hoy: si atendemos a los resultados de los meta-análisis citados en el epígrafe anterior, que miden igualmente el valor predictivo (Fazel et al., 2012; Singh et al., 2011), así como a los trabajos publicados en los últimos años sobre tasas reales de reincidencia observadas en delincuentes tenidos por muy peligrosos puestos en libertad (Alex, 2013; Kinzig, 2010; Müller y Stolpmann, 2015; Rusche, 2004), la evidencia empírica sigue demostrando que los porcentajes de acierto en la predicción de la violencia – especialmente cuando se trata de los sujetos a los que se asigna un riesgo alto de reincidencia violenta – siguen siendo muy modestos, y la proporción de falsos positivos muy elevada, hasta el punto de que no puede considerarse que la situación haya mejorado sustancialmente respecto de la que evidenciaron los casos Baxstrom y Dixon, y así lo reconocen expresamente algunos autores, por ejemplo Müller y Stolpmann (2015, p. 43).

Pero con el desarrollo de los instrumentos estructurados de valoración del riesgo se han ido generalizando otras formas de medir su capacidad predictiva, métodos estadísticos que no dependen de los números totales de falsos positivos y falsos negativos que resultan de una predicción dada, sino que cuantifican la capacidad de cada herramienta para discriminar, dentro de la muestra, entre sujetos reincidentes y no reincidentes. Seguramente el más conocido en la actualidad es la AUC o área bajo la curva ROC.¹¹ Una curva ROC (acrónimo de *Receiver Operating Characteristic*, o Característica Operativa del Receptor) es la representación gráfica, en un sistema de coordenadas, de la sensibilidad en el eje de abscisas frente a (“1–especificidad”) en el eje de ordenadas para un sistema clasificador binario según varía el umbral de

¹¹ Otro estadístico también utilizado como medida de la capacidad de predicción del riesgo relativo es la *d* de Cohen: “*d* mide la diferencia promedio entre los reincidentes y los no reincidentes y compara esta diferencia con el grado en que los reincidentes se diferencian entre sí y con el grado en que los no reincidentes se diferencian entre sí” (Hanson y Morton-Bourgon, 2009, p. 5).

discriminación. Es decir, dibuja la curva que forman los puntos correspondientes a las diversas combinaciones de valores de sensibilidad y de “1–especificidad” que se obtienen según vayamos situando el punto de corte entre riesgo alto y bajo en cada uno de los distintos niveles de riesgo posibles que contiene un instrumento de predicción. También puede describirse afirmando que es la representación de la ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la ratio de falsos positivos (FPR = Razón de Falsos Positivos) en función de dónde se sitúe el umbral de discriminación. El área que queda bajo esta curva (AUC por sus siglas en inglés: *area under the curve*) varía entre 0 y 1: un valor de 1 para la AUC significa que el instrumento clasifica perfectamente todos los casos (todos los futuros reincidentes y sólo ellos han sido considerados peligrosos, y todos los futuros no reincidentes y sólo ellos han sido considerados no peligrosos), mientras que una AUC por debajo de 0.5 indica una clasificación que no es mejor que el azar.

La AUC ha sido muy utilizada desde los años 90 para medir la calidad de los diversos instrumentos de valoración del riesgo de violencia (por ej. Heilbrun, 2009), y los numerosos estudios publicados al respecto aseguran que alcanza valores altos, lo que indicaría que la capacidad predictiva relativa de la mayoría de los instrumentos más utilizados es buena, porque es superior al azar (*better than chance*) (por ejemplo, Barnett et al., 2010; Hanson y Morton-Bourgon, 2009; Mossman 1994; Urbaniok, Noll, Grunewald, Steinbach y Endrass 2006). Sin embargo, para entender correctamente lo que estas valoraciones positivas significan hay que tener muy presente cuál es la información que proporciona la AUC. La AUC y el resto de parámetros que miden el riesgo *relativo* de reincidencia reflejan la capacidad de un instrumento para clasificar a los sujetos en función de que su riesgo de reincidencia sea mayor o menor *que el de los demás sujetos de la muestra* a los que se ha evaluado, pero no dicen nada sobre cuál es la probabilidad de reincidencia asociada a cada nivel de riesgo en el que se clasifica a los sujetos. Por ejemplo, una AUC de 0.75, que generalmente se considera en este tipo de estudios como satisfactoria en cuanto medida de riesgo relativo¹², lo que nos dice es que si escogemos al azar un sujeto que efectivamente ha reincidido, hay un 75% de

¹² Con todo, es discutible cuáles son los valores de la AUC que deberían considerarse satisfactorios, de la misma manera que también es discutible que la AUC ‘en bruto’ sea por sí misma una medida adecuada y útil según los propósitos que se persigan con la valoración del riesgo (mera predicción; predicción más tratamiento; predicción diferenciada no sólo de reincidencia en general sino de tipos, gravedad, frecuencia, etc, de la reincidencia), cfr. sobre ello Sjöstedt y Grann (2002).

probabilidades de que haya obtenido con ese instrumento una puntuación indicativa de mayor riesgo que un individuo no reincidente también escogido al azar. Por eso se denomina a estos parámetros medidas de riesgo *relativo*, porque lo que indican es la capacidad del instrumento para distribuir a los sujetos evaluados en categorías de mayor o menor riesgo de reincidencia unas respecto de las otras; o dicho con otras palabras, el que un individuo obtenga una puntuación indicativa de riesgo alto nos está diciendo que tiene mayor riesgo de reincidir que los demás sujetos evaluados con ese instrumento y que con él obtuvieron puntuaciones inferiores (Mossman, 2006).

El tipo de información que proporcionan estas medidas sobre la precisión con que se estima el riesgo relativo puede ser, desde luego, muy útil a ciertos efectos: por ejemplo, la curva ROC cuya área es la AUC permite calcular el punto de corte que en un instrumento determinado consigue producir al mismo tiempo el menor número de falsos positivos y también el menor número de falsos negativos. Además, la AUC tiene la ventaja de que es un parámetro independiente de las tasas de prevalencia, por lo que permite efectuar comparaciones entre el funcionamiento de diversos instrumentos sin que la prevalencia del fenómeno que se mide oscurezca los resultados, a diferencia de lo que ocurre cuando se comparan directamente los valores predictivos positivos, que están muy influenciados por las tasas de prevalencia (Mossman, 1994).

Pero es importante recordar que la AUC no dice nada sobre cuál es la probabilidad de riesgo *absoluto*, es decir, la probabilidad de reincidencia asociada a cada nivel de riesgo en el que la herramienta de valoración clasifica a los sujetos. Una AUC de 0.75 de ninguna manera significa que un sujeto incluido en la categoría de riesgo alto tenga un 75% de probabilidades de volver a delinquir, ni tampoco una AUC de 0.75 indica que ésa sea la fracción de predicciones correctas que se consiguen con dicho instrumento de valoración. Es más, al ser la AUC una propiedad del instrumento que depende de los porcentajes de sensibilidad y especificidad, y que es independiente del número total de errores que se generan al elegir el punto en el que situar la frontera entre riesgo alto y bajo, un instrumento con una AUC elevada puede generar predicciones con números muy altos de falsos positivos y valores predictivos positivos muy bajos. El ejemplo imaginario que utiliza Mossman en su trabajo de 1994 es muy ilustrativo: supongamos que disponemos de una herramienta de valoración que clasifica a los individuos en 5 niveles de riesgo, siendo el 1 el de riesgo más bajo y el 5 el de

riesgo más alto. Aplicado a una muestra de 1.000 sujetos (de los que el 10% actuó violentamente), los números de errores en la predicción serán distintos en función de que coloquemos el punto de corte entre riesgo alto y bajo en el nivel 1, 2, 3, 4 o 5; no sólo habrá menos falsos positivos y más falsos negativos si el punto de corte es el 5, y por el contrario muchos falsos positivos y menos falsos negativos si el punto de corte lo situamos en el 1, sino que el número total de errores (sumando falsos positivos y falsos negativos) también irá variando. Así, si consideramos de alto riesgo sólo a los sujetos que alcanzan el nivel cinco y los internamos, tendremos un número total de errores de 110 (50 falsos negativos y 60 falsos positivos), mientras que si consideramos peligrosos a todos los sujetos a partir del nivel 1 y los internamos tendremos 457 errores en total (7 falsos negativos y 450 falsos positivos). Sin embargo, para todas las predicciones hechas con ese instrumento, tanto la que produjo 457 errores como la que generó sólo 110, la AUC es la misma: 0.856 (Mossman, 1994, Tabla 2 en la p. 785 y Figura 1 en la p. 786).

Frente a los numerosos estudios de validación de los instrumentos de predicción del riesgo de violencia en los que se ha contrastado su buena capacidad de predicción del riesgo relativo, son mucho menos abundantes los estudios que analizan la estabilidad en la estimación de las probabilidades absolutas de reincidencia asociadas a cada nivel de riesgo (Helmus et al., 2012). Es decir, los que tratan de medir la precisión con que estas herramientas estiman las probabilidades absolutas de reincidencia, comprobando si la probabilidad de reincidencia que evidenciaron en el estudio inicial los sujetos clasificados en cada categoría de riesgo efectivamente se mantiene en sucesivos estudios de validación con grupos de sujetos distintos a los de la muestra inicial. Algunos meta-análisis publicados en los últimos años han comprobado que las tasas reales de violencia evidenciadas por sujetos clasificados como de alto riesgo con un mismo instrumento varían enormemente de unas muestras a otras (Helmus et al., 2012; Singh et al., 2014). Esto significa que el hecho de que en un estudio los sujetos clasificados como de alto riesgo según, por ejemplo, el Static-99R, hayan tenido una tasa de reincidencia del 35% no nos permite inferir sin más que si aplicamos el mismo instrumento a una nueva muestra de sujetos la tasa de violencia que evidenciarán los que clasifiquemos como de alto riesgo vaya a ser igual, porque los estudios realizados demuestran que unas veces es bastante mayor, y otras bastante menor.

Estos resultados corroboran un aspecto de los estudios sobre el valor predictivo de los instrumentos de valoración del riesgo a los que nos hemos referido en el epígrafe anterior (3.1): la gran variabilidad en los valores predictivos positivos que allí se constataba es lo que en estos últimos se analiza como variabilidad en las tasas absolutas de reincidencia estimadas. En otras palabras, tanto unos como otros confirman que muchas veces las personas clasificadas en una categoría de riesgo alto según una determinada herramienta de predicción no muestran en realidad después el grado de comportamiento violento que según dicha herramienta era esperable. De manera que, en definitiva, la evidencia empírica hasta ahora disponible permite afirmar que si bien las herramientas estructuradas de valoración del riesgo de violencia tienen una capacidad moderadamente satisfactoria para estimar con precisión el riesgo relativo de reincidencia (o de violencia, según los casos), evidencian sin embargo una capacidad muy pobre para estimar el riesgo absoluto de reincidencia (o de violencia, en su caso).

La buena capacidad para estimar el riesgo relativo de violencia o de reincidencia que tienen los instrumentos estructurados de valoración del riesgo puede constituir una información útil para la toma de determinadas decisiones en el ámbito judicial, especialmente aquellas que tengan que ver con la asignación de recursos escasos a un grupo de personas. Supongamos que un centro penitenciario dispone de 20 plazas en el programa de tratamiento para los agresores sexuales, pero tiene 40 internos condenados por esos delitos. Un instrumento que pueda discriminar bien los 20 entre dichos 40 que tienen más probabilidades de reincidencia (estimación de riesgo relativa) representaría una buena base sobre la que tomar la decisión de asignar las plazas disponibles. El hecho de que en el grupo de internos a los que el instrumento asigne un riesgo elevado de reincidencia pueda haber muchos falsos positivos no importa excesivamente, pues también ellos pueden beneficiarse en alguna medida de los efectos positivos del tratamiento, aunque éste en su caso fuera menos necesario. Aquí el problema es, ante todo, económico: al haber falsos positivos en el grupo de los considerados peligrosos se destinarán recursos al tratamiento de algunas personas que en principio no lo necesitarían, pero si el instrumento tiene una elevada sensibilidad nos aseguramos de

que entre los 20 a los que se lo ofreceremos estarán la mayoría de los que sí lo necesitan¹³.

Sin embargo, hay otras muchas decisiones que han de adoptar los operadores jurídicos para las que la información pertinente no es el riesgo relativo que representa un sujeto respecto de los demás integrantes del grupo al que pertenece, sino que requieren estimaciones fiables del riesgo absoluto de reincidencia. Por ejemplo, en la regulación legal de la suspensión de la condena, los permisos penitenciarios, la libertad condicional, e incluso la revisión de la prisión permanente revisable que ha introducido en el Código penal español la Ley Orgánica 1/2015, uno de los criterios fundamentales para tomar estas decisiones es la peligrosidad de los sujetos. Aquí no hay un número máximo de permisos o de libertades condicionales que se puedan conceder al año, y lo que interesa saber no es si el interno sobre el que hay que tomar la decisión es más o menos peligroso que otros, sino si su probabilidad de reincidencia es baja o elevada¹⁴. Y cuando el juez solicita auxilio a un perito para estimar el riesgo de reincidencia violenta de cada sujeto concreto sobre el que tiene que decidir lo que necesita saber es qué grado de acierto tiene el instrumento de valoración que se le ha aplicado y que le asigna una determinada probabilidad de reincidencia. Es decir, lo relevante aquí es hasta qué punto es cierto que este sujeto tiene las probabilidades de reincidir que dice el instrumento. Adviértase que aquí el problema, además de económico¹⁵, tiene también o incluso sobre todo otra dimensión: ¿es *justo* limitar derechos fundamentales a los individuos sobre la base de estimaciones de riesgo (*absoluto*) con muy pocas probabilidades de acierto y sometidas a grandes márgenes de error? Aunque evidentemente la respuesta a esta pregunta ya no es técnica sino valorativa, y no debe darla el perito sino el juez, precisamente porque es muy limitada la capacidad de las herramientas de valoración del riesgo para proporcionar la información que aquí es pertinente (es decir, para estimar con precisión la tasa de reincidencia esperable de los sujetos clasificados en cada nivel

¹³ Un contexto de decisión parecido en el que también podrá ser útil aplicar estos instrumentos es el ofrecimiento de programas de apoyo y prevención de la reincidencia en la comunidad, una vez excarcelado el sujeto, como los “Círculos de Apoyo y Responsabilidad” puestos en marcha recientemente en Cataluña. En este programa los criterios de selección son – entre otros – tanto un riesgo moderado o alto de reincidencia sexual como la participación voluntaria del sujeto (Nguyen et al., 2014, p. 4).

¹⁴ Aunque ambas cuestiones puedan estar en cierta forma relacionadas.

¹⁵ Que también, por los costes que supone una decisión equivocada: tanto por lo caro que resulta mantener en prisión – quizá por muchos años aún – a quien podría haber salido en libertad, como por los costes que supondría la comisión de un nuevo delito (los daños materiales y personales que éste conlleva, los costes de la tramitación del nuevo proceso, etc.).

de riesgo; o dicho de otra manera: para estimar con precisión el riesgo absoluto de reincidencia), es por lo que diversos autores han advertido en los últimos años de lo cuestionable que puede resultar adoptar decisiones que limitan de manera muy importante los derechos fundamentales de los individuos basadas en estimaciones de un riesgo futuro de delincuencia o de delincuencia violenta sometidos a un grado de incertidumbre tan elevado (Fazel et al., 2012; Singh et al., 2014). Y ello, aunque los instrumentos con que se valore el riesgo de reincidencia hayan evidenciado repetidamente valores elevados de precisión en la predicción del riesgo *relativo*.

3.3. Otros problemas que afectan al grado de precisión de las valoraciones del riesgo de violencia

Por último, y desde esta misma perspectiva (es decir, teniendo en cuenta el tipo de decisiones para las que las leyes penales exigen efectuar valoraciones del riesgo de violencia o de reincidencia), me referiré muy brevemente a otros tres problemas relacionados con la precisión de las estimaciones de riesgo. El primero es antiguo y muy conocido: la influencia que tiene en las posibilidades de predicción correcta del riesgo de reincidencia o de violencia la baja prevalencia del fenómeno que se quiere predecir. Es sabido que las tasas de prevalencia bajas afectan negativamente al valor predictivo positivo, y el Derecho penal suele estar interesado en obtener predicciones del riesgo de violencia respecto de los responsables de delitos más graves, que son los menos frecuentes, por lo que precisamente en las predicciones que más interesan al Derecho penal es muy probable que los valores predictivos positivos sean muy bajos, y haya por tanto muchos falsos positivos. Por otro lado, el desconocimiento de las tasas base de reincidencia para diversos grupos de delitos y de delincuentes en una determinada población genera una notable incertidumbre sobre la proporción de falsos positivos y falsos negativos que se obtendrá al aplicar un instrumento con una sensibilidad dada¹⁶ (cfr. sobre todo ello Martínez Garay, 2014).

¹⁶ Por ejemplo, según la regulación de la revisión de la prisión permanente revisable en el art. 92 del Código Penal ésta podrá dejarse en suspenso tras 25 años de cumplimiento si el sujeto presenta un pronóstico favorable de reinserción social. Teniendo en cuenta que la media de edad de los que puedan ser candidatos a la revisión rondará los 50 años más o menos, ¿cuál es la tasa de reincidencia de los sujetos de entre 50 y 60 años de edad autores de delitos contra la vida tras el cumplimiento de más de 20 años de prisión)? No hay estudios en España que proporcionen información sobre esto.

En segundo lugar, en los últimos años se ha cuestionado seriamente que las probabilidades numéricas que asignan los instrumentos estructurados de valoración del riesgo de violencia para los *grupos* de sujetos que entran en las diferentes categorías de riesgo se pueden utilizar sin más para estimar las probabilidades de reincidencia de los *individuos* concretos, porque se aduce que cuando se aplica la probabilidad calculada para un grupo a uno de los individuos de ese grupo se incrementan exponencialmente los márgenes de error en la estimación (Cooke y Michie, 2010; Cooke y Michie 2011; Hart y Cooke, 2013). Aunque esta tesis ha sido cuestionada por otros autores y es en la actualidad objeto de debate (cfr., en sentido crítico respecto de los anteriores, los trabajos de Harris, Lowenkamp y Hilton, 2015, o Mossman, 2015), siembra en todo caso nuevas dudas sobre la precisión de las estimaciones de riesgo absoluto hechas sobre individuos, que son las decisivas en la mayor parte de las decisiones que los operadores jurídicos han de adoptar.

Por último, comparto la opinión de quienes sugieren que, dado el tipo de condicionantes de los que deriva la limitada capacidad predictiva de los instrumentos de valoración del riesgo de violencia, no parece que al menos en un futuro cercano dicha capacidad vaya a aumentar notablemente, y ello aun cuando deban seguir explorándose por supuesto técnicas o vías novedosas que permitan aumentar nuestro conocimiento en este campo¹⁷. Como afirman Yang, Wong y Coid (2010):

After almost five decades of developing risk prediction tools, the evidence increasingly suggests that the ceiling of predictive efficacy may have been reached with the available technology. Other approaches such as tree modeling (Steadman et al., 2000) and Neural Networks (Price et al., 2000) require further exploration, but it is unlikely that a very high level of predictive accuracy is achievable because of theoretical constraints (p. 759).

3.4. Conclusiones

Las conclusiones que a mi juicio cabe extraer de todo lo expuesto son las siguientes:

¹⁷ Sobre las posibilidades que ofrecen las neurociencias en este contexto cfr. por ej. Aharoni et al. (2014), según los cuales los marcadores neurobiológicos podrían aportar información que mejorara la utilidad de los instrumentos de valoración del riesgo actualmente existentes, aunque se muestran bastante escépticos sobre la posibilidad de que dicha información pueda llegar a garantizar predicciones acertadas sobre individuos concretos.

1. El documento *Tasa de reincidencia penitenciaria 2014* en Cataluña (Capdevila Capdevila et al. 2015) es un excelente estudio que ofrece mucha información y muy valiosa sobre la reincidencia y sobre los factores que están asociados a ella. Sin embargo, al informar sobre los resultados que arrojó la aplicación del RisCanvi a una muestra de 638 sujetos para predecir el riesgo de reincidencia, presenta los datos de una forma que puede inducir a confusión al lector, porque no diferencia suficientemente entre sensibilidad y especificidad, por un lado, y valor predictivo por otro. El "acierto en el pronóstico" que se obtuvo al aplicar dicho instrumento a esa muestra fue del 77.2% si por tal se entiende la sensibilidad, pero sólo del 17.94% si lo entendemos como valor predictivo positivo. Al omitir esta última información se genera probablemente una imagen excesivamente optimista sobre la capacidad predictiva del RisCanvi.

2. No obstante, el valor predictivo del RisCanvi en este caso no difiere en exceso del que evidencian muchos otros instrumentos de valoración del riesgo muy conocidos y utilizados a nivel internacional.

3. Existen diferentes formas de medir la capacidad predictiva de los instrumentos de valoración del riesgo de violencia o de reincidencia, y no todas ellas aportan la misma clase de información. Estos instrumentos pueden proporcionar información muy valiosa para el progreso de la criminología y pueden ser también de utilidad en la práctica para la adopción de determinadas decisiones relacionadas con la política criminal y la gestión penitenciaria, pero siempre que se sea consciente de sus limitaciones y de qué es lo que se está midiendo o estimando en cada caso. Si bien alcanzan unos niveles de precisión probablemente satisfactorios en la estimación del riesgo relativo de reincidencia, no hay que olvidar que la evidencia empírica sigue demostrando que son mucho más imprecisos en la estimación del riesgo absoluto de reincidencia, y que producen una notable y sistemática sobreestimación de la peligrosidad especialmente cuando se aplican a la predicción de fenómenos con bajas tasas de prevalencia como lo es la delincuencia violenta. En este contexto, en el que en muchas de las decisiones judiciales que requieren la estimación de la peligrosidad del sujeto está en juego la libertad de los ciudadanos, es en mi opinión peligroso generar demasiadas expectativas acerca de las posibilidades reales de predicción de la reincidencia y especialmente de la reincidencia violenta, cuando los datos existentes a

nivel nacional e internacional evidencian que la capacidad predictiva (de riesgo absoluto) de todas estas herramientas sigue siendo muy limitada, al menos en la actualidad y probablemente también en el futuro a medio plazo.

4. Es aconsejable que los investigadores sean extremadamente cautelosos en la interpretación de los resultados de sus estudios y también al transmitir dicha información, de manera que se eviten confusiones sobre el significado de los valores obtenidos. Es deseable que se especifique siempre qué es exactamente lo que se está midiendo y que se hagan explícitas las limitaciones y los márgenes de error a que están sometidas las predicciones. Hay que tener en cuenta que en un contexto en el que el legislador hace depender de los pronósticos de peligrosidad cada vez más decisiones – y cada vez más gravosas¹⁸ – sobre la libertad de los delincuentes, los operadores jurídicos que se ven obligados a adoptarlas y que cargan con la responsabilidad de los errores que se puedan cometer (ya sea por falsos positivos o por falsos negativos) recurrirán sin cuestionarlo a cualquier dato que aparente ofrecer una base científica para estas decisiones, aun cuando no lo hayan entendido correctamente, o el dato no proporcione en realidad la información que ellos crean que aporta. Seguramente sería conveniente que los investigadores advirtieran expresamente – al menos en los estudios que estén directa o indirectamente relacionados con el ámbito forense – que el hecho de que un instrumento de valoración del riesgo evidencie un acierto elevado en la estimación del riesgo relativo de reincidencia o de delincuencia no significa que sea capaz de estimar la probabilidad absoluta de reincidencia o de delincuencia de los individuos o de los grupos con un grado de precisión aceptable. Además, en la presentación de los resultados se debería diferenciar con claridad entre la sensibilidad y el valor predictivo positivo.

Agradecimientos

Agradezco muy especialmente al Dr. Francisco Montes Suay, Catedrático de Estadística e Investigación Operativa de la Universidad de Valencia, su inestimable ayuda al haber resuelto mis dudas estadísticas y revisado varios epígrafes del trabajo. Agradezco también a los evaluadores anónimos de la REIC y al editor de la revista su concienzuda

¹⁸ El ejemplo más paradigmático es como ya hemos señalado *supra* la revisión de la pena de prisión perpetua revisable, introducida en nuestro país por la Ley Orgánica 1/2015, que depende de que quepa realizar “un pronóstico favorable de reinserción social”.

revisión del manuscrito inicial y las sugerencias que me hicieron a lo largo del proceso de evaluación, lo que ha contribuido a mejorar notablemente el trabajo. Finalmente, quiero agradecer a la Dra. Clara Viana Ballester, Profesora de Derecho Penal de la Universidad de Valencia, sus valiosos consejos y su apoyo, al inicio y a lo largo de esta investigación.

Financiación

Este trabajo se ha realizado en el marco de los proyectos de I+D+I “Ejecución de penas y medidas de seguridad. Tendencias político-criminales en las modernas reformas penales” (DER2013-47859-R), financiado por el Ministerio de Economía y Competitividad, y “La gestión de los antecedentes penales como medio de prevención del delito” (GV/2016/085), financiado por la Generalitat Valenciana.

4. Referencias

- Aharoni, E., Mallett, J., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W... Kiehl, K. A. (2014). Predictive accuracy in the neuroprediction of rearrest. *Social Neuroscience*, 9(4), 332-336. doi: 10.1080/17470919.2014.907201
- Aitken, C. G. G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists* (2ª ed.). Chichester: John Wiley & Sons. [OX Reader version].
- Albrecht, G. (2004). Sinn und Unsinn der Prognose von Gewaltkriminalität. En W. Heitmeyer & H. G. Soeffner (Eds.), *Gewalt. Entwicklungen, Strukturen, Analyseprobleme* (pp. 475-524). Frankfurt: Suhrkamp.
- Alex, M. (2013). *Nachträgliche Sicherungsverwahrung – ein rechtsstaatliches und kriminalpolitisches Debakel* (2ª ed.). Holzkirchen: Felix Verlag.
- Andrés-Pueyo, A. (2013). Valoració del risc i gestió de la reincidència: la utilitat del RisCanvi en la reinserció. En J. Cid, M. Ferrer & A. Ibáñez (Coords.), *De l'execució de penes a la reinserció* (pp. 67-70). Barcelona: UAB.
- Andrés-Pueyo, A. & Echeburúa, E. (2010). Valoración del riesgo de violencia: instrumentos disponibles e indicaciones de aplicación. *Psicothema*, 22(3), 403-409.

- Andrés-Pueyo, A. & Redondo, S. (2007). Predicción de la violencia: entre la peligrosidad y la valoración del riesgo de violencia. *Papeles del Psicólogo*, 28(3), 157-173.
- Barnett, G. D., Wakeling, H. C. & Howard, P. D. (2010). An Examination of the Predictive Validity of the Risk Matrix 2000 in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, 22(4), 443-470. doi: 10.1177/1079063210384274
- Buchanan, A, Binder, R., Norko, M. & Swartz, M. (2012). Resource Document on Psychiatric Violence Risk Assessment. *American Journal of Psychiatry*, 169(3), 340. <http://dx.doi.org/10.1176/appi.focus.130402>.
- Capdevila Capdevila, M. (Coord.), Blanch Serentill, M., Ferrer Puig, M., Andrés-Pueyo, A., Framis Ferrer, B., Comas López, N.,...Mora Encinas, J. (2015), *Tasa de reincidencia penitenciaria 2014*. Centro de Estudios Jurídicos y Formación Especializada, Generalitat de Catalunya.
- Cooke, D. J. & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: a challenge for forensic practice. *Law and Human Behaviour*, 34, 259-274. doi: 10.1007/s10979-009-9176-x.
- Cooke, D. J. & Michie, C. (2011). Violence risk assessment. Challenging the illusion of certainty. En B. McSherry & P. Keizer (Eds.), *Dangerous people. Policy, prediction and practice* (pp. 147-161). New York-London: Routledge.
- Díez Ripollés, J. L. (2012). La política criminal española. *Claves de Razón Práctica* 219, 40-49 (cit. por la versión recogida en Díez Ripollés, J. L. (2015). *Delitos y penas en España*. Madrid: Catarata).
- Díez Ripollés, J. L., Giménez-Salinas Framis, A., Andrés-Pueyo, A., Barberet Havicam, R., Cerezo Domínguez, A., Cid Moliné, J., ... Redondo, S. (2010, Marzo 24). Los datos de la criminalidad. *El Imparcial*. Recuperado de <http://www.elimparcial.es/noticia/59836/nacional/Investigadores-y-academicos-denuncian-el-secretismo-de-Interior-sobre-criminalidad.html>, cit. por la versión recogida en Díez Ripollés, J. L. (2015). *Delitos y penas en España*, Madrid: Catarata, pp. 93-97.
- Fazel, S., Singh, J. P., Doll, H., & Grann, M. (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. doi: 10.1136/bmj.e4692.
- Hanson, R. K. & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1-21. doi: 10.1037/a0014421.
- Harcourt, B. E. (2007). *Against Prediction: profiling, policing and punishing in an actuarial age*. The University of Chicago Press.
- Harris, G. T., Lowenkamp, C. T. & Hilton, N. Z. (2015). Evidence for Risk Estimate Precision: Implications for Individual Risk Communication. *Behavioral Sciences and the Law*, 33, 111-127. doi: 10.1002/bsl.2158.
- Hart, S. D. & Cooke, D. J. (2013). Another look at the (im-)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences and the Law*, 31, 81-102. doi: 10.1002/bsl.2049.
- Heilbrun, K. (2009). *Evaluation for risk of violence in adults*. New York: Oxford University Press [OX Reader version]. Recuperado de <http://site.ebrary.com/lib/universvaln/detail.action?docID=10300134>.
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M. & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: a meta-analysis. *Criminal Justice and Behaviour*, 39, 1148-1171. doi: 10.1177/0093854812443648.
- Kinzig, J. (2010). *Die Legalbewährung gefährlicher Rückfalltäter. Zugleich ein Beitrag zur Entwicklung des Rechts der Sicherungsverwahrung* (2ª ed.). Berlin: Duncker & Humblot.
- Martínez Garay, L. (2014). La incertidumbre de los pronósticos de peligrosidad: consecuencias para la dogmática de las medidas de seguridad (epígrafe 5 redactado en coautoría con Montes Suay, F.). *InDret. Revista para el análisis del Derecho* 2/2014, 1-77.
- Monahan, J. (1981). *Predicting violent behaviour. An assessment on clinical techniques*. London: Sage Publications.
- Mossman, D. (2015). From Group Data to Useful Probabilities: The Relevance of Actuarial Risk Assessment in Individual Instances. *The Journal of the American Academy of Psychiatry and the Law*, 43(1), 93-102.

- Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment*, 18(1), 41-63. doi: 10.1177/107906320601800104.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62(4), 783-792.
- Müller, J. L. & Stolpmann, G. (2015). Legalbewährung nach rechtskräftiger Ablehnung einer nachträglichen Anordnung der Unterbringung in der Sicherungsverwahrung. *Monatsschrift für Kriminologie und Strafrechtsreform*, 98(1), 35-47.
- Nguyen, T., Frerich, N., García, C. Soler, C., Redondo, S. & Andrés-Pueyo, A. (2014). Reinserción y gestión del riesgo de reincidencia en agresores sexuales excarcelados: el proyecto 'Círculos de Apoyo y Responsabilidad' en Cataluña. *Boletín Criminológico* nº 151, 1-5.
- Rusche, S. (2004). *In Freiheit gefährlich? Eine Untersuchung zu Häufigkeit und Gründen falscher Kriminalprognosen bei psychisch kranken Gewaltverbrechern*. Regensburg: Roderer.
- Singh, J. P., Fazel, S., Gueorguieva, R. & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk assessment instruments. *The British Journal of Psychiatry*, 204(3), 180-197. doi: 10.1192/bjp.bp.113.131938.
- Singh, J. P., Grann, M. & Fazel, S. (2011). A comparative study of violence risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25.980 participants. *Clinical Psychology Review*, 31, 499-513.
- Singh, J.P. & Petrila, J. (2013). Measuring and Interpreting the Predictive Validity of Violence Risk Assessments: An Overview of the Special Issue. *Behavioral Sciences and the Law*, 31, 1-7. doi: 10.1002/bsl.2054.
- Sjöstedt, G. & Grann, M. (2002). Risk Assessment: What is Being Predicted by Actuarial Prediction Instruments? *International Journal of Forensic Mental Health*, 1(2), 179-183. doi: 10.1080/14999013.2002.10471172.
- Skeem, J. L. & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, 20(1), 38-42. doi: 10.1177/0963721410397271.

- Soler, C. (2013, abril). *RisCanvi. Protocolo de evaluación y gestión del riesgo de violencia con población penitenciaria*. Comunicación presentada al Foro internacional de buenas prácticas en prevención de la delincuencia juvenil. Bogotá. Recuperado de [http://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/138A3DBF8E8A85B905257C9F00803A14/\\$FILE/LinkClick6.pdf](http://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/138A3DBF8E8A85B905257C9F00803A14/$FILE/LinkClick6.pdf)
- Steadman, H. J. (2000). From Dangerousness to Risk Assessment of Community Violence: taking stock at the turn of the century. *Journal of the American Academy of Psychiatry and the Law*, 28, 265-271.
- Urbaniok, F., Noll, T., Grunewald, S., Steinbach, J. & Endrass, J. (2006). Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. *The Journal of Forensic Psychiatry & Psychology*, 17(1), 23-31. doi: 10.1080/02699200500297799.
- Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136(5), 740-767. doi: 10.1037/a0020473.

Lucía Martínez Garay es Doctora en Derecho y Licenciada en Criminología por la Universidad de Valencia. Ha trabajado como investigadora en la Universitat Jaume I de Castellón y como profesora ayudante en la Universidad de Alicante. En la actualidad es Profesora Titular en el Departamento Derecho penal de la Universidad de Valencia. Sus principales áreas de interés son la imputabilidad penal, la peligrosidad y las medidas de seguridad, y los delitos de terrorismo y de pertenencia a organizaciones criminales.