THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF EXPERT CLINICAL AND ACTUARIAL TESTIMONY[1]

Andrés Páez

Professor Associado do Departamento de Filosofia da Universidade de los Andes, Bogotá, Colômbia; Presidente da Associação Latino-americana de Filosofia Analítica (ALFAn). apaez@uniandes.edu.co

A predição do comportamento futuro: as promessas vazias das perícias atuarial e clínica

ABSTRACT

Testimony about the future dangerousness of a person has become a central staple of many judicial processes. In settings such as bail, sentencing, and parole decisions, in rulings about the civil confinement of the mentally ill, and in custody decisions in a context of domestic violence, the assessment of a person's propensity towards physical or sexual violence is regarded as a deciding factor. These assessments can be based on two forms of expert testimony: actuarial or clinical. The purpose of this paper is to examine the scientific and epistemological basis of both methods of prediction or risk assessment. My analysis will reveal that this kind of expert testimony is scientifically baseless. The problems I will discuss will generate a dilemma for factfinders: on the one hand, given the weak predictive abilities of the branches of science involved, they should not admit expert clinical or actuarial testimony as evidence; on the other hand, there is a very strong tradition and a vast jurisprudence that supports the continued use of this kind of expert testimony. It is a clear case of the not so uncommon conflict between science and legal tradition.

Keywords: Behavior Prediction, Clinical Testimony, Actuarial Testimony, Risk Assessment, Mental Illness.

Resumo

O testemunho sobre a periculosidade futura de uma pessoa tem sido uma marca central de muitos processos judiciais. Em contextos tais como pagamento de fiança, proferimento de sentença e decisões sobre liberdade condicional, confinamento civil do doente mental e costódia em casos de violência doméstica, a avaliação da propensão de uma pessoa à violência física ou sexual é considerada um fator fundamental. Essas avaliações podem ser baseadas em duas formas de perícia: atuarial ou clínica. O propósito deste artigo é examinar a base científica e epistemológica de ambos os métodos de predição ou de avaliação de risco. Minha análise revelará que este tipo de perícia não possui fundamento. Os problemas que serão discutidos gerarão um dilema para os investigadores dos fatos no processo judicial: por um lado, dadas as habilidades preditivas fracas dos ramos da ciência envolvidos, não deveriam admitir a perícia atuarial e clínica como meios de prova; por outro lado, há uma tradição muito forte e uma vasta jurisprudência que dá suporte ao uso contínuo deste tipo de prova pericial. Este é um caso claro do não tão incomum conflito entre a ciência e a tradição jurídica.

Keywords: Predição do comportamento, Perícia clínica, Perícia atuarial, Avaliação de risco, Doença mental.

1. INTRODUCTION

In October 2012, six Italian seismologists and a government official were found guilty of manslaughter and sentenced to six years each for failing to adequately warn residents of the risk before the earthquake that hit the city of L'Aquila in 2009 killing 308 people. The verdict sent shockwaves throughout the scientific community. *Nature* published an editorial calling the verdict "perverse and the sentence ludicrous"[2]. In 2014, an appeals court acquitted the seismologists and reduced to two years the sentence of the government official.

Unreasonable expectations about the predictive abilities of a branch of science are not uncommon in law. They are at the basis not only of unwarranted negligence claims, but also of legal decisions that have a scientific veneer but that make use of predictions that the scientific community at large does not sanction. Consider the *Adam Walsh Child Protection and Safety Act* (2006), an American federal statute that establishes among its provisions a controversial post-conviction

---

2 Editorial, 2012, p. 446.

civil commitment scheme. The Commitment Provision[3] allows the federal Bureau of Prisons to keep inmates in prison past their release date if the government can prove "by clear and convincing evidence"[4] that the inmate is a "sexually dangerous person", i.e. one "who has engaged or attempted to engage in sexually violent conduct or child molestation and who is sexually dangerous to others"[5]. A person is sexually dangerous to others insofar as he "suffers from a serious mental illness, abnormality, or disorder as a result of which he would have serious difficulty in refraining from sexually violent conduct or child molestation if released"[6]. The American Psychiatric Association has repeatedly objected to the civil commitment of sex offenders on scientific grounds. In 1999, a task force created by the APA declared:

> [S]exual predator commitment laws represent a serious assault on the integrity of psychiatry, particularly with regard to defining mental illness and the clinical conditions for compulsory treatment. Moreover, by bending civil commitment to serve essentially non-medical purposes, sexual predator commitment statutes threaten to undermine the legitimacy of the medical model of commitment. (...) [P]sychiatry must vigorously oppose these statutes, to preserve the moral authority of the profession and to ensure continuing societal confidence in the medical model of civil commitment[7].

Most of the debate has focused on the clinical methods used to assess the risk that the offender is likely to commit future acts of sexual violence, and on the type of information that can be used in that assessment. During the process, the Department of Justice can consider past conduct that did not result in an arrest, prosecution or conviction. In fact, offenders can be certified for civil commitment even if they have no prior criminal record of sex offenses: "It is not

The American Psychiatric Association has repeatedly objected to the civil commitment of sex offenders on scientific grounds.

---

3 The Commitment Provision has its roots in two previous Supreme Court decisions: *Kansas v. Hendriks* (1997) and *Kansas v. Crane* (2002).

4 18 U.S.C. §4248(d). The standard is lower than the one required for the original criminal conviction, viz. "beyond all reasonable doubt".

5 18 U.S.C. §4247(a)(5).

6 18 U.S.C. §4247(a)(6).

7 American Psychiatric Association Task Force on Sexually Dangerous Offenders, 1999, p. 173. The Commitment Provision has also been challenged on constitutional grounds. In 2009, the United States Court of Appeals for the Fourth Circuit, in Richmond, Va., ruled that none of the powers granted to Congress in the Constitution empowered it to authorize such civil commitments. In 2010 the Supreme Court upheld the Adam Walsh Act in *United States v. Comstock*. Justice Breyer made it clear that the court was not ruling on the separate question of whether the Commitment Provision violated the Constitution's due process clause. See Baker (2009) for a discussion of the constitutional issues involved.

necessary that a person have been charged with or convicted of any criminal act related to the conduct being considered — a limitation that could prevent a mental health professional from considering probative and relevant evidence such as long-established patterns of behaviour, admissions of criminal activity previously undetected by authorities, and statements of intent to commit future sexually violent crimes or acts of child molestation"[8].

The use of scientific predictions in law is limited to the two types of cases illustrated by the previous two examples. The first type is constituted by negligence claims. Negligence covers a wide territory, including claims related to a doctor's liability for a patient's lost chance to recover from illness, claims about the unforeseen negative impact of a product or a construction project on the environment or on people's health, or about the unanticipated dire financial effects of a stock market transaction on stockholders' investments. Negligence claims are very fact-specific and the examples could be multiplied *ad nauseam*. If one is interested in the *limits* of scientific prediction, one must distinguish between liability caused by scientific errors and mistakes due to carelessness, inattention, or purposeful deception, from liability caused by scientific predictions that could have reasonably been made at the time but were not[9]. It is the latter type of liability that will be relevant[10].

The second type of case involves the prediction of people's future behavior, especially of a violent or sexual nature, in settings such as capital sentencing, bail and parole decisions, rulings about the civil confinement of the mentally ill, and custody decisions in a context of domestic violence.[11] These predictions can be based on two forms of expert testimony: actuarial or clinical. By far, the most common is clinical prediction, either by a psychiatrist or a psychologist, although actuarial testimony is often used in parole board hearings.

---

8 Bureau of Prisons, 2007, p. 43207.

9 The difference between a scientifically predictable and a reasonably foreseeable event, as defined in the law of torts, is a question of degree. Almost all the cases I have in mind require expert testimony given the mathematical (probabilistic) nature of the facts involved.

10 For a clarification of the basic concepts involved, see WRIGHT, 1988.

11 The two types of cases came together in the 1970s, when courts imposed tort liability on clinicians who negligently failed to predict their patients' future violent behavior. The therapist's duty to protect potential victims stems from the Supreme Court of California's decision in *Tarasoff v. Regents of the University of California 1976*.

(...) my strategy will be to focus on the slim scientific and epistemological basis of the prediction of future behavior in general.

The purpose of this paper is to examine this second type of prediction in legal contexts. The legal scenarios in which future behavior is relevant vary widely, from capital sentencing to child custody, and the decisionmakers range from juries and judges, to parole boards. It is therefore difficult to take into account all the specific details and circumstances involved in the use of predictions of future behavior in particular cases. Instead, my strategy will be to focus on the slim scientific and epistemological basis of the prediction of future behavior in general. The problems I will discuss will generate a dilemma for judges: on the one hand, given the weak predictive abilities of the branches of science involved, they should not admit expert clinical or actuarial testimony as evidence[12]; on the other hand, there is a very strong tradition and a vast jurisprudence that supports the continued use of this kind of expert testimony. The resolution of this dilemma lies in an epistemically responsible revision of the admissibility rules for this kind of evidence.

In the final section I will examine some consequences of adopting either method of behavior prediction. I will focus on the strong epistemic dependence on expert testimony in this type of decisions, a dependence that exceeds almost any other legal scenario because it is very difficult to find ancillary evidence to support a prediction about future behavior that does not involve more expert testimony. I will argue that in some cases this dependency on expert testimony jeopardizes the independence between the scientific and the legal standard of proof, between the level of confidence required to make the prediction and the standard of proof used to decide the case. I will also argue that reliance on actuarial methods in particular makes it more likely that the resulting judicial policies that emerge will be more an artifact of the instrument used than of sound political and moral decisions. Although my analysis will be circumscribed to American law, the lessons for other legal systems are fairly obvious.

---

12 *Daubert v. Merrell Dow Pharmaceuticals 1993.*

2. CLINICAL TESTIMONY

In this section I will examine the scientific and epistemological foundations of clinical testimony, the first of the two types of expert testimony regularly used to determine the future behavior of a defendant or an offender[13]. Clinical testimony is widely regarded as the more problematic of the two generators of future behavior predictions[14].

There is little controversy regarding the kind of future criminal behaviour that the legal system ought to prevent. But it is far from clear what findings of fact will enable it to predict the behaviour it seeks to prevent. Let us return initially to the Commitment Provision of the *Adam Walsh Act* (AWA)[15] and examine the methodology used to assess the risk that the inmate is likely to commit future acts of sexual violence if released. According to the AWA, a finding of likelihood of sexual recidivism based on the diagnosis of a volitional impairment is a necessary requirement for civil commitment. The problem is that there is no conceptual clarity on what exactly is being diagnosed. "The field of risk assessment [of the likelihood of sexual recidivism] is lacking consistent empirical support defining volitional impairments relevant to a threshold for legal civil commitment"[16]. The statute does not define the terms "serious mental illness", "abnormality", or "disorder", mentioned in the definition of a sexually dangerous person, expecting perhaps that expert clinical testimony, the DSM (Diagnostic and Statistical Manual of Mental Disorders) or the ICD (International Statistical

---

13 Clinical testimony can be either completely unstructured, based solely on the clinician's experience, or aided by actuarial instruments. Most of the discussion in this section refers to unstructured clinical assessments, which are the most common basis for expert testimony in court. Since the next section criticizes the epistemological basis of current actuarial instruments in general, I consider it unnecessary to devote a separate section to "mixed" methods that include both clinical and actuarial criteria.

14 For a defense of the view that individualized clinical assessments of the likely dangerousness of people are superior to actuarial ones, see LITWACK, 2001; for the opposing view targeted by LITWACK, see QUINSEY et al., 1999. HARCOURT, 2007, also opposes actuarial methods, but instead of individualized assessments of future dangerousness, he proposes a turn to randomness in punishment and policing.

15 Under Canadian law a court can impose a similar sentence called "indeterminate detention". It is imposed, among other reasons, when "the offender, by his or her conduct in any sexual matter including that involved in the commission of the offence for which he or she has been convicted, has shown a failure to control his or her sexual impulses and a likelihood of causing injury, pain or other evil to other persons through failure in the future to control his or her sexual impulses" (*Criminal Code, X*XIV, 753 (1)(b)).

16 FABIAN, 2012, p. 309.

Classification of Diseases and Related Health Problems)[17] will be able to fill in the blanks.

Fitch reports that 85% of offenders committed under sexually violent predatory laws in the United States have been diagnosed with paraphilias, which are abnormal sexual behaviors listed both in the DSM and the ICD[18]. But a paraphilia is a highly controversial concept that depends on what is considered sexually deviant at a particular place and time. Up until 1973, for example, homosexuality was classified as a paraphilia under the DSM-II. Any proposed definition of "paraphilia'' is thus "vulnerable to societal pressures rather than advances in science"[19]. More recent versions of the manuals have not provided further clarity about the concept. There are significant differences between the DSM-5 (American Psychiatric Association, 2013) and the ICD-10 (World Health Organization, 2010) regarding which disorders are included and how they are categorized[20]. It is not surprising, therefore that there is a growing chorus calling for the removal of paraphilias from the DSM[21].

Philosophers of science have also questioned the theoretical basis of these classifications. Murray, for example, argues that the current literature on mental illness lacks a coherent concept of the mental and a satisfactory account of disorder[22]. In his view, the DSM classifies mental illnesses according to their discernible symptoms, ignoring the underlying causal structure of the mind. This classification is useful for treatment purposes, but without an understanding of the mental mechanisms involved it seems woefully unsuitable for the purpose of predicting future behavior[23].

Aside from the controversial and highly debated definitional issues regarding paraphilias and mental illnesses, there are many other straightforward and noncontroversial methodological problems with the prediction of future behaviour based on clinical diagnosis. In what follows I will briefly present some of the main issues.

---

17 All WHO member countries are required to follow the definitions in the ICD, but the DSM carries more weight among psychiatrists and psychologists.

18 FITCH, 2003.

19 ZONANA, 2011, p. 249.

20 REED, 2010.

21 MOSER and KLEINPLATZ, 2005.

22 MURRAY, 2006.

23 I am grateful to Santiago Amaya for calling my attention to this issue.

(...) a paraphilia is a highly controversial concept that depends on what is considered sexually deviant at a particular place and time.

The first problem with clinical expert testimony is that experts often overestimate base rates[24], which in turn leads to over-conservative decisions regarding the release of prisoners. Prediction of an event becomes more difficult the further downwards the base rate departs from 0.5. The base rate for recidivism of violent crimes, according to many studies, is never higher than 20%[25], and for sexual crimes it is between 10% and 15% after 5 years, and 20% after 10 years[26]. Additionally, people with medical training, such as psychiatrists, tend to overestimate the base rate to a higher degree than psychologists[27]. Decisionmakers also often overestimate how much information they possess. This leads them to overconfidence and to decisions that are not warranted by the data[28]. They also rely on highly salient information that has relatively little predictive value, such as history of institutional violence[29].

Often mental health practitioners testifying in court shore up their judgments by mentioning their accumulated experience. In statements such as: "In my 20 years of experience judging similar cases…" the listener is supposed to find a good reason to believe the expert witness' testimony. Although accumulated experience is a necessary condition for expertise and a good reason to believe a person's judgment in many contexts, research shows that in the case of clinical prediction accumulated experience is basically irrelevant. Several authors have shown that there is little empirical evidence supporting expert status for clinicians on the basis of their training, experience, or information processing ability[30]. Often experts with the same degree of experience disagree in their diagnosis of psychiatric patients[31], and despite the lack of correlation between experience and predictive accuracy, the confidence of clinicians in their diagnosis increases with experience[32]. Several studies have

---

24 Base rates refer to the relative frequency of a state or condition in a population or a series of events. The overestimation of base rates might be the result of the fact that clinical experts rely on their professional practice, in which the frequency of mental illnesses is much higher than in the general population.

25 YANG et al., 2010.

26 HANSON, 2003.

27 QUINSEY, 1981.

28 WIGGINS, 1973.

29 QUINSEY, 1979.

30 WIGGINS, 1973; SCHINKA & SINES, 1974.

31 QUINCEY and AMBTMAN, 1979.

32 GOLDBERG, 1968.

compared the predictive abilities of experienced clinicians with that of graduate students and lay people. In a famous study, Quincey and Ambtman compared the predictions of nine teachers and four senior forensic psychiatrists regarding the behavior of 30 people who had been released from prison. Based on the same data set, the teachers were more accurate in their predictions than the experts[33].

In view of all of these methodological problems, in an *amicus curiae* addressed to the Supreme Court of the United States in *Barefoot* (1983), the APA rejected the use of clinical predictions of long-term future behaviour:

> Psychiatrists should not be permitted to offer a prediction concerning the long-term future dangerousness of a defendant in a capital case, at least in those circumstances where the psychiatrist purports to be testifying as a medical expert possessing predictive expertise in this area. Although psychiatric assessments may permit short-term predictions of violent or assaultive behavior, medical knowledge has simply not advanced to the point where long-term predictions —the type of testimony at issue in this case— may be made with even reasonable accuracy. The large body of research in this area indicates that, even under the best of conditions, psychiatric predictions of long-term future dangerousness are wrong in at least two out of every three cases[34].

Most of the problems that led the APA to this conclusion were well known in the medical and psychological literature at the time, but they seem to have had very little effect on changing the culture of admitting clinical predictions as an acceptable form of expert testimony. Quite the contrary: In *Barefoot*, the Supreme Court stated that since the APA did not claim "that psychiatrists are always wrong with respect to future dangerousness, only most of the time"[35], it would not exclude such testimony: "we are no more convinced now that the view of the APA should be converted into a constitutional rule barring an entire category of expert testimony. (...) The suggestion that no psychiatrist's testimony may be presented with respect to a defendant's future dangerousness is somewhat like asking us to disinvent the wheel"[36]. The Court's

(...) the APA rejected the use of clinical predictions of long-term future behaviour.

---

33 QUINCEY and AMBTMAN, 1979.

34 Brief for *Amicus Curiae of the American Psychiatric Association*, 1983, p. 3.

35 *Barefoot*, p. 901.

36 *Barefoot*, p. 899 and 896.

decision made clinical predictions of future dangerousness[37] unavoidable in many court proceedings, and the issue is not likely to be challenged any time soon[38].

3. ACTUARIAL TESTIMONY

In its *amicus curiae* in *Barefoot*, the American Psychiatric Association also asserted that the most reliable predictors of long-term future dangerousness were factors that have nothing to do with psychiatric disorders or illnesses, such as age, sex, and previous convictions, among others. In consequence:

> [T]he long-term prediction of future dangerousness is an essentially lay determination that should be based not on the diagnoses and opinions of medical experts, but on the basis of predictive statistical or actuarial information that is fundamentally non-medical in nature. The psychiatric gloss on such data furnished by expert medical testimony provides little, if any, additional information to the jury[39].

The use of psychiatric expert testimony to present actuarial data can cause great damage to the defendant. By dressing up the actuarial information in a medical disguise, a psychiatrist´s testimony will receive undeserved credibility. Expert clinical testimony is more persuasive to jurors than actuarial testimony, even after cross-examination and after being confronted with testimony from a rival expert[40]. Psychiatric testimony also spares the jury the difficult task of interpreting statistical data, relying instead on the expert's interpretation, which is provided using the same expertise possessed by a layman. For these reasons, the APA considers that psychiatric testimony should be eliminated from the fact-finding process if the goal is to predict long-term future dangerousness, and that it should be replaced with actuarial instruments. Nowadays, over 60% of general psychiatric patients are routinely

---

37 It should be noted that the mysterious-sounding expression "the prediction of future dangerousness" has fallen into disuse in recent times, and most predictions are now stated in the vocabulary of "risk assessment". This reflects, according to HAMILTON (2015, p. 7-8), the shift from clinical towards actuarial methods.

38 The Court reached similar decisions in *Schall v. Martin* (1984) and *United States v. Salerno* (1987).

39 Brief for *Amicus Curiae of the American Psychiatric Association*, 1983, p. 5.

40 KRAUSS & SALEs, 2001.

assessed for violence risk using actuarial instruments[41], rising to above 80% in forensic psychiatric hospitals[42].

Policymakers in the judicial system have also embraced the use of actuarial methods. Their goal is to find cost-effective solutions to criminal offending. Actuarial methods allow officials to reserve prison resources for high-risk offenders and to identify good candidates for rehabilitation in cheaper community-based programs[43]. According to a 2004 survey, out of the 32 US states that granted parole at the time, 23 had used actuarial instruments as part of these decisions[44].

Policymakers in the judicial system have also embraced the use of actuarial methods.

There are over 150 actuarial instruments in use, and they are starting to be used in developing countries[45]. In this section I will examine some of the best-known actuarial instruments used in the prediction of future violent and sexual behavior, seeking to establish whether the faith placed in them by the judicial system and by the APA is guaranteed from an epistemological point of view. To anticipate, I will argue that the use of these methods to predict future dangerous is epistemically unacceptable and that their indiscriminate use in the judicial system can lead to serious and dire consequences.

Actuarial methods are based on statistical data about risk factors that are known to predict recidivism across contexts and individuals. These factors are combined and each factor is assigned a weighted score. The goal is to provide a final score for each individual, which is then associated to a specific risk level according to pre-established criteria. For example, an individual score between 15 and 20 on a questionnaire could mean that the individual will be classified as "high risk". Some methods are applied across the board; others are designed for specific types of offenders or crimes. Here I will focus only on violent or sexual offenders. Risk factors can be static or dynamic. Static factors include age, educational level, occupation, sexual and nonsexual criminal history, and previous alcohol problems, among others. Dynamic factors are those that can vary

---

41 HIGGINS et al., 2005.

42 KHIROYA et al., 2009.

43 HAMILTON, 2015.

44 HARCOURT, 2007.

45 FAZEL et al., 2012.

in time, in particular during the reclusion period. They include a tendency towards interpersonal conflicts, substance abuse, and tolerance towards sexual violence.

The best known and most widely used actuarial instruments for violence and sexual risk assessment are the Violence Risk Appraisal Guide (VRAG) and the Static-99[46]. VRAG assesses the risk of further violent behavior among people who have already committed violent crimes. It includes 12 items including one that requires determining psychopathy using the Psychopathy Checklist (revised) (PCL-R)[47]. VRAG is perhaps the most researched instrument in terms of replications and cross-validation[48]. SORAG[49] is a variation of VRAG for sexual offenders, in particular, for men convicted for offences of rape or child molestation. Static 99 was developed to assess the risk of recidivism among patients convicted for sexual crimes. It only considers static factors; hence the name[50].

An initial problem with these methods stems from the very nature of probabilistic phenomena: If the base rate of an event is very low, no statistical predictive instrument will be very effective[51]. As we saw in the previous section, the base rates for violent and sexual recidivism in most populations is at most 20%. It is true that the base rates for the samples used to design VRAG and Static 99 were higher: 31% for the former after 7 years, and 21% for the latter after 10 years[52]. However, there is plenty of evidence that these results do not reflect the general recidivism base rates in the general population of violent and sexual offenders. When more realistic base rates are used, the predictive success of both instruments falls dramatically. For example, when a 10% base rate is used for patients classified

---

46 HARRIS et al., 1993 and 2003.

47 See HARE, 1991. As LITWACK point out (2001, p. 413), this makes VRAG crucially dependent on clinical assessments since PCL-R measures a personality variable that requires clinical judgment to obtain. Litwack's ultimate purpose is to show that actuarial methods have not been proven to be superior to clinical ones, but if, as I argue here, both methods are equally suspect, it becomes irrelevant whether one is slightly less inaccurate than the other.

48 SKEEM and MONAHAN, 2011.

49 QUINSEY et al., 1998.

50 Other well-known actuarial instruments include the SIR (Statistical Information on Recidivism) scale and SONAR (Sex Offender Need Assessment Record). The former only includes static factors. It is modestly successful in predicting general recidivism but it is not very effective in the prediction of violent or sexual crimes; the latter focuses on changes in dynamic factors during incarceration and serves as a complement for other actuarial instruments.

51 MOSSMAN, 2008.

52 HASTINGS et al., 2011.

as "high risk" using VRAG, 69% would represent false predictions of recidivism[53]. The problem is not just that violence and sexual risk prediction is an inexact science[54]. The discussion below will show that it is seriously flawed and that it introduces undesirable distortions in the legal system.

The financial system and the insurance industry make effective use of actuarial predictive instruments. The effectiveness and confidence of the decisions they make for individual cases is based on their knowledge of all the relevant variables with a proven relationship with future risk, for example, the probability that people within a certain population will default on their mortgages. The situation is entirely different within the legal system. In particular, there is no reliable information about the true recidivism rates for several violent offenses. Recidivism rates for sexual or violent crimes are based on officially recorded information, such as an arrest, a criminal conviction, or incarceration[55]. This information is unreliable for two reasons. The first one is that it constitutes only a fraction of the true reoffense rates. It is well known that most sexual crimes, for example, are not reported. According to the National Crime Victimization Survey conducted by the US Justice Department, 68% of sexual assaults committed between 2008 and 2012 were left unreported. An earlier report by the US Bureau of Justice Statistics showed that the majority of rapes and sexual assaults perpetrated against females between 1992 and 2000 were not reported to the police. Only 36 percent of rapes, 34 percent of attempted rapes, and 26 percent of sexual assaults were reported[56]. The very nature of these crimes, which are often committed by family members, makes it unlikely that these statistics will improve.

The second reason is that there is a lack of clarity regarding what counts as a violent crime. For example, in some jurisdictions entering

---

53 HAMILTON, 2015, p. 40.

54 General recidivism is easier to predict, i.e. predictions can be made with higher statistical confidence, because the base rate of minor or moderately serious offenses is much higher than the base rate for violent or sexual crimes.

55 Different actuarial instruments use different methods to determine recidivism: convictions, arrests, probation/parole violations, or self-reports. This methodological difference makes them incommensurable.

56 RENNISON, 2002.

(...) there is no reliable information about the true recidivism rates for several violent offenses.

an empty house to commit burglary counts as a violent crime[57]. Counting such crimes as violent offenses offers a distorted picture of the base rate for violent recidivism. Furthermore, Quinsey et al. claim that the VRAG score is "positively related to the probability of at least one violent reoffense"[58]. However, without a more detailed definition of what counts as a violent reoffense, the VRAG score can generate widespread injustice: "Even an almost 100% probability that an offender, if released, will commit a simple assault within the next 10 years would not justify that offender's continued retention by any reasonable cost-benefit analysis"[59]. The uncertainty about recidivism rates does not allow an adequate analysis of the actuarial instrument, in particular, regarding the rate of false negatives and false positives. Without this information it is impossible to correctly assess the sensitivity and specificity of the actuarial instrument, and *a fortiori*, to calculate its error rate[60].

Perhaps the most problematic feature of most actuarial methods is that their actual success rate is very modest. Consider initially the study that provided the basis for the development of VRAG[61]. The study was conducted among 618 juvenile and adult male offenders who had been released from secure confinement after being treated in the maximum-security Oak Ridge psychiatric facility in Ontario or briefly assessed there prior to imprisonment. After 7 years their recidivism rate was determined. When the offenders were divided into high and low risk according to the 12 predictor variables that constitute VRAG, using a cutoff score of 7, they obtained the following results:

Perhaps the most problematic feature of most actuarial methods is that their actual success rate is very modest.

---

57 The US Supreme Court has regarded burglary as violent in some cases (*James v. United States*, 2007; *Taylor v. United States,* 1990), and as non-violent in others (*Solem v. Helm,* 1983; *Tennessee v. Garner*, 1985). It is counted as a violent crime under the Sentencing Guidelines of the United States (United States Sentencing Commission, 2013). The main argument for counting it as violent is that the potential for violence is always present. The classification is liable to being manipulated for political purposes, in particular, by fear-mongering politicians.

58 QUINSEY et al., 1998, p. 149.

59 LITWACK, 2001, p. 429.

60 Sensitivity and specificity are statistical measures of the performance of a classificatory instrument such as VRAG. The sensitivity of a risk assessment instrument is its ability to detect true positives, people classified as high risk who will recidivate, while its specificity is its ability to detect true negatives, people classified as low risk who will not recidivate. Technically, sensitivity is defined as the number of true positives divided by the total number of people in the population who will recidivate, including the false negatives: TP/(TP + FN). Specificity is defined as the number of true negatives divided by the total number of people in the population who will not recidivate, including the false positives: TN/(TN + FP).

61 HARRIS et al., 1993.

THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF
EXPERT CLINICAL AND ACTUARIAL TESTIMONY

Table 1: The basis for VRAG

	Recidivists	Nonrecidivists	Total
High Risk	114	94	208
Low Risk	77	333	410
Total	191	427	618

These results indicate that among those patients classified as "high risk", there was a recidivism rate of 55% (114/208), while the recidivism rate for those classified as "low risk" was 19% (77/410). The former result is very troubling. It indicates that the instrument was only slightly better than chance at predicting future dangerousness. Furthermore, if the resulting 94 false positives were placed in civil commitment, 15% of the total population would be unfairly committed. And since VRAG did not detect 40% of the total of recidivists (77/191), it failed to protect society by releasing a large number of future recidivists. These troubling results are reflected in the sensitivity and specificity of the instrument. Based only on this initial sample, VRAG has a very modest sensitivity as a predictor of violence (TP/(TP + FN) = 114/(114+77) = 0.6). Its specificity is acceptable (TN/(TN + FP) = 333/(333+ 94) = 0.78).

During the last three decades, VRAG and many other actuarial instruments have been used in many different populations with varying results. A complete discussion of the technical details regarding the statistical methods used to assess the predictive validity of VRAG and other instruments falls beyond the scope of this essay[62]. The following paragraphs can only offer a faint flavor of a highly technical discussion. Many defenders of actuarial instruments measure predictive accuracy using an index of sensitivity and specificity across score thresholds known as "area under the curve" (AUC). AUC values lie between 0 and 1, with 1 indicating perfect discriminatory ability and 0.5 indicating discriminatory ability no better than chance. In most studies, AUC values for actuarial instruments fall between 0.5 and 1[63], with the most popular of them

---

62 For a detailed technical discussion of the assessment of predictive validity, see HAMILTON (2015, pp. 23-35).

63 RETTENBERGER et al., 2010.

Many meta-analyses of these instruments have confirmed that they lead to an unacceptable number of false positives.

falling between 0.7 and 0.75[64]. What this means is that actuarial instruments have been able to classify violent and sexual recidivists at higher levels of risk than nonrecidivists about 70 to 75% of the time. Defenders of actuarial methods argue that these results represent moderate or large size effects. In statistics, an effect size is a measure of the strength of a phenomenon, such as the correlation between two variables. However, there is no consensus within the statistics community about the exact relationship between effect sizes and AUC values, and the former is often used inconsistently[65]. This lack of consensus within the statistics community weakens the claims of the defenders of actuarial methods.

It must be kept in mind that AUC is only an index of discrimination at the group level. It does not signify the probability that a particular individual was correctly classified: a high AUC value does not mean that a person classified as high risk will most likely become a recidivist. Furthermore, an actuarial instrument can have a high AUC even if the instrument is not well calibrated, that is, even if the percentage of predicted outcomes is significantly different than the proportion of actual outcomes[66]. Another way to put it is that AUC measures are not affected by the actual base rates of offending. Thus an instrument can have a high AUC and at the same time it can offer predictions that differ widely from the true rates of recidivism. Defenders of actuarial methods argue that the use of AUC values solves the low base rate problem discussed in previous paragraphs, but it is more adequate to say that it simply ignores the problem[67], with dire consequences: "Even with relatively high AUC values (e.g., 0.8), predictions of the occurrence of low base rate events (e.g., 10% base rate or lower) will almost always result in a very large number of false positives"[68].

Many meta-analyses of these instruments have confirmed that they lead to an unacceptable number of false positives[69]. One of those studies, which offered a systematic review and meta-analysis of 251

---

64 Singh et al, 2011. AUC for VRAG's initial sample was 0.72. This is not very impressive if we consider that a good AUC for a medical diagnosis test is normally above 0.8.

65 KellEY & PREACHER, 2012.

66 HAMILTON, 2015.

67 AMENTA et al., 2003.

68 ROSENFELD et al., 2011, p. 41.

69 SINGH et al., 2011.

validation studies of nine popular actuarial instruments —including VRAG, SORAG, Static 99, and PCL-R— offered the following results: Only 41% of people classified as moderate or high risk by *violence* risk assessment tools violently reoffended, 23% of people classified as moderate or high risk by *sexual* risk assessment tools sexually reoffended, and 52% of people classified as moderate or high risk by *generic* risk assessment tools went on to commit unspecified offenses[70]. On a brighter note, the instruments were able to identify low risk individuals with high levels of accuracy. The authors concluded: "One implication of these findings is that, even after 30 years of development, the view that violence, sexual, or criminal risk can be predicted in most cases is not evidence based. (…) [R]isk assessment tools in their current form can only be used to roughly classify individuals at the group level, and not to safely determine criminal prognosis in an individual case"[71].

4. UNFORESEEN CONSEQUENCES OF ACTUARIAL METHODS

A very *sensitive* statistical instrument is optimal when one prefers to obtain a much larger number of false positives than of false negatives, i.e., when one prefers that the number of undetected recidivist inmates be minimized, while paying the price of keeping many future nonrecidivists in prison. In contrast, a very *specific* instrument is optimal when one prefers to obtain a much larger number of false negatives than of false positives, i.e., when one prefers to reduce the number of imprisoned future nonrecidivists while taking the risk of releasing many future recidivists. The decision to adopt a very sensitive or a very specific instrument, or to adjust its parameters to make it more sensitive or specific, should be guided by a previous theoretical analysis regarding the price that we, as a society, are willing to pay in terms of civil liberties to guarantee our safety. It should also be informed by a theory of fair punishment. Unfortunately, the rush to adopt actuarial method has been guided by neither. As Harcourt points out:

---

70 FAZEL et al., 2012.

71 FAZEL et al., 2012, p. 5.

[T]he proliferation of actuarial methods has begun to bias our conception of just punishment. (...) [T]hese actuarial instruments represent nothing more than fortuitous advances in technical knowledge from disciplines, such as sociology and psychology, that have no normative stake in the criminal law. These technological advances are, in effect, exogenous shocks to our legal systems, and this raises very troubling questions about what theory of just punishment we would independently embrace and how it is, exactly, that we have allowed technical knowledge, somewhat arbitrarily, to dictate the path of justice[72].

A related point has to do with the cutoff points in the instrument's scoring system. Why is 7 instead of 8 the cutoff point for "high risk" in VRAG? Neither clinicians nor statistician have a commonly agreed definition of "high risk." In the case of VRAG, the cutoff point was chosen because it offered, in the authors' opinion, the best trade-off between sensitivity and specificity. But that trade-off was judged under the assumption that false positives and false negatives have "an equivalent cost"[73]. However, there is a long tradition in law that regards a false positive as much more costlier than a false negative. This tradition is captured by Blackstone's ratio ("the law holds it better that ten guilty persons escape, than that one innocent party suffer") and by the idea that justice must err on the side of innocence. Laudan has argued that "a standard of proof is best conceived as a mechanism for distributing errors. That, in turn, suggests that if we could figure out the relative cost to society of false convictions and false acquittals, we might be able to use the ratio of these costs as a mechanism for determining the height of the SoP"[74]. Given the consequences to the prisoner of being classified as high risk using an actuarial instrument, the cut-off point can also be regarded as a sort of standard of proof, and its determination must also be based on our societal conception of the cost of a false positive and a false negative.

Another related consequence of classifying individuals as high- or low-risk is that the cutoff point might distort the standard of proof used in trial. If the standard of proof is preponderance of the

---

72 HARCOURT, 2007, p. 3.

73 HARRIS et al., 1993, p. 329. In more recent times, Harris and Rice have added that "it can be reasonable for public policy to operate on the basis that a miss (e.g., failing to detain a violent recidivist beforehand) is twice as costly as a false alarm (e.g., detaining a violent offender who would not commit yet another violent offense)" (2013, p. 106).

74 LAUDAN, 2006, p. 68.

(...) there is a long tradition in law that regards a false positive as much more costlier than a false negative.

THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF EXPERT CLINICAL AND ACTUARIAL TESTIMONY

evidence, being classified as "high risk" will tilt the scales against the defendant, while under a standard of clear and convincing evidence, being classified as "low risk" will tilt them in favor of the defendant. Furthermore, there is a risk that actuarial instruments will usurp the factfinder's role. In a criminal trial, it is the judge's role to make judgments on the question of reliability and truthfulness of an actuarial instrument. In Allen's words, "there is no such thing as 'naked statistical evidence.' All evidence presented at trial will be tested by the factfinder's epistemology"[75]. The problem is that most factfinders do not have the required knowledge to undertake an epistemological analysis of actuarial instruments. To do so, they must understand the underlying statistical techniques, which are far from intuitive. In addition, in most cases there will be no more available evidence upon which to base a decision, other than additional clinical or actuarial expert testimony, and a factfinder might be tempted to simply rely on the score provided by the instrument, resulting in what Roberts and Zuckerman call "trial by scientific expert"[76]. Even when there is ancillary evidence, the use of actuarial evidence can generate an anchoring effect. An anchoring bias occurs when a person places too heavy a weight on a single piece of information. It is a cognitive heuristic that helps decisionmakers face complex judgments, but it is also liable to generate errors. If decisionmakers anchor their decisions on actuarial tools, the potential for error is thereby increased.

A further problem with the classificatory standards of actuarial tools is that some of them will classify an individual as "high risk" while others will classify the same individual as "low risk"[77]. The future of an individual therefore depends on the choice of actuarial instrument! The lack of convergence on the same cutoff point is further evidence of their inability to adequately track the rate of recidivism. Furthermore, it is not clear either why falling into the "high risk" category automatically corresponds to a decision of continued imprisonment. Why is a specific probability of recidivism associated with that decision? And how is that probability connected to the duration of a prolonged sentence?

------

[75] ALLEN, 1991, p. 1098.

[76] ROBERTS and ZUCKERMAN, 2010, p. 489.

[77] BARBAREE et al., 2006.

THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF EXPERT CLINICAL AND ACTUARIAL TESTIMONY

> **(...) actuarial instruments have not been designed to help us understand "low risk" individuals.**

Additionally, actuarial instruments have not been designed to help us understand "low risk" individuals. "Low risk" individuals are just those people who happen not to be "high risk". The instruments provide no information regarding the factors that diminish their propensity towards violent or sexual recidivism, and are therefore useless in helping policymakers evaluate how effective rehabilitation and socialization programs really are[78]. From a pragmatic point of view it is understandable that the instruments focus on factors that serve as predictors for recidivism, but decisions are being made about low risk individuals despite the fact that these instruments offer no information about them.

Finally, it seems epistemically unacceptable to use actuarial instruments developed in North America in other continents with very different populations. Some studies suggest that the results vary according to population. For example, VRAG works better in the UK and Canada than in the United States[79]. There is also evidence that the predictive validity of VRAG is gender dependent: it works better with male inmates[80]. Currently there are no actuarial risk assessment instruments designed specifically for use with female inmates. If the use of actuarial instruments cannot be avoided, they should at least be tailored to take into account the specific socioeconomic and clinical circumstances of female delinquents[81].

5. CONCLUSIONS

There are several lessons that can be learned from this dire landscape. To avoid the violation of due process generated by the use of clinical or actuarial predictions in judicial proceedings, one option is to convert *Daubert* into a constitutional principle, as suggested by Beecher-Monas[82]. The *Daubert* standard is only required in federal courts in the United States; its adoption at the state level, and in any kind of judicial proceeding, would go a long ways towards cutting the

---

[78] HAMILTON, 2014.

[79] YANG et al., 2010.

[80] HASTINGS et al., 2012.

[81] Interestingly, clinicians are also far worse at predicting future violent offenses committed by females. According to one study, their accuracy did not differ from chance. Apparently, their inaccuracy appeared to be a function of their underestimation of the base rate of violence among mentally disordered women (Lidz et al., 1993).

[82] BEECHER-MONAS, 2007.

ground from under that particular type of expert testimony. There is a risk, however, that courts will be willing to lower the *Daubert* standard for evidence that originates in the social sciences. In fact, this has already happened. The Texas Court of Criminal Appeals in *Nenno v. State* (1998) stated that "[w]hen addressing fields of study aside from the hard sciences, such as the social sciences or fields that are based primarily upon experience and training as opposed to the scientific method, [the law's] requirement of reliability applies but with less rigor than to the hard sciences"[83].

Perhaps the main response offered by defenders of actuarial methods when confronted with arguments such as the ones presented in this paper is that there is nothing better in the offing: "Demonstrably less accurate methods of risk appraisal are widely used instead [...] The only ethical course of action is to use the most accurate system available, even if it is imperfect[84]. But there is a more ethical course of action, and it is to stop using these methods *tout court*. One cannot but agree with Beecher-Monas when she says that the current state of affairs undermines the rule of law: "Admitting scientifically baseless expert testimony on future dangerousness into evidence is not only cynical, it also undermines law´s moral authority. The very least we can do in system that aspires to do justice is to be sure that the scientific testimony admitted in our courts has been tested, scrutinized, and properly limited"[85].

Eliminating the prediction of future dangerousness as a factor in sentencing and establishing instead the use of metrics based on the harm caused, the degree of responsibility of the accused, and previous convictions will be a more just course of action than using demonstrably flawed decision methods. The prediction of future dangerousness can also be eliminated as a factor in deciding parole. The only criteria available for parole decisions ought to be statistically solid factors such as an advanced age. The recidivism rate of rapists steadily decreases with age, and prisoners over 60 are very unlikely to commit violent or sexual crimes if released[86]. The risk of error in releasing such prisoners is so low that it is worth

---

83 *Nenno v. State*, p. 561 quoted by MONAHAN, 2000, p. 912.

84 QUINSEY et al., 1998, p. 176.

85 BEECHER-MONAS, 2007, p. 167.

86 HANSON, 2002.

taking, especially in light of the obvious benefit or resocializing an individual and freeing up sorely needed space to detain truly dangerous individuals. Furthermore, if rehabilitation programs that establish equivalences between work/study and sentence reduction are to have a real impact, their results must be a starting point in all bail decisions.

I would like to end by clarifying that none of the objections to actuarial methods presented in this paper appeal to what is often called the "G2i" problem: the intrinsic difficulties involved in the application of tools designed at the group level to individual cases[87]. I have no concerns regarding many of the applications in the individual case of conclusions derived from the general case when the price to be paid in terms of errors is socially or scientifically acceptable.[88] But actuarial instruments, as we have seen in this paper, demand an extremely high price in terms of false positives, and that by itself is sufficient reason to reject their use in the individual case, independently of whether, in general, statistical generalizations devoid of any causal support can be the basis for individualized legal decisions.

> **(...) none of the objections to actuarial methods presented in this paper appeal to what is often called the "G2i" problem: the intrinsic difficulties involved in the application of tools designed at the group level to individual cases.**

---

87 FAIGMAN et al., 2014.

88 See SCHAUER, 2003, for a general defense of the use of generalizations in legal decision-making. Schauer is careful to acknowledge that error rates that are scientifically acceptable need not be legally acceptable: "Science can tell us that a certain scientific process has, say, a 12 percent error rate (or specific rates of Type I and Type II errors or false positives and false negatives). And scientists must decide for their own scientific purposes whether such rates are sufficient, for example, to assert that something is the case, conclude that a finding is adequate for publication, or find a research program promising enough to renew a research grant. But whether such an error rate is sufficient for a trier of fact to hear it, put someone in jail, keep someone out of jail, justify an injunction, or award damages is not itself a scientific question" (2010, p. 1214).

THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF
EXPERT CLINICAL AND ACTUARIAL TESTIMONY

REFERENCES

AMENTA, Amy E.; GUY, Laura S.; EDENS, John F. Sex Offender Risk Assessment: A Cautionary Note Regarding Measures Attempting to Quantify Violence Risk. *Journal of Forensic Psychology Practice*, n. 3, p. 39-50, 2003.

AMERICAN PSYCHIATRIC ASSOCIATION. *Dangerous Sex Offenders: A Task Force Report of the American Psychiatric Association.* Washington: American Psychiatric Association, 1999.

AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and Statistical Manual of Mental Disorders DSM-5*. Washington: American Psychiatric Association, 2013.

ALLEN, Ronald. On the Significance of Batting Averages and Strikeout Totals: A Clarification of the "Naked Statistical Evidence" Debate, the Meaning of "Evidence," and the Requirement of Proof Beyond Reasonable Doubt. *Tulane Law Review,* n. 65, p. 1093-1110, 1991.

BAKER, Emily. The Adam Walsh Act: Un-Civil Commitment. *Hastings Constitutional Law Quarterly*, n. 37, p. 143-165, 2009.

BARBAREE, Howard; LANGTON, Calvin; PEACOCK, Edward. Different Actuarial Risk Measures Produce Different Risk Rankings for Sexual Offenders. *Sexual Abuse*, n. 18, p. 423-440, 2006.

BEECHER-MONAS, Erica. *Evaluating Scientific Evidence. An Interdisciplinary Framework for Intellectual Due Process*. New York: Cambridge University Press, 2007.

BRIEF for *Amicus Curiae of the American Psychiatric Association*, *Barefoot v. Estelle*, 1983 U.S. S. Ct. Briefs LEXIS 1529 (March 4, 1983) (No. 82-6080).

BUREAU OF PRISONS. Civil Commitment of a Sexually Dangerous Person. *Federal Register, 72*, p. 43205-43209, 2007.

EDITORIAL: Shock and Law [Editorial]. *Nature*, *490*, 446, 2012.

FABIAN, John. The Adam Walsh Child Protection and Safety Act: Legal and Psychological Aspects of the New Civil Commitment Law for Federal Sex Offenders. *Cleveland State Law Review*, n. 60, p. 307-364, 2012.

FAIGMAN, David; MONAHAN, John; SLOBOGIN, Christopher. Group to Individual (G2i) Inference in Scientific Expert Testimony. *University of Chicago Law Review*, 81, p. 417-480, 2014.

FAZEL, Seena et al. Use of Risk Assessment Tools to Predict Violence and Antisocial Behaviour in 73 Samples Involving 24827 People: Systematic Review and Meta-analysis. *BMJ*, 345, e4692, 2012.

FITCH, W. Lawrence. Sexual Offender Commitment in the United States: Legislative and Policy Concerns. *Annals of the New York Academy of Science*, n. 989, p. 489–501, 2003.

GOLDBERG, Lewis. Simple Models or Simple Processes: Some Research on Clinical Judgments. *American Psychologist*, n. 23, p. 482-496, 1968.

HAMILTON, Melissa. Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law. *Arizona State Law Journal*, n. 47, p. 1-62, 2015.

HANSON, Karl. Recidivism and Age. Follow-up Data from 4673 Sexual Offenders. *Journal of Interpersonal Violence*, n. 17, p. 1046-1062, 2002.

HARCOURT, Bernard. *Against Prediction: Profile, Policing, and Punishing in an Actuarial Age*. Chicago: University of Chicago Press, 2007.

HARE, Robert. *Manual for the Hare Psychopathy Checklist-Revised*. Toronto: Multi-Health Systems, 1991.

HARRIS, Andrew et al. *Static-99 Coding Rules: Revised 2003*. Solicitor General Canada, 2003.

HARRIS, Grant; RICE, Marnie; QUINSEY, Vernon. Violent Recidivism of Mentally Disordered Offenders. The Development of a Statistical Prediction Instrument. *Criminal Justice and Behavior*, n. 20, p. 315-335, 1993.

HARRIS, Grant; RICE, Marnie. Bayes and Base Rates: What Is an Informative Prior for Actuarial Violence Risk Assessment? *Behavioral Sciences & the Law*, n. 31, p 103-124, 2013.

HASTINGS, Mark et al. Predictive and Incremental Validity of the Violence Risk Appraisal Guide Scores with Male and Female Jail Inmates. *Psychological Assessments*, n. 23, p. 174-183, 2012.

HIGGINS, Nicola et al. Assessing Violence Risk in General Adult Psychiatry. *Psychiatry Bulletin*, n. 29, p. 131-133, 2005.

KELLEY, Ken; PREACHER, Kristopher. On Effect Size. *Psychological Methods*, n. 17, p. 137-152, 2012.

KHIROYA, Reena; WEAVER, Tim; MADEN, Tony. Use and Perceived Utility of Structured Violence Risk Assessment in English Medium Secure Forensic Units. *Psychiatry Bulletin*, n. 33, p. 129-32, 2009.

KRAUSS, Daniel; SALES, Bruce D. The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing. *Psychology, Public Policy and Law*, n. 7, p. 267-310, 2001.

LAUDAN, Larry. *Truth, Error, and Criminal Law. An Essay in Legal Epistemology*. New York: Cambridge University Press, 2006.

LIDZ, Charles W.; MULVEY, Edward Patrick; GARDNER, William. The Accuracy of Predictions of Violence to Others. *Journal of the American Medical Association*, n. 269, p. 1007-1011, 1993.

LITWACK, Thomas R. Actuarial versus Clinical Assessments of Dangerousness. *Psychology, Public Policy and Law*, n. 7, p. 409–443, 2001.

MONAHAN, John. Violence Risk Assessment: Scientific Validity and Evidentiary Admissibility. *Washington and Lee Law Review*, 57, 901-918, 2000.

MOSER, Charles; KLEINPLATZ, Peggy. DSM-IV-TR and the Paraphilias: An Argument for Removal. *Journal of Psychology & Human Sexuality*, n. 17, p. 91-109, 2005.

MOSSMAN, Douglas. Connecting Which Dots? Problems in Detecting Uncommon Events. In: Harris, Andrew; PAGÉ, Caroline (Eds.). *Sexual Homicides and Paraphilias: The Correctional Service of Canada's Experts Forum 2007*. Ottawa: Correctional Service of Canada, 2008.

MURRAY, Dominic. *Psychiatry in the Scientific Image*. Cambridge: MIT Press, 2006.

QUINSEY, Vernon L. Demographic and Clinical Variables Associated with Release from a Maximum Security Psychiatric Institution. *Criminal Justice and Behavior*, n. 6, p. 390-399, 1979.

_____. The Long-Term Management of the Mentally Disordered Offender. In HUCKER, Stephen J.; Webster, Christopher D.; BEN- ARON, Mark H. (Eds.). *Mental Disorder and Criminal Responsibility*. Toronto: Butterworths, 1981, p. 137-155.

QUINSEY, Vernon L.; AMBTMAN, Rudolf. Variables Affecting Psychiatrists' and Teachers' Assessments of the Dangerousness of Mentally Ill Offenders. *Journal of Consulting and Clinical Psychology*, n. 47, p. 353-362, 1979.

QUINSEY, Vernon et al. *Violent Offenders: Appraising and Managing Risk*. Washington: American Psychological Association, 1998.

REED, Geoffrey M. Toward ICD-11: Improving the Clinical Utility of WHO's International Classification of Mental Disorders. *Professional Psychology: Research and Practice*, n. 41, p. 457–464, 2010.

RENNISON, Callie M. *Rape and Sexual Assault: Reporting to Police and Medical Attention, 1992–2000.* Washington: U.S. Department of Justice, Bureau of Justice Statistics, 2002.

RETTENBERGER, Martin et al. Prospective Actuarial Risk Assessment: A Comparison of Five Risk Assessment Instruments

in Different Sexual Offender Subtypes. *International Journal of Offender Therapy and Comparative Criminology*, n. 54, p. 169-186, 2010.

ROBERTS, Paul; ZUCKERMAN, Adrian. *Criminal Evidence*. 2nd edition. Oxford: Oxford University Press, 2010.

ROSENFELD, Barry; EDENS, John; LOWMASTER, Sara. Measure Development in Forensic Psychology. In: ROSENFELD, Barry; PENROD, Stephen D. (Eds.). *Research Methods in Forensic Psychology*. Hoboken: Wiley, 2003, p. 26-42.

SCHAUER, Frederick. *Profiles, Probabilities and Stereotypes*. Cambridge: Harvard University Press, 2003.

_____. Can Bad Science Be Good Evidence? Neuroscience, Lie Detection, and Beyond. *Cornell Law Review*, n. 95, p. 1191-1219, 2010.

SCHINKA, John A.; SINES, Jacob O. Correlates of Accuracy in Personality Assessment. *Journal of Clinical Psychology*, n. 30, p. 374-377, 1974.

SINGH, Jay P., GRANN, Martin; FAZEL, Seena. A Comparative Study of Violence Risk Assessment Tools: A Systematic Review and Metaregression Analysis of 68 Studies Involving 25,980 Participants. *Clinical Psychology Review*, n. 31, p. 499-513, 2011.

SKEEM, Jennifer L.; MONAHAN, John. Current Directions in Violence Risk Assessment. *Current Directions in Psychological Science*, 20, 38-42, 2011.

UNITED STATES SENTENCING COMMISSION. *Guidelines Manual*, 2013. Disponível em: http://www.ussc.gov/Guidelines/2013_Guidelines/index.cfm. Acessado em: maio 2016.

WORLD HEALTH ORGANIZATION. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*, 2010. Disponível em: http://apps.who.int/classifications/icd10/browse/2016/en. Acessado em: maio 2016.

WIGGINS, Jerry S. *Personality and Prediction: Principles of Personality Assessment*. Reading: Addison-Wesley, 1973.

WRIGHT, Richard W. Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts. *Iowa Law Review*, n. 73, p. 1001-1077, 1988.

YANG, Min; WONG, Stephen C. P.; COID, Jeremy Weir. The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools. *Psychological Bulletin*, n. 136, p. 740–767, 2010.

ZONANA, Howard. Sexual Disorders: New and Expanded Proposals for the DSM-5 – Do we Need Them? *Journal of the American Academy of Psychiatry and the Law*, n. 39, p. 245–9, 2011.

Laws and Cases Cited

Adam Walsh Child Protection and Safety Act of 2006, Public Law No. 109-248 (2006).

Barefoot v. Estelle 463 US 880 (1983).

Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 US 579 (1993).

James v. United States 550 US 192 (2007).

Kansas v. Hendriks 521 US 346 (1997).

Kansas v. Crane 534 US 407 (2002).

Nenno v. State 979 S.W.2d 549 (Tex. Crim. App. 1998).

Schall v. Martin 467 US 253 (1984).

Solem v. Helm 463 US 277 (1983).

Tarasoff v. Regents of the University of California 17 Cal. 3d 425, 551 P.2d 334 (1976).

Taylor v. United States 495 US 575 (1990).

Tennessee v. Garner et.al. 471 US 1 (1985).

United States v. Comstock 560 US 128 (2010).

United States v. Salerno 481 US 739 (1987).

THE PREDICTION OF FUTURE BEHAVIOR: THE EMPTY PROMISES OF EXPERT CLINICAL AND ACTUARIAL TESTIMONY