



Estimating the Probability of Sexual Recidivism Among Men Charged or Convicted of Sexual Offences: Evidence-Based Guidance for Applied Evaluators

L. Maaïke Helmus¹

[1] *School of Criminology, Simon Fraser University, Vancouver, British Columbia, Canada.*

Sexual Offending: Theory, Research, and Prevention, 2021, Vol. 16, Article e4283, <https://doi.org/10.5964/sotrap.4283>

Received: 2020-09-01 • **Accepted:** 2021-04-20 • **Published (VoR):** 2021-06-15

Handling Editor: Martin Rettenberger, Centre for Criminology (Kriminologische Zentralstelle – KrimZ), Wiesbaden, Germany

Corresponding Author: L. Maaïke Helmus, 10322 Saywell Hall, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6. E-mail: lmaaïkehelmus@gmail.com

Abstract

Risk assessment is routinely applied in forensic decision-making. Although relative risk information from risk scales is robust across diverse samples and settings, estimates of the absolute probability of sexual recidivism are not. Nonetheless, absolute recidivism estimates are still necessary in some evaluations. This paper summarizes research and offers guidance on evidence-based practices for assessing the probability of recidivism, organized largely around questions commonly asked in court. Overall, estimating the probability of sexual recidivism is difficult and should be undertaken with humility and circumspection. That being said, research favours empirical-actuarial risk tools for this task, more structured scales, and the use of multiple scales. Professional overrides of risk scale results should not be used under any circumstances. Paradoxically, however, professional judgement is still required in some circumstances. Risk scales do not consider all relevant risk factors, but the added value of external risk factors reaches a point of diminishing returns and may or may not be incremental (or worse, can degrade accuracy). There are reasons actuarial risk scales may both underestimate recidivism (e.g., undetected offending, short follow-ups) and overestimate recidivism (e.g., inclusion of sex offences not of interest in some referral questions, data on declining crime and recidivism rates, newer studies demonstrating overestimation of recidivism). Given all these considerations and the need for humility, in the absence of exceptional circumstances, I would not deviate too far from empirical estimates.

Keywords

risk assessment, recidivism probability estimates, sexual offences, civil commitment



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License, CC BY 4.0](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

Highlights

- Evidence-based guidance for the difficult task of estimating the probability of sexual recidivism
- Research favours use of actuarial, structured, and multiple risk scales
- Discussion of the role and added value of professional judgement
- Review of ways risk assessment scales may both overestimate and underestimate recidivism

Virtually all decisions impacting people charged or convicted of a sexual offence should heavily consider their risk to reoffend; the risk principle of effective correctional practice tells us that greater reductions in recidivism will be achieved by prioritizing our treatment, supervision, and management resources proportionate to risk (Bonta & Andrews, 2017). For almost all decisions (e.g., resource allocation, treatment/supervision intensity), relative risk information is sufficient to inform such evidence-based decision-making (G. T. Harris et al., 2015; Helmus, 2018). This is useful, as it appears as though structured risk scales are robust in assessing relative risk (a.k.a., discrimination; for brief review, see Helmus, 2018).

In contrast, structured risk scales do not appear consistent across samples in identifying the absolute probability of recidivism (calibration; Helmus, 2018). Nonetheless, decision-makers have an affinity for them (Blais & Forth, 2014; Chevalier et al., 2015) and in some assessment contexts, such as civil commitment in the United States, they may even be legally required (Association for the Treatment of Sexual Abusers [ATSA] & Sex Offender Civil Commitment Programs Network [SOCCPN], 2015). As of 2015, 20 U.S. states and the District of Columbia have established some form of civil commitment; in roughly half of these states (ATSA & SOCCPN, 2015). The exact legislative criteria vary by state, but most have some form of criteria that imply an absolute threshold of risk (e.g., “likely” to reoffend), with some having fairly specific criteria in law or case law, such as “more likely than not,” which can be defined as a probability of recidivism that exceeds 50% (e.g., in Washington state, see *re Detention of Brooks*, 145 Wn.2d 275, 36 P.3d 1034, 2001).

It is ironic that one of the most liberty-restricting decisions (i.e., preventative involuntary detention post-sentence) is most strongly tied to the *least* reliable piece of information that risk scales can provide (recidivism probabilities). As I have previously argued, given the instability of recidivism estimates, evaluators should be “humble and circumspect when reporting absolute recidivism probabilities (if they report them at all), with a cautionary note indicating that research demonstrates that recidivism estimates may not generalize well across diverse samples and settings” (Helmus, 2018, p. 3).

Despite the limitations of absolute probability estimates, professionals conducting risk assessments are often still required to opine on this issue. Consequently, it is crucial

that evaluations are as evidence-based as possible, while always acknowledging the limitations of current evidence. Humility and circumspection are important principles, but additional guidance may also be helpful.

The purpose of this paper is to summarize research relevant to estimating the probability of sexual recidivism and to offer evidence-based guidance. This includes considering factors that may influence recidivism estimates derived from risk scales. Much of this paper overlaps with content I have submitted in court cases where I was asked to comment on the quality and evidence base of the risk assessment methodology employed by a forensic evaluator. Section headings are often titled and organized around the most common questions I receive from lawyers, judges, or jurors. I will preface the paper with some foundational discussion of sexual recidivism base rates. Then I will outline some empirical premises that should inform the basis of all risk assessment decisions (with a focus on evidence from sex offence risk assessment). Lastly, I will cover some considerations that should inform specific estimates of the probability of sexual reoffending.

What Is the Overall Rate of Sexual Offence Recidivism?

This is a seemingly simple question, but it lacks a precise answer. Setting aside the issue of undetected recidivism (to be discussed further below), the public generally believes recidivism rates are higher than the data suggest (Helmus, 2016; Krauss et al., 2018; Levenson et al., 2007). One of the earliest and most frequently cited meta-analyses (Hanson & Bussière, 1998) found a sexual recidivism rate of 13% among 23,393 men charged or convicted of sexual offences across 61 different studies, with an average follow-up period between 4-5 years. More recently, examining 7,225 men charged or convicted of sex offences across 20 diverse studies, Hanson, Harris, Letourneau, Helmus, and Thornton (2018) found 5-year sexual recidivism rates of 9%, 10-year rates of 13%, 15-year rates of 16%, 20-year rates of 18% and 25-year rates of 18.5%.

These long-term findings suggest two things. Firstly, long-term rates of recidivism (e.g., 20-25 years) are roughly double the rate of 5-year estimates, although the gap reduces as risk increases (Thornton et al., 2021). Secondly, recidivism is most likely to occur in the early years of release. The longer someone stays sex-offence free in the community, the less the likelihood of sexual recidivism. After 10-15 years of release, most individuals charged or convicted of sexual offences pose no more risk of sexual recidivism than people with a criminal history but who have no known sexual offences (for more detailed analyses and generation of lifetime residual risk calculations, see Thornton et al., 2021).

One of the complications in assessing sexual recidivism base rates from the above reviews is that the rates vary across studies. Obviously longer follow-ups are associated

with higher recidivism, but the variability extends beyond that. In addition to random chance, studies differ on so many different factors, such as where they came from (e.g., what country, what setting – such as prison or probation sample), how they define recidivism (e.g., charges, convictions, the types of criminal records they obtained), and what types of individuals were included in that sample (e.g., more high risk cases, or more low-risk cases).

Overall, previously cited meta-analyses and reviews are likely to overestimate detected recidivism rates. The reason for this is because we tend to conduct more research on higher risk samples. This is often due to practicalities: settings with higher risk individuals (e.g., prisons, lengthier and more intensive treatment programs, individuals under intensive supervision, screened for civil commitment) tend to collect more detailed and richer information, which is a better source of data for researchers. For example, the recidivism norms for Static-99R (Hanson et al., 2016) have samples of routine/representative cases, and samples of people preselected as high risk/needs, but there is no collection of samples preselected to be low risk. This is because little to no research is conducted on these types of samples; if anything, they are often diverted away from settings where we would obtain detailed risk assessments on them (notably, this is consistent with the risk principle; Bonta & Andrews, 2017).

So what are the base rates of recidivism if we examine studies that are more representative of the population of all individuals convicted of sexual offences? Restricting it to more representative samples (12 studies, $n = 7,244$), the 5-year sexual recidivism rate is closer to 7% (Lee & Hanson, 2021), rather than the 9-13% range presented above. More recent studies tend to find even lower rates. For example, a study of over 17,000 men convicted of sexual offences in Texas found that 4% were re-arrested for a sexual offence within 5 years (Boccaccini et al., 2017). Internationally, a 5-year sexual recidivism rate of 6% was found for a population-based sample of inmates in Austria (the proportion of individuals convicted for sexual offences who are not incarcerated is unknown; Rettenberger et al., 2015).

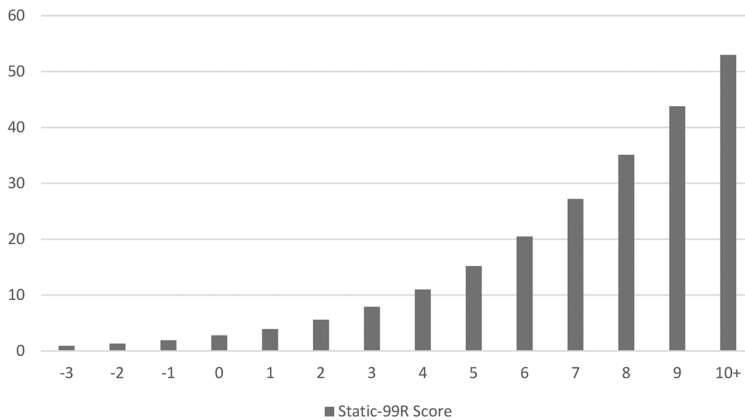
Consequently, answering the question “what is the recidivism rate” depends on a lot of factors. Our most precise way of estimating this is to consider things like the length of follow-up, and the risk level of the individual (while also ensuring we understand how most studies define and measure recidivism, which is discussed further below). Recidivism estimates from actuarial risk scales consider both follow-up length and the individual’s risk factors, and tend to provide the best estimates.

As an example, Figure 1 presents the 5-year recidivism rates based on their Static-99R (Helmus, Thornton, et al., 2012) scores across a collection of fairly routine and representative samples (Hanson, Thornton, et al., 2016). Here, recidivism refers to new charges or convictions for sexual offences. Higher Static-99R scores are associated with higher recidivism rates. Depending on their score, expected recidivism rates are as low as 1% and as high as 53%. Even with this breakdown, however, the recidivism estimates

per score vary significantly across different samples in ways that are not yet fully understood (Hanson, Thornton, et al., 2016). In other words, the probability of recidivism associated with a Static-99R score of 2 was not consistent across samples and it's hard to know why. Helmus (2009) found that some common methodological differences (e.g., year of release, type of offence, recidivism criteria) did not meaningfully explain the base rate variation. Although many factors likely contribute small amounts to the variation, Hanson, Thornton, et al. (2016) have concluded that the most likely explanation refers to the density of dynamic risk factors. Specifically, for individuals with a Static-99R score of 2, for example, higher recidivism rates will be found for individuals with higher levels of dynamic risk.

Figure 1

Five-Year Recidivism Estimates per Static-99R Score



Empirical Premises of Recidivism Risk Assessment

This section discusses four evidence-based principles for conducting risk assessment, including my recommendations within these topics.

1) The More Structured, the Better

Research finds that the most accurate approaches to predicting behaviour are empirical-actuarial. Following Meehl's (1954) definition, actuarial prediction scales are mechanical methods where the items are explicitly identified, with a clear algorithm for computing total scores, and probabilities of the outcome (e.g., recidivism) are provided

for each total score. [Hanson and Morton-Bourgon \(2009\)](#) have further clarified that empirical-actuarial scales are ones in which the scale items are supported by research.

There is a large and cross-disciplinary body of research regarding recidivism and many other outcomes, suggesting that mechanical prediction schemes outperform unstructured professional judgement ([Ægisdóttir et al., 2006](#); [Bonta et al., 1998](#); [Dawes et al., 1989](#); [Grove et al., 2000](#); [Hanson & Morton-Bourgon, 2009](#); [Mossman, 1994](#)). As I have discussed previously ([Lehmann et al., 2016](#)), this cross-disciplinary literature contradicts the intuitive belief that the expertise of professionals should be better equipped to handle complex situations and case-specific factors (e.g., [Boer et al., 1997](#)). Paradoxically, it appears to be simultaneously correct that although level of expertise matters in predicting many outcomes (e.g., experts generally outperform novices), actuarial decision algorithms outperform experts, but only under some conditions ([Kahneman & Klein, 2009](#); [Shanteau, 1992](#)). An important question, then, is under what conditions?

In summarizing the decision-making and cognitive science literature, [Shanteau \(1992\)](#) found evidence for good expert performance in weather forecasters, livestock judges, astronomers, test pilots, soil judges, chess masters, physicists, mathematicians, accountants, grain inspectors, photo interpreters, and insurance analysts. Poor professional judgements were noted for clinical psychologists, psychiatrists, student admissions evaluators, court judges, behavioural researchers, counselors, personnel selectors, parole officers, polygraph judges, intelligence analysts, and stock brokers. Mixed performance was found for nurses, physicians, and auditors. [Shanteau \(1992\)](#) proposed a variety of task features that were associated with poorer performance from experts. He concluded that human behaviour is inherently more unpredictable than physical phenomena, and that decision-making is particularly difficult for unique tasks, when feedback is unavailable, and when the environment is intolerant of error.

[Kahneman \(2011\)](#) provided a more updated summary of the performance of experts across a variety of tasks, with similar conclusions. According to [Kahneman and Klein \(2009\)](#), expert opinion can be expected to outperform actuarial decisions when the environment is regular (i.e., highly predictable), the expert has considerable practice, and there are opportunities to get timely feedback on decisions to learn from errors or false cues. These conditions are generally not present in recidivism risk assessment. The sheer number of diverse predictors of recidivism (e.g., see [Bonta & Andrews, 2017](#); [Hanson & Morton-Bourgon, 2005](#)) suggest that criminal behaviour is not highly predictable. The number of contingencies are infinite ([Hanson, 2009](#)) and evaluators typically receive no feedback on their decision, much less timely feedback. Additionally, risk assessment environments are arguably one of the most risk-averse and least tolerant of error.

Given these general and well-replicated findings supporting mechanical decision-making, I recommend using empirical-actuarial approaches, wherever valid scales exist. Additionally, the more structured, the better (e.g., scales with more detailed scoring rules). More structured scales will decrease (but not eliminate) variability in scores and

predictive accuracy across evaluators and settings and they will increase transparency and objectivity, all of which are beneficial for high stakes decisions. More structure in the scale's coding rules is particularly helpful in cases with extensive file information, where it becomes increasingly difficult for clinical judgement to focus on the most relevant information. As an example of more structure reducing variability, studies examining Structured Professional Judgement (SPJ) scales have found that variability in the accuracy of the scales is higher when the final assessment is presented as a professional opinion as opposed to a total score (Hanson & Morton-Bourgon, 2009; Helmus & Bourgon, 2011). Additionally, higher interrater reliability, which is often a byproduct of structured scoring guidelines, has been generally related to higher predictive accuracy (Hanson & Morton-Bourgon, 2009; Smid et al., 2014).

Additionally, actuarial risk scales provide more precise quantitative risk communication information than other methods, such as SPJ scales. For example, SPJ scales cannot provide empirically validated recidivism probability estimates, reducing their utility for civil commitment evaluations. As another example, research has found that Static-99R, which has a very structured and detailed coding manual, was much less susceptible to adversarial allegiance biases (i.e., scores that tend to favour the side that retained the evaluator) than the Psychopathy Checklist-Revised (PCL-R; Hare, 2003), which is structured, but still has much more open-ended and subjective scoring criteria (Murrrie et al., 2013).

There is one important caveat to this suggestion: This section presumes that the empirical-actuarial scales are being applied to appropriate populations (i.e., individuals broadly similar to the types of cases that were included in the development and validation research on the scale). If this precondition is not met, alternative risk assessment methods may be required. For example, there are currently no actuarial risk assessment tools validated for females who have sexually offended; consequently, professional judgement is the best option available.

2) More Scales Are Better (But the Best Way to Combine Them Isn't Clear)

No single risk assessment scale captures all relevant information, and no single scale has been consistently demonstrated to be superior to any other scale (Hanson & Morton-Bourgon, 2009, with the exception that the RRASOR is statistically inferior to several other scales). In other words, there are multiple scales available, with different strengths/weaknesses/purposes, with no clear winner.

A more comprehensive and balanced assessment would consider multiple risk assessment scales. Meta-analyses have found that dynamic risk scales add incrementally to static risk scales, and more specifically, Static-99R and STABLE-2007 add incrementally to each other (Brankley et al., 2021; Van den Berg et al., 2018). Even highly similar static risk scales for individuals charged or convicted of sexual offences add unique

information in predicting sexual recidivism (Babchishin et al., 2012; Lehmann et al., 2013; note that Seto, 2005 did not find incremental validity but this study was insufficiently powered).

In these studies, however, the incremental validity tended to be small in magnitude. Consequently, just because multiple scales significantly improve accuracy does not necessarily mean that the improvement provides sufficient practical gains in accuracy to warrant the added time involved in using multiple scales; this is a separate consideration that requires weighing the extra time/resources required, practical considerations, and the consequences of the decision both to individual liberty and public safety. Nonetheless, given these findings, Babchishin et al. (2012) recommended that for high stakes decisions such as civil commitment, *both* Static-99R and Static-2002R should be used. If different scales each predict the same recidivism outcome and measure at least some different things (or measure those things differently, or weight them differently), then we would expect value added in using multiple risk scales. Multiple scales converging with similar information should increase our confidence in the assessment results. Multiple scales providing meaningfully different results should encourage us to be more cautious in our assessments and necessitate thoughtful analysis and interpretation.

What is not yet fully established in the research literature is exactly *how* the results of different risk scales should be combined. Conceptual arguments could be made for methods such as trusting the highest risk estimate (i.e., being most cautious), trusting the lowest (i.e., giving the individual the benefit of the doubt), and trying to identify the best scale (although research demonstrates that there likely is no “best” scale). Assuming all scales are relatively equal, an averaging approach makes the most sense psychometrically; each scale assesses risk with some error, and averaging them should provide a better, more reliable assessment. This assumption, however, cannot be taken for granted.

For combining Static-99R and Static-2002R, the findings of Babchishin et al. (2012) broadly supported an averaging approach but without specific guidance on how to do so; the Lehmann et al. (2013) study found the best approach was from averaging (their method involved averaging risk ratios). The averaging approach, however, is unlikely to be ideal in some situations, such as one scale meaningfully predicting better than another¹, one scale having lower quality recidivism estimates, or a greater mismatch between the individual being assessed and the validation research (e.g., a person >60 years old assessed on a scale that does not do a good job accounting for age, an Indigenous individual assessed with multiple risk scales demonstrating considerable variability in cross-cultural validity)

1) For example, in the case of static risk factors substantially outpredicting dynamic risk factors, a scale that combines the two without giving much less weight to the dynamic factors may lead to degradation in accuracy (Helmus et al., 2019).

3) Professional Overrides Degrade Accuracy

As I have previously discussed (Helmus & Quinsey, 2020), it is well known (and moreover, obvious) that actuarial risk scales do not include all relevant/possible risk factors. Does that mean that predictive accuracy could be improved by adjusting actuarial results based on external, empirical risk factors not included in the scale? This seems intuitively defensible, but research has consistently found that overriding actuarial results degrades their predictive accuracy. In their meta-analysis, Hanson and Morton-Bourgon (2009) found three studies showing degradations in predictive accuracy when clinicians were allowed to override risk scale results based on what they considered to be important for that case.

In a subsequent and impressively large study, Wormith et al. (2012) examined the accuracy of the Level of Service/Case Management Inventory (LS/CMI, one of the most frequently used actuarial risk scales for general recidivism; see Bonta & Andrews, 2017) in predicting general, violent, and sexual recidivism among individuals convicted for non-sex offences ($N > 24,000$) and sexual offences ($N > 1,900$). The results were the same for both offence types and all three recidivism outcomes: the overrides degraded the accuracy of the scale. Interestingly, the degradations were consistently more pronounced for individuals convicted of sexual offences. Similar results were found for juveniles convicted of sexual and non-sexual offences, with the largest degradations in predictive accuracy found for overrides when trying to predict sexual recidivism (Schmidt et al., 2016). Even when staff in one jurisdiction were trained on the research regarding overrides and requested to use them judiciously (e.g., no more than 5% of cases), predictive accuracy was still lower when overrides were applied, compared to the original actuarial estimates (Guay & Parent, 2018).

A further study on overrides examined 441 individuals released from Minnesota prisons in 2012 (Duwe & Rocque, 2018). Corrections staff scored cases on the MnSOST-3. Additionally, an End of Confinement Review Committee assigned a risk level. This multi-stakeholder committee consists of the prison warden, a law enforcement officer, a sex offender treatment professional, a caseworker experienced in supervising individuals convicted of sex offences, and a victim services professional. This process is unique in that risk level decisions are made jointly by this diverse committee representing different stakeholders in risk management for those convicted of sexual offences. They considered actuarial risk assessment scales combined with other information they considered relevant. The risk level determined solely by the MnSOST-3 risk assessment scale was strongly related to sexual rearrests. However, the risk level determined by consensus from this multi-disciplinary committee showed little to no relationship with sexual rearrest and was substantially worse than the results of the risk scale.

If we know that actuarial risk scales do not incorporate all relevant information, then why don't overrides work? There are many potential reasons for this. Although some evaluators may be capable of identifying empirically-supported risk factors external to

the scale, they may not be good at integrating this information. Perhaps the additional risk factors already correlate with existing information in the scale such that the new information does not add incremental accuracy (i.e., does not provide any meaningful improvement in accuracy above the information already considered). Or perhaps they overweight the new information – given how much is already included in the risk scales, additional relevant factors may make only a tiny additional contribution to what is already measured, in most cases not changing the individual's risk category.

The ability to override results may exacerbate a punitive or risk-averse tendency in evaluators, which is seen in findings that evaluators are far more likely to use overrides to increase than to decrease an individual's risk score (Hanson et al., 2015; Schmidt et al., 2016; Wormith et al., 2012), although in one study where staff used overrides judiciously, they were used roughly equally in both directions (Guay & Parent, 2018). Furthermore, Wormith et al. (2012) found preliminary evidence that, unlike overrides to increase risk, overrides to decrease risk may have merit, although so few cases were overridden down to make strong conclusions. This suggests that evaluators may not be good at assessing circumstances that would increase risk, but they may be able to assess circumstances that decrease risk.

Additionally, where subjectivity is increased, so is the potential for bias. Consequently, it is probable that overrides are influenced at least to some extent by factors irrelevant to recidivism risk. This may also explain why degradation in accuracy is strongest for people convicted of sexual offences, where there is a strong negative emotional reaction to their offences (colloquially referred to as the “ick factor”). This “ick factor” would explain why evaluators were more likely to upgrade risk excessively for individuals convicted of sex offences compared to non-sexual offences (Wormith et al., 2012). In addition to increasing the potential for bias, subjectivity also increases unreliability. With so many possible reasons to override, it is not surprising that Hanson et al. (2007) found near-chance levels of agreement in override decisions ($ICC = .15$). This alone could explain the lack of predictive value of overrides; if evaluators cannot agree on when and why to override, the overrides are unlikely to predict outcomes.

Regardless of the reason for the degradation in accuracy, the research findings are clear that actuarial results should not be adjusted. These findings are also consistent with the broader literature in other areas of decision-making involving predictions of human behaviour, whereby mechanical prediction approaches outperform professional judgement.

4) A Paradox: Professional Judgement Degrades Accuracy, But Is Still Necessary

Paradoxically, even though research supports mechanical approaches to prediction over professional judgement, and that professional overrides degrade the accuracy of actuarial risk scales, this does not remove professional judgement and expertise from the task

of risk assessment. As discussed above, risk assessment scales do not incorporate all relevant information. Additionally, rarely is a risk assessment scale specifically designed to assess the referral question in a legal case, particularly when it comes to civil commitment.

So where does this leave the conscientious evaluator who knows that all relevant information is not included in the risk scale, but trying to adjust the scale will only make things worse? It is important to remember that a risk assessment scale is one piece of information that is used in case management decisions. It is an important piece of information, but it is not the sole piece of information. As noted by Hanson (2009), “scoring an actuarial risk tool is not a risk assessment” (p. 174).

Even when actuarial scales are used, professional expertise is needed to pull everything together in the risk assessment. It is needed to determine whether any unique factors exist that may reduce confidence in the application of a particular risk scale with a particular case. For example, for perhaps the person is not representative of the samples used to develop/validate the scale, or there is an important construct not measured in the scale that has demonstrated incremental accuracy in predicting the relevant outcome. Professionals must synthesize and interpret discrepant results of risk scales, which may involve considering the amount of empirical support, the purpose, and the construct validity and composition of the various scales. Lastly, professionals must make extrapolations and form an overall opinion given that risk scales rarely provide a yes/no answer to the specific referral question (e.g., particularly in civil commitment, there are differences in the outcome relevant to the legal criteria compared to the outcome being predicted by the scale).

For example, a risk assessment scale like Static-99R can tell you (with important limitations) the probability of an individual with this score being charged or convicted of any new sexual offence (including non-contact sex offences) within five years. And this information should not be altered via professional judgement. However, this is not the same as indicating the lifetime probability of an individual committing (detected or undetected) a new predatory act of sexual violence (if that is the criterion of civil commitment). Consequently, some professional judgement is required in using information from actuarial risk scales to inform the specific referral question. Given these considerations and the general findings supporting the higher accuracy for mechanical prediction approaches, I recommend that empirical-actuarial risk scales should be given the most weight in predicting likelihood of recidivism, and that professional judgement needs to be incorporated, but sparingly and with humility, and separately from the actuarial results.

How Much Should Factors Outside Structured Risk Scales Influence Overall Conclusions About Recidivism Probability?

Given that no structured risk scales consider *all* relevant risk factors, overrides degrade accuracy, but some professional judgement is still required, I am often asked how much consideration should be given to external, evidence-based risk factors. The challenging part in forming an overall evaluation of risk is to understand how much this external information should influence your overall evaluation. My basic opinion is absent the truly exceptional case (i.e., the person who tells you his plan to reoffend; severe mobility impairments), some but not a lot. Here, it's important to add some comments on incremental validity, which refers to how much the new information adds to existing information in predicting recidivism.

We know that many risk factors on their own (i.e., if that's the *only* risk factor you consider) have a small to moderate effect in predicting recidivism (e.g., [Hanson & Morton-Bourgon, 2005](#)). By combining risk factors together into an overall scale, you can get moderate to large predictive accuracy. But risk factors are not purely additive. They tend to be related to each other, so someone who has one risk factor (e.g., a risk factor not included in the risk scale you're using), likely has other related risk factors, which likely will be captured in existing risk scales ([Brouillette-Alarie & Hanson, 2015](#)). In other words, the unique value of a risk factor above and beyond risk scales is generally much smaller than the value of that information if that was the *only* thing you considered (i.e., you did not use any risk scales).

Consequently, adding more risk factors will reach a point of diminishing returns. That's generally why most risk scales stop after including 10-20 items. So even a factor that predicts well on its own and is not included in a risk assessment scale may add little to nothing in predicting recidivism after you account for the risk scale. This is probably part of why overrides don't work: because without structured scales developed from research, people do not have an easy, intuitive way to consider the unique, added value of a new risk factor above and beyond all the risk factors they have already considered in their risk assessment scales, and so they tend to give too much weight to the external information.

Example: Number of Sex Offence Victims

The most common example I see in risk assessment reports is evaluators opining that the risk scale underestimates the individual's risk because it considered their charges or convictions for sex offences but did not take into account a large number of sex offence victims, including unadjudicated victims. This argument makes a number of assumptions: a) that having more sex offence victims than what is reflected by charges or convictions is unusual or noteworthy; b) that the number of victims predicts recidivism; c) that it is not adequately addressed by other methods of accounting for criminal history; and d) that it adds incrementally to existing risk scales. I have seen cases where

it seems the number of victims is given equal (if not more) weight than the risk scale results in the evaluator's final opinion of risk, but looking at these assumptions it seems the threshold for a factor like this to appreciably change the risk assessment results is actually quite high.

It is not uncommon for individuals with a history of sex offending to have more victims than their criminal record captures. Findings vary but individuals convicted of sexual offences typically report on average, between 2-10 times as many sex offence victims as their official criminal records would indicate (Ahlmeier et al., 2000; Emerick & Dutton, 1993; Groth et al., 1982; Weinrott & Saylor, 1991). In a recent study, the number of sex offence victims was significantly related to sexual recidivism (Stephens et al., 2018), although it is unknown whether this variable would add to predictive accuracy above and beyond existing risk scales and factors, and if so, how much.

Whether victim count adds incrementally to existing risk scales has not been directly tested, but this relates to the issue of whether it is an optimal method of measuring criminal history. In analyses leading to the development of Static-2002R, Hanson and Thornton (2003) found that the number of offences in the current sentencing occasion for sex offences was not related to sexual recidivism. In other words, looking at the most recent sex offence charges/convictions, individuals who were convicted of multiple offences were not more likely to reoffend than individuals convicted of one offence. When capturing sexual offending history, the number of sentencing occasions was slightly more predictive than the number of charges or the number of convictions (where the latter two would be more sensitive to the number of offences committed or the number of victims). This suggests that when predicting recidivism, it is better to count the number of sentencing occasions rather than the number of offences/victims/charges.

This result may seem counter-intuitive. The number of victims is not included in most risk scales and it feels important. One possible reason for this finding is that sexual offending behaviour might be very different for people who have not been detected versus those who have been detected by the criminal justice system. People are more likely to continue bad behaviour if they are not experiencing consequences for it. From that perspective, the number of occasions sanctioned would be far more informative about risk than the number of victims, particularly the number of victims before being caught for the first time.

This example demonstrates that a seemingly intuitive position ("this scale underestimates risk because it does not take into account these undetected victims") rests on a number of assumptions which could contribute to degradations in accuracy. That being said, of course further research could always suggest improvements and structured ways to consider this information. For example, although it can be presumed that people convicted of sexual offences have more victims than what is noted on their criminal record, perhaps there is some threshold for victim count that really does override risk scale results. Where this threshold resides, however, is unknown and would need to be

empirically investigated. Overall, this example illustrates the potential danger of placing much emphasis on risk factors not incorporated in risk scales.

How Good Are Absolute Recidivism Estimates From Actuarial Scales? Do They Over-Estimate or Under-Estimate Recidivism?

Empirically derived recidivism estimates from empirical-actuarial risk scales are the strongest and most defensible way of making predictions about an individual's likelihood of reoffending. That being said, however, this does not mean that recidivism estimates are without error. It is important to acknowledge the strengths and limitations of the recidivism estimates and any assessment measure used ([American Psychological Association, 2013](#)), consider how the estimates map on to the outcome of interest for a particular case or jurisdiction, and to consider the likely direction of external factors (e.g., would they result in the recidivism estimates over-estimating or under-estimating the outcome of interest?).

Importantly, research has consistently found that although risk assessment scales are fairly consistent across diverse samples and settings in terms of ranking individuals according to their relative risk to reoffend, the exact probability estimates of recidivism tend not to generalize as well ([Hanson, Thornton, et al., 2016](#); [Helmus, Hanson, et al., 2012](#); [Helmus & Thornton, 2016](#); [Mills et al., 2005](#); [Olver et al., 2014](#); [Snowden et al., 2007](#)). Often, there are empirically based reasons why recidivism estimates from existing actuarial risk scales may be too low in predicting this outcome, but also other reasons that they may be too high.

Reasons Why Actuarial Recidivism Estimates May Be Too Low

Underreporting

This is an obvious one and almost universally (and appropriately) mentioned. All recidivism estimates from evidence-based risk tools are based on new charges and/or convictions for a sexual offence and they will be an underestimate of true recidivism because sexual offences are notoriously underreported. The challenge, however, is that we do not know how much this underreporting might influence officially detected recidivism estimates.

Combined across many victimization surveys, a reasonable estimate is that approximately 10% of sex crimes may be reported to police (for review, see [Hanson et al., 2003](#)), although this information is becoming dated and reporting rates are likely to change over time (e.g., after the #MeToo movement). Taking 10% as a heuristic, however, that does not mean that real recidivism rates are 10 times higher than observed recidivism rates. There are several reasons for this.

Firstly, over 95% of individuals arrested for sexual offences have no detected sex offence history (Sandler et al., 2008), so it is hard to generalize from overall rates of reported crime to a small subset of already-detected individuals. In other words, those already detected and convicted for sex offences might be much more likely to be caught for future offences than perpetrators who are unknown to the criminal justice system. Additionally, many individuals who reoffend will commit multiple sex offences, increasing their likelihood of being detected overall. They may not get caught for their first or second victim, but if they keep reoffending, they will likely be caught eventually, regardless of low detection rates. This means that the detection rate per offence is lower than the detection rate per individual (whereby the detection rate per individual is influenced by the number of victims they have, as well as other factors).

Consequently, understanding how underreporting will impact recidivism estimates requires considering reporting rates and the average number of victims per offender, as well as potential differences between individuals who haven't been caught versus those who have. An additional complication is that higher risk offenders are more likely to be higher-frequency offenders, which means that they may be more likely to be eventually detected for recidivism. And yet another complication is that some types of offences have higher detection rates (e.g., violent stranger rapes or attempted rapes versus exhibitionism or incest; Hanson et al., 2003).

One example of research that attempts to quantify this effect is a 2003 conference presentation from Hanson and colleagues. Examining empirically derived estimates of plausible detection rates as well as offence frequency rates, depending on an individual's risk level, their probability of being detected for a sexual recidivism incident could be as low as 5% (in which case observed recidivism rates would severely underestimate real recidivism) or as high as 90% (in which case the underestimation may only be very small).

More recently, Scurich and John (2019) explored undetected sex offending with a variety of different statistical modeling assumptions and suggested that the gap between detected and real sexual offending may be quite large. Unfortunately, there is insufficient evidence to evaluate many of the assumptions of their models (e.g., Kelley, 2019; note that some of these limitations would be shared in the Hanson et al., 2003 estimates as well). For example, Scurich and John's (2019) analyses presume that reporting rates, successful prosecution rates, frequency of offending rates, etc., do not meaningfully differ for individuals who have never been caught vs those already known to the system. Their modeling also assumed these factors would be similar across risk levels and offence types (e.g., contact versus non-contact offences), which is unlikely. Their models also assumed 100% of allegations are substantiated. While false reports are hopefully uncommon, they do exist.

Their analyses also assumed frequency of offending per individual does not change over time, which contradicts key literature on desistance (Hanson, 2018). Much of the

recidivism research they relied upon came from out-of-date and non-representative samples. Some of these timing issues may be critically important as detection rates could change considerably over time with cultural shifts. For example, Canada has witnessed increases in sex offending reporting rates in the 1980s (along with increases in the legal definition of what constitutes a sex offence; Kong et al., 2003) and again since the #MeToo movement started in 2017 (Rotenberg & Cotter, 2018).

Where does this leave us? Scurich and John's (2019) models indicate that the gap between detected and undetected offending varies considerably based on these unknown parameters. So in other words, we really don't know with confidence what this gap is until we develop better research on the underlying parameters. This research should be sensitive to time (i.e., length of follow-up, changes in sex offending rates and detection over time), case features (type of offences, frequency of offending), and differences between individuals never detected vs detected (and for the latter group, differences prior to and post-detection). For example, based on the parameters used by Scurich and John (2019), for an overall recidivism base rate of 15% after 5 years, estimates of actual reoffending range as low as 19% to as high as 82%. Based on the limitations of their analyses, the gap is unlikely to be as large as they fear. But there is still much unknown. Given the large number of factors and assumptions influencing these types of estimates, we do not yet have a sufficient empirical basis to provide reliable structured extrapolations to undetected recidivism. But we can identify certain principles, such as for those who offend with higher frequency, the gaps between detected and undetected offending will be smaller. In other words, high rates of underreporting do not necessarily mean high rates of undetected recidivists; it is possible that many unreported offences are attributed to already-detected recidivists as well as perpetrators completely unknown to the criminal justice system.

Shortage of Long-Term Follow-Up Studies

Not surprising, it is more difficult and time consuming to conduct long-term recidivism studies. Consequently, most actuarial risk assessment scales provide recidivism estimates for 5 and/or 10 year follow-up periods. If you were to factor in longer follow-ups (e.g., 20 years, lifetime recidivism), the cumulative recidivism rates would increase, but it is difficult to know exactly by how much (and similar to accounting for undetected offending, this pattern would likely differ based on risk levels). For people convicted of all types of offences, including sex offences, they are most likely to reoffend shortly after release. The longer they are able to remain offence-free, the less likely recidivism becomes (Blumstein & Nakamura, 2009; Bushway et al., 2011; Hanson et al., 2014; Howard, 2011). A recent and important advance in this area has been the development of statistical models to calculate lifetime and residual estimates of sexual recidivism based on Static-99R (Thornton et al., 2021).

Reasons Why Actuarial Recidivism Estimates May Be Too High

Type of Sexual Offences Included (e.g., Non-Contact)

Most current risk assessment scales included non-contact sexual offences in their outcome, which may not be part of some referral questions. For example, indeterminate sentences may be based on risk for contact sexual offences, violent sexual offences, or even more narrow terms such as “predatory acts of sexual violence” (e.g., Washington state RCW 71.09.020(18)). Consequently, the outcome in existing risk scales may overestimate of the types of offences in the referral. For example, in a representative sample of released inmates from Austria, five-year sexual recidivism rates decreased from 6% to 4% when only contact offences were considered (Rettenberger et al., 2015).

Crime Rates and Recidivism Rates Are Declining Over Time

There is considerable research suggesting that crime rates and recidivism rates are declining over time. This is true of all crimes, as well as sexual offences specifically (for a review, see Helmus, 2009). To obtain 10-year recidivism estimates, any research must be examining individuals released at least a decade ago. However, given how long it takes to collect and publish data, most current 10-year recidivism studies contain cases released at least 15 years ago. Given declining crime and recidivism rates, this may mean that research studies perpetually overestimate current recidivism rates.

Some Studies Are Finding That Static-99R Overestimates Recidivism

This is specific to Static-99R (I have yet to see similar research conducted on other scales) and could be partly related to the declining recidivism rates over time or may reflect other factors such as a shortage of large recidivism studies from the United States in the existing Static-99R recidivism norms. Either way, there are two new large field studies from the United States, both of which are finding lower recidivism rates per Static-99R score compared to the estimates provided by the routine correctional samples. In a study of 1,626 individuals convicted of sexual offences in California (Lee et al., 2016) and in another study of 17,455 cases released from prison in Texas (Boccaccini et al., 2017), both studies found that the observed 5-year recidivism rates were approximately half of what the Static-99R routine norms would have predicted. A more recent study from Minnesota (Duwe & Rocque, 2018) with roughly 500 individuals convicted of sexual offences also found sexual recidivism rates were less than half of what would be predicted by Static-99R, although they used a 4-year follow-up instead of 5 years. This overestimation of recidivism is quite substantial. One thing that is unclear is how well the recidivism data in these large studies would capture sexual recidivism. For example, if individuals are charged with violent offences for sexually motivated behaviour, or if their parole is breached for a new sexual offence but the criminal record only indicates “parole revocation,” then these types of sexual recidivism would not be captured. These

considerations may mean that the overestimation is not as bad as what was observed, but nonetheless, these are still three credible studies that suggest that the routine Static-99R norms may be overestimating recidivism in more recent samples in the United States.

Conclusions

Estimating the probability of sexual recidivism is difficult. Empirical-actuarial risk assessment scales are currently the only tools that provide evidence-based recidivism probability estimates, but they are not without their limitations. In particular, they demonstrate significant variability across samples. Nonetheless, they are the best estimate we have. For evaluators who need to comment on the absolute probability of sexual recidivism, research favours empirical-actuarial risk scales over professional judgement. Research also favours more structured scales and the use of multiple scales. Professional overrides of risk scale results should not be used under any circumstances. Nonetheless, there are still additional factors not included in the risk scales, and there is never a perfect match between the risk scale data and the specific referral question in a given evaluation. Consequently, evaluators still need to exercise some judgement, and consider ways in which actuarial recidivism estimates may both overestimate and underestimate recidivism.

Actuarial recidivism estimates will necessarily underestimate recidivism because they cannot account for undetected offending, and they also generally do not provide long-term (e.g., 15+ years) recidivism estimates, although Static-99R now has lifetime estimates. However, the amount of underestimation is difficult to quantify and may not be as high as some would think. Additionally, there is also some evidence to suggest that they may overestimate officially recorded sexual recidivism in modern samples (particularly from the United States), and they would be overestimates of the types of offences typically of interest for civil commitment.

Given considerations on both sides of this equation, risk scale estimates are likely a plausible mid-range estimate. In the absence of exceptional circumstances, in an overall professional opinion, I would not place the estimated likelihood of sexual recidivism too far from the range of estimates provided by actuarial risk scales. How far is “too far” is difficult to say, and it depends on many factors. I think anything more than plus or minus 10 percentage points from the empirical estimate would require a pretty strong justification. Such an approach is also consistent with my previous recommendation to comment on recidivism probabilities with humility and circumspection (Helmus, 2018).

Funding: The author has no funding to report.

Acknowledgments: This work was completed on the traditional and unceded territories of the Coast Salish Peoples (where the city of Burnaby currently resides), specifically the Squamish, Tsleil-Waututh, Musqueam, and Kwikwetlem Peoples. I would like to thank David Thornton for helpful comments on an early draft of this paper.

Competing Interests: Although I am a co-developer of some risk tools, I do not receive royalties from the use of the tools (although I am occasionally paid for training or consultation).

References

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.
<https://doi.org/10.1177/0011000005285875>
- Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse, 12*, 123-138.
<https://doi.org/10.1177/107906320001200204>
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *The American Psychologist, 68*, 7-19. <https://doi.org/10.1037/a0029889>
- Association for the Treatment of Sexual Abusers, & Sex Offender Civil Commitment Programs Network. (2015). *Civil commitment of sexual offenders: Introduction and overview*.
<https://www.atsa.com/sites/default/files/%5bCivil%20Commitment%5d%20Overview.pdf>
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment, 19*, 442-461.
<https://doi.org/10.1177/1073191112458312>
- Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior, 38*, 531-543. <https://doi.org/10.1037/lhb0000082>
- Blumstein, A., & Nakamura, K. (2009). Redemption in the presence of widespread criminal background checks. *Criminology, 47*, 327-359. <https://doi.org/10.1111/j.1745-9125.2009.00155.x>
- Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual offenders. *Psychological Assessment, 29*, 611-623. <https://doi.org/10.1037/pas0000377>
- Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: Guidelines for clinical practice. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 326-342). New York, NY, USA: Guilford Press.

- Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.) Abingdon, United Kingdom: Routledge.
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, *123*, 123-142. <https://doi.org/10.1037/0033-2909.123.2.123>
- Brankley, A. E., Babchishin, K. M., & Hanson, R. K. (2021). STABLE-2007 demonstrates predictive and incremental validity in assessing risk-relevant propensities for sexual offending: A meta-analysis. *Sexual Abuse*, *33*(1), 34-62. <https://doi.org/10.1177/1079063219871572>
- Brouillette-Alarie, S., & Hanson, R. K. (2015). Comparaison de deux mesures d'évaluation du risque de récidive des délinquants sexuels. *Canadian Journal of Behavioural Science*, *47*(4), 292-304. <https://doi.org/10.1037/cbs0000019>
- Bushway, S. D., Nieuwebeerta, P., & Blokland, A. (2011). The predictive value of criminal background checks: Do age and criminal history affect time to redemption? *Criminology*, *49*, 27-60. <https://doi.org/10.1111/j.1745-9125.2010.00217.x>
- Chevalier, C. S., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior*, *39*, 209-218. <https://doi.org/10.1037/lhb0000114>
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674. <https://doi.org/10.1126/science.2648573>
- Duwe, G., & Rocque, M. (2018). The home-field advantage and the perils of professional judgment: Evaluating the performance of the Static-99R and the MnSOST-3 in predicting sexual recidivism. *Law and Human Behavior*, *42*(3), 269-279. <https://doi.org/10.1037/lhb0000277>
- Emerick, R. L., & Dutton, W. A. (1993). The effect of polygraphy on the self-report of adolescent sex offenders: Implications for risk assessment. *Annals of Sex Research*, *6*, 83-103. <https://doi.org/10.1007/BF00849301>
- Groth, A. N., Longo, R. E., & McFadin, J. B. (1982). Undetected recidivism among rapists and child molesters. *Crime and Delinquency*, *28*(3), 450-458. <https://doi.org/10.1177/001112878202800305>
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30. <https://doi.org/10.1037/1040-3590.12.1.19>
- Guay, J.-P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior*, *45*, 82-100. <https://doi.org/10.1177/0093854817719482>
- Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology*, *50*, 172-182. <https://doi.org/10.1037/a0015726>
- Hanson, R. K. (2018). Long-term recidivism studies show that desistance is the norm. *Criminal Justice and Behavior*, *45*, 1340-1346. <https://doi.org/10.1177/0093854818793382>
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*(2), 348-362. <https://doi.org/10.1037/0022-006X.66.2.348>

- Hanson, R. K., Harris, A. J. R., Helmus, L., & Thornton, D. (2014). High risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence, 29*, 2792-2813.
<https://doi.org/10.1177/0886260514526062>
- Hanson, R. K., Harris, A. J. R., Letourneau, E., Helmus, L. M., & Thornton, D. (2018). Reductions in risk based on time offense free in the community: Once a sex offender, not always a sex offender. *Psychology, Public Policy, and Law, 24*, 48-63. <https://doi.org/10.1037/law0000135>
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (User Report No. 2007-05). Public Safety Canada.
<http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/index-eng.aspx>
- Hanson, R. K., Helmus, L. M., & Harris, A. J. R. (2015). Assessing the risk and needs of supervised sexual offenders: A prospective study using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior, 42*, 1205-1224. <https://doi.org/10.1177/0093854815602094>
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163. <https://doi.org/10.1037/0022-006X.73.6.1154>
- 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*, 1-21.
<https://doi.org/10.1037/a0014421>
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002* (Corrections Research User Report No. 2003-01). Ottawa, ON, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K., Thornton, D., Helmus, L. M., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sexual Abuse, 28*, 218-252.
<https://doi.org/10.1177/1079063215574710>
- Hanson, R. K., Thornton, D., & Price, S. (2003). *Estimating sexual recidivism rates: Observed and undetected*. Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, St. Louis, MI, USA.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised technical manual* (2nd ed.). Toronto, Canada: Multi-Health Systems.
- Harris, G. T., Lowenkamp, C. T., & Hilton, N. Z. (2015). Evidence for risk estimate precision: Implications for individual risk communication. *Behavioral Sciences & the Law, 33*(1), 111-127.
<https://doi.org/10.1002/bsl.2158>
- Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MR58443)
- Helmus, L. M. (2016). What does the general public know and want to know about sex offenders? [Newsletter]. *The Forum, 28*(2).
- Helmus, L. M. (2018). Sex offender risk assessment: Where are we and where are we going? *Current Psychiatry Reports, 20*, Article 46. <https://doi.org/10.1007/s11920-018-0909-8>

- Helmus, L., & Bourgon, G. (2011). Taking stock of 15 years of research on the Spousal Assault Risk Assessment guide (SARA): A critical review. *International Journal of Forensic Mental Health, 10*, 64-75. <https://doi.org/10.1080/14999013.2010.551709>
- 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 Behavior, 39*, 1148-1171. <https://doi.org/10.1177/0093854812443648>
- Helmus, L. M., Johnson, S., & Harris, A. J. R. (2019). Developing and validating a tool to predict placements in administrative segregation: Predictive accuracy with inmates, including Indigenous and female inmates. *Psychology, Public Policy, and Law, 25*(4), 284-302. <https://doi.org/10.1037/law0000201>
- Helmus, L. M., & Quinsey, V. L. (2020). Predicting violent reoffending with the VRAG-R: Overview, controversies, and future directions for actuarial risk scales. In J. S. Wormith, L. A. Craig, & T. Hogue (Eds.), *What works in violence risk management: Theory, research and practice* (pp. 119-144). Chichester, United Kingdom: Wiley.
- Helmus, L., & Thornton, D. (2016). The MATS-1 risk assessment scale: Summary of methodological concerns and an empirical validation. *Sexual Abuse, 28*, 160-186. <https://doi.org/10.1177/1079063214529801>
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse, 24*, 64-101. <https://doi.org/10.1177/1079063211409951>
- Howard, P. (2011). *Hazards of different types of reoffending* (Ministry of Justice Research Series 3/11). London, United Kingdom: UK Ministry of Justice.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY, USA: Macmillan.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *The American Psychologist, 64*, 515-526. <https://doi.org/10.1037/a0016755>
- Kelley, S. M. (2019, July). *How much sexual offending goes undetected?* Paper presented at the 9th Biennial conference of the Australia and New Zealand Association for the Treatment of Sexual Abuse, Brisbane, Australia.
- Kong, R., Johnson, H., Beattie, S., & Cardillo, A. (2003). Sexual offences in Canada (85-002-XIE). *Juristat, 23*(6). Canadian Centre for Justice Statistics. <https://www150.statcan.gc.ca/n1/pub/85-002-x/85-002-x2003006-eng.pdf>
- Krauss, D. A., Cook, G. I., & Klapatch, L. (2018). Risk assessment communication difficulties: An empirical examination of the effects of categorical versus probabilistic risk communication in sexually violent predator decisions. *Behavioral Sciences & the Law, 36*(5), 532-553. <https://doi.org/10.1002/bsl.2379>
- Lee, S. C., & Hanson, R. K. (2021). Updated 5-year and new 10-year sexual recidivism rate norms for Static-99R with routine/complete samples. *Law and Human Behavior, 45*(1), 24-38. <https://doi.org/10.1037/lhb0000436>

- Lee, S. C., Restrepo, A., Satariano, A., & Hanson, R. K. (2016). *The predictive validity of Static-99R for sex offenders in California: 2016 update*. State Authorized Risk Assessment Tools for Sex Offenders (SARATSO).
http://saratso.org/pdf/ThePredictiveValidity_of_Static_99R_forSexualOffenders_inCalifornia_2016v1.pdf
- Lehmann, R. J. B., Fernandez, Y., & Helmus, L. M. (2016). Strengths of actuarial risk assessment. In D. R. Laws & W. O'Donohue (Eds.), *Treatment of sex offenders: Strengths and weaknesses in assessment and intervention* (pp. 45-81). Cham, Switzerland: Springer.
- Lehmann, R. J. B., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K.-P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment, 25*, 1019-1024. <https://doi.org/10.1037/a0033098>
- Levenson, J. S., Brannon, Y. N., Fortney, T., & Baker, J. (2007). Public perceptions about sex offenders and community protection policies. *Analyses of Social Issues and Public Policy (ASAP), 7*(1), 137-161. <https://doi.org/10.1111/j.1530-2415.2007.00119.x>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN, USA: University of Minnesota Press.
- Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice and Behavior, 32*, 565-585.
<https://doi.org/10.1177/0093854805278417>
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783-792. <https://doi.org/10.1037/0022-006X.62.4.783>
- Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, 24*, 1889-1897.
<https://doi.org/10.1177/0956797613481812>
- Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale – Sexual Offender version. *Sexual Abuse, 26*, 472-499. <https://doi.org/10.1177/1079063213502679>
- Rettenberger, M., Briken, P., Turner, D., & Eher, R. (2015). Sexual offender recidivism among a population-based prison sample. *International Journal of Offender Therapy and Comparative Criminology, 59*(4), 424-444. <https://doi.org/10.1177/0306624X13516732>
- Rotenberg, C., & Cotter, A. (2018, November). Police-reported sexual assaults in Canada before and after #MeToo, 2016 and 2017. *Juristat*. Canadian Centre for Justice Statistics.
<https://www150.statcan.gc.ca/n1/pub/85-002-x/2018001/article/54979-eng.htm>
- Sandler, J. C., Freeman, N. J., & Socia, K. M. (2008). Does a watched pot boil? A time-series analysis of New York State's sex offender registration and notification law. *Psychology, Public Policy, and Law, 14*, 284-302. <https://doi.org/10.1037/a0013881>
- Schmidt, F., Sinclair, S. M., & Thomasdóttir, S. (2016). Predictive validity of the Youth Level of Service/Case Management Inventory with youth who have committed sexual and non-sexual offences. *Criminal Justice and Behavior, 43*, 413-430. <https://doi.org/10.1177/0093854815603389>

- Scurich, N., & John, R. S. (2019). The dark figure of sexual recidivism. *Behavioral Sciences & the Law*, 37, 158-175. <https://doi.org/10.1002/bsl.2400>
- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, 17(2), 156-167. <https://doi.org/10.1037/1040-3590.17.2.156>
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, 252-266. [https://doi.org/10.1016/0749-5978\(92\)90064-E](https://doi.org/10.1016/0749-5978(92)90064-E)
- Smid, W. J., Kamphuis, J. H., Wever, E. C., & van Beek, D. J. (2014). A comparison of the predictive properties of nine sex offender risk assessment instruments. *Psychological Assessment*, 26(3), 691-703. <https://doi.org/10.1037/a0036616>
- Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, 37, 1539-1549. <https://doi.org/10.1017/S0033291707000876>
- Stephens, S., Seto, M. C., Goodwill, A. M., & Cantor, J. M. (2018). The relationship between victim age, gender, and relationship polymorphism and sexual recidivism. *Sexual Abuse*, 30(2), 132-146. <https://doi.org/10.1177/1079063216630983>
- Thornton, D., Hanson, R. K., Kelley, S. M., & Mundt, J. C. (2021). Estimating lifetime and residual risk for individuals who remain sexual offense free in the community: Practical applications. *Sexual Abuse*, 33(1), 3-33. <https://doi.org/10.1177/1079063219871573>
- Van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, 30(2), 179-191. <https://doi.org/10.1037/pas0000454>
- Weinrott, M. R., & Saylor, M. (1991). Self-report of crimes committed by sex offenders. *Journal of Interpersonal Violence*, 6, 286-300. <https://doi.org/10.1177/088626091006003002>
- Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511-1538. <https://doi.org/10.1177/0093854812455741>



Sexual Offending: Theory, Research, and Prevention (SOTRAP) is the official journal of the International Association for the Treatment of Sexual Offenders (IATSO).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.