

A case of the pot calling the kettle black: A rejoinder to Quinsey's canard

Brian R. Abbott
Independent Practice

[Sexual Offender Treatment, Volume 14 (2019), Issue 2]

Abstract

In the Sexual Offender Treatment Journal, Quinsey (2017) responded to an article written by Abbott (2017), published in the International Journal of Law and Psychiatry, regarding the application of the Violence Risk Appraisal Guide- Revised ("VRAG-R") in civil confinement psychological evaluations and legal proceedings for sexually violent predators in the United States. Quinsey's criticisms are based largely on questionable presentation and interpretation of source material. Quinsey offers no substantive information or data that would lead me to alter my analysis, findings, conclusions, or recommendations (Abbott, 2017). The VRAG-R, at its current state of development, is not an appropriate actuarial measure to use in SVP civil confinement risk assessments.

Keywords: sexually violent predators, actuarial risk assessment, VRAG-R, actuarial prediction

I authored an article published in the International Journal of Law and Psychiatry (Abbott, 2017) that laid out a detailed legal, psychological, and statistical analysis related to the application of the Violence Risk Appraisal Guide-Revised (VRAG-R) to address the legally defined threshold of sexual recidivism proscribed by sexually violent predator ("SVP") laws in the United States. The examination focused on two main areas, including the fit of the VRAG-R standardization sample with those individuals being prosecuted as SVPs and the fit of the hands-on sexual and nonsexual violent recidivism predicted by the VRAG-R in comparison to the more narrowly defined legal threshold of sexual recidivism as contained in SVP statutes. The analysis led me to conclude that the VRAG-R was ill-suited for addressing the SVP legally defined threshold for sexual recidivism ("likely threshold") and forensic evaluators would be hard pressed to justify its use under generally accepted standards of psychological practice.

In response to my article, Quinsey posted a draft paper online in July 2017 (<http://www.vrag-r.org/wp-content/uploads/2017/07/Comment-on-Abbott-July-25.pdf>) that he revised into a working paper in August 2017, also posted online (https://www.researchgate.net/publication/319141638_The_Art_of_Misdirection_Abbott's_Shaky_Take_on_the_VRAG-R), which he modified once again as published by Sexual Offender Treatment (Quinsey, 2017). I direct my rejoinder to Quinsey (2017). Upon first read, Quinsey's response evoked reminders of a climate where contrary points of view are not debated on their merits. Upon further reflection, Quinsey's ad hominem response reveals how he employs the very tactic of which he accuses me (i.e., misdirection and misinformation) when attempting to rehabilitate the VRAG-R as an appropriate instrument to address the likely threshold as defined by SVP statutes.

Misstatements

In the first of the many misstatements in his response, Quinsey (2017) asserts that my analysis focuses on the application of the VRAG-R in Washington State SVP trials. This is simply false. In

the introduction to Abbott (2017), I express my intent to examine the application of the VRAG-R as applied to the twenty-two jurisdictions in the United States with SVP laws. Based on this false premise, Quinsey (2017) repeatedly references information from the state of Washington in an attempt to rebut or criticize my conclusions. This is nothing more than a distraction from the real issues.

It is interesting that Quinsey (2017) finds fault that the abstract of the article does not reference a citation as to declining base rates of sexual recidivism and how this makes it increasingly difficult to substantiate the likely threshold. What he fails to mention, however, is that I offer attribution for my conclusion in section 1.3 of the article as it relates to the Static-99R, which is the customary actuarial instrument used in SVP risk assessments (Jackson, Travia, & Schneider, 2009). I did not specify the actual rate of decline but will take the opportunity to do so here.

The sexual recidivism base rate for the original Static-99 actuarial sample (Hanson & Thornton, 2000) was 18% over five years. The five-year base rate of sexual recidivism observed in the Static-99R routine corrections sample is 8.3% (Hanson, Thornton, Helmus, & Babchishin, 2016), which represents a 57% reduction in the observed sexual reoffense rate. Similarly, the Minnesota sexual offenders comprising the MnSOST-3 actuarial sample (Duwe & Freske, 2012) demonstrated an approximate two-thirds decline in the observed sexual recidivism rate as compared to the MnSOST-R standardization sample (Epperson, Kaul, Huot, Hesselton, Alexander, & Goldman, 2003). As I delineated in my article (Abbott, 2017), score-wise sexual recidivism rates in actuarial measures decline at a rate that is roughly equal to the proportional decrease in the base rate. Consequently, base rate differences between the actuarial sample and the group representing the individual being assessed is a critical consideration when forensic evaluators are concerned about the accuracy of prediction.

The Case for Misdirection

Quinsey (2017) argues early in his response that a probabilistic term describing risk, such as high, is meaningless without associating a probability estimate with it. I agree with Quinsey on this point, and it is consistent with the introduction to Abbott (2017; section 1.3), where I describe how SVP laws define the risk threshold using probabilistic terms such as likely, more likely than not, or highly probable (hereinafter referred to as "likely"). For this reason, the score-wise probability estimates contained in sexual recidivism actuarial tables have been widely recognized as providing the best fit when evaluating whether the individual being assessed meets the likely threshold set by statute (Scurich, 2018; Abbott, 2017; Scurich & Krauss, 2014; Abbott, 2013; Donaldson & Abbott, 2011; Woodworth & Kadane, 2004; Janus & Prentky, 2003).

But, Quinsey (2017) reverses his position when arguing against the point I made regarding the importance of calibration, in general, and, in particular, how calibration studies lacked for the VRAG-R. In section 2.8 of my article, I describe how calibration studies are important because we cannot assume that the score-wise risk estimates contained in actuarial tables, like the VRAG-R, are accurate when applied to offenders from a group different from the actuarial standardization sample. For example, assume a forensic evaluator assesses an individual who is a member of a sexual offender group with an 8.0% base rate of sexual recidivism over five years. The evaluator decides to use the VRAG-R to assess the individual where the five-year base rate for violent reoffense is 32%. As a rule of thumb, it would be expected that the VRAG-R would inflate recidivism risk by a magnitude of four times. Rather than to address this issue, Quinsey (2017) argues the wholly unrelated point of selection accuracy.

Quinsey (2017) engages in obfuscation by confusing selection accuracy (i.e., ROC statistic) with

accuracy of score-wise probability estimates. While the former is useful when selecting a risk assessment tool, it is the latter that a forensic psychologist considers when evaluating whether an individual being prosecuted as an SVP meets the legally defined likely threshold. The ROC appraises discrimination performance across measures at various thresholds and this is not relevant evidence in forensic proceedings (Howard, 2017). The comparison of selection accuracy between the VRAG-R and other violent recidivism actuarial instruments does not assist the psychologist or trier of fact to understand the accuracy of the VRAG-R bin-wise risk estimate in predicting whether the individual being assessed meets the likely threshold. Conspicuously absent from Quinsey's contention is any reference to statistical or legal authorities where a probabilistic threshold of risk can be answered in a binary fashion; not to mention such an argument is illogical statistically and inconsistent with his prior assertion about the fit of probability estimates when assessing probabilistic statements of risk. The ROC does not inform about the accuracy of the probability of risk at each score on an instrument (Singh, 2013). Rather, the point estimate and the confidence interval about it, or the E/O analysis (Hanson, 2017) are examples of the type of statistical information to ascertain the accuracy of score-wise probability estimates.

The data produced by the two by two contingency table advocated by Quinsey (2017) can be used to compute selection accuracy statistics such as sensitivity and specificity. The use of a contingency table is not appropriate way to assess prognostic accuracy of risk assessment measures (Helmus & Babchishin, 2017). Rather, the point estimate at the actuarial score assigned to the individual and the 95% confidence interval about it informs the psychologist and the trier of fact about the accuracy of the prediction as it relates to the individual being assessed. As I explain (Abbott, 2017), the VRAG-R fails to provide an estimate of error about the bin-wise probability estimates, so it is not possible to understand the accuracy of the VRAG-R's probability estimates. It is my view that the absence of confidence intervals prevents psychologists from complying with standards of practice to consider and report the error associated with assessment procedures and to qualify their opinions accordingly (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014; American Psychological Association ["APA"], 2013; APA, 2010).

It is intuitively apparent that a forensic evaluator ultimately arrives at a binary decision about whether the individual meets the likely threshold. This decision; however, is the product of careful consideration of whether the probability estimate from the actuarial table, at the score assigned to the individual, satisfies the likely threshold as defined by law and whether the evaluator has sufficient confidence about the decision when considering the 95% confidence interval. The VRAG-R fails in these respects.

At the time that I wrote my article, no VRAG-R calibration studies had been published. Quinsey (2017) references two studies presented at conferences by Olver (2017, October) and by Gregório-Hertz, Rettenberger, and Eher (2016). He contends the results for both studies confirm that the VRAG-R violent recidivism probability estimates at each risk bin are accurate when applied to the offenders from the two study samples. Since Quinsey (2017) was published in *Sexual Offender Treatment*, both papers have since been published (Gregório-Hertz, Eher, Etzler, & Rettenberger, 2019; Olver & Sewall, 2018). The data reported by Olver and Seawall (2018) and Gregório-Hertz et al. (2019) establish calibration of the VRAG-R bin-wise violent recidivism rates, but as I explain below, the results appear to present meaningful forensic implications when assessing the risk for sexual reoffense among individuals undergoing SVP civil confinement proceedings. Both studies tested calibration using the expected to observed ("E/O") index as presented by Hanson (2017). Before addressing the results of the two studies, I first provide a brief primer on the E/O index as foundation for the reader to understand my conclusions.

The expected value (E) is a function of multiplying the total bin-wise sample from the comparison study by the corresponding proportion of recidivists from the VRAG-R standardization sample, as reported by Harris, Rice, Quinsey, and Cormier (2015) in Table C.1, at page 298. The observed value (O) represents the number of actual recidivists from the comparison study at the each VRAG-R bin. For each VRAG-R bin, E is divided by O to produce the E/O index. The E/O value of 1.0 indicates perfect calibration. Values more than 1.0 represent that the VRAG-R overestimates violent recidivism in the comparison sample. Values less than 1.0 denote that the VRAG-R underestimates violent recidivism in the comparison sample. A 95% confidence interval around the E/O index is computed using the formula from Hanson (2017). A confidence interval without 1.0 means the expected to observed values are significantly different (i.e., does not support calibration) at $p < .05$ (Hanson, 2017).

Gregório-Hertz et al. (2019) examined 534 male sex offenders released from Austrian prisons. They present the E/O analysis in Table 8 of their article. They reported expected proportions of violent recidivists by each VRAG-R risk bin (Cf. Table 8, column 3) that differ from those reported by Harris et al. (2015). I recomputed the E/O indexes exchanging the bin-wise proportions of violent offenders from Harris et al. (2015) for those expected rates reported in Table 8, column 3 of Gregório-Hertz et al. (2019). standardization sample (Harris et al., 2015) for the expected proportion of violent recidivists. I entered the data into an excel spreadsheet that computes the E/O index and the 95% confidence interval as specified by Hanson (2017). The recomputed results were not substantially different except at risk bins 2 and 3. The analysis reported by Gregório-Hertz et al. (2019) indicated that the VRAG-R expected rate of violent reoffense was lower than the observed rate of violent recidivism among the Austrian offenders ($E/O = .96$ and $.69$, respectively), whereas the recomputed values demonstrated that the VRAG-R expected rate of violent recidivism was greater than the observed rate among the Austrian group ($E/O = 2.31$ and 1.23 , respectively). The recomputed E/O indexes reflected that the VRAG-R over-estimated sexual reoffense among the Austrian sexual offenders at the VRAG-R risk bins of 2 and 3. Nonetheless, the E/O indexes were not considered significantly different and this supports that the bin-wise violent recidivism rates from the VRAG-R developmental sample (Harris et al., 2015) calibrated with the Austrian group.

The Olver (2017, October) conference presentation, which Quinsey (2017) relied upon, examined the use of the VRAG-R among a group of Canadian sexual offenders who were treated in a prison-based sexual offender treatment program known as Clearwater. The E/O analysis did not compare the violent recidivism experience between the VRAG-R standardization sample and the Clearwater sample. The findings reported by Olver (2017, October), in my opinion, did not support the assertion by Quinsey (2017) that the violent recidivism rates from the VRAG-R standardization sample calibrated with the violent recidivism rated observed in the Clearwater sample. Since the conference presentation in 2017, Olver and Sewall (2018) presented a more comprehensive presentation of their data in a peer reviewed journal, including an E/O analysis comparing the VRAG-R standardization sample to 295 Canadian sexual offenders from the Clearwater sample. I recomputed the E/O analysis by using the expected proportions of violent recidivism reported for the VRAG-R (Harris et al., 2015).

I used the following methodology when conducting the E/O analysis using data from Olver and Seawall (2018). I obtained the total number of Clearwater sexual offenders and the corresponding number of observed violent recidivists over five years at each VRAG-R risk bin from Table 4, panel 1, at page 754 from Olver and Seawall (2018). I relied upon the current bin-wise proportions of violent recidivists over five years from the VRAG-R standardization sample (Harris et al., 2015) for the expected proportion of violent recidivists. I entered the data into an excel spreadsheet that computes the E/O index and the 95% confidence interval as specified by Hanson (2017). Table 1 presents the results. The notation "R+" represents the term recidivists, the notation "E" refers to

expected number of recidivists, and "O" specifies the observed number of recidivists.

Table 1: E/O Analysis of Five-Year Violent Recidivism Rates between the VRAG-R Standardization Sample and Sexual Offenders from the Clearwater Study

VRAG-R Risk Bin	A Proportion of VRAG-R R+	B Total N from Clearwater	A X B E	Clearwater N of R+ O	E/O	95% Confidence Interval	
						Lower	Upper
1	.09	3	0.3	0	-	-	-
2	.12	5	0.6	0	-	-	-
3	.16	18	2.7	1	2.70	0.38	19.17
4	.20	15	3.0	1	3.00	0.42	21.30
5	.26	45	11.7	10	1.17	0.63	2.17
6	.34	27	9.2	6	1.53	0.69	3.41
7	.45	24	10.8	7	1.54	0.74	3.24
8	.58	38	22.0	17	1.30	0.81	2.09
9	.76	120	91.2	66	1.38	1.09	1.76

Since no Clearwater sexual offenders reoffended violently in risk bins 1 and 2, the E/O analysis could not be performed. The data for risk bin 9 indicates that the results from the VRAG-R standardization sample overestimated violent recidivism in the Clearwater sample at a magnitude of 1.38 times. This difference is statistically significant ($p < .05$) because the 95% confidence interval does not contain 1.0. The 95% confidence intervals for remaining risk bins (3-8) contain the value of 1.0, which indicates that observed number of recidivists from Clearwater were not different from the VRAG-R standardization sample to a statistically significant level, although there are noteworthy trends in the analysis that I address next. The expected violent recidivism rates from the VRAG-R standardization sample at risk bins 3 and 4 overestimate violent recidivism in the Clearwater sample by a magnitude of 2.7 and 3.0 times, respectively. Despite the lack of statistically significant results, the wide 95% confidence intervals at the scores of 3 and 4, in my view, raise uncertainty about the accuracy of the VRAG-R probability estimates. The wide variation in the 95% confidence intervals at risk bins 3 and 4 indicate the VRAG-R violent recidivism rates may be about 40% of what is observed in the Clearwater population, may overestimate violent recidivism on the order of twenty times, or fall anywhere between those values. The wide confidence interval results from having a single Clearwater recidivist in each risk bin. This effect results from the width of the 95% confidence interval in the E/O analysis being inversely related to the number of observed recidivists (Hanson, 2017). As the number of observed recidivists decrease, the width of the confidence interval increases and vice versa. The elevated E/O index and extreme variation in the confidence intervals, in my opinion, raises meaningful uncertainty about the accuracy of the E/O index at risk bins 3 and 4 as applied to the Clearwater sample.

Risk bin 5 from the Clearwater sample ($E/O = 1.17$) appears to display the best fit between the expected and observed number of violent recidivists. The E/O index values at risk bins 6-8, while not dissimilar at a statistically significant level, may have meaningful practical implications when rendering forensic decisions about whether individuals meet the likely threshold. A statistically nonsignificant finding may still portend consequential forensic differences (Donaldson, Abbott, & Michie, 2012). An example illustrates this point. The likely threshold in the state of Washington is considered as greater than 50% to commit predatory acts of sexual violence. Assume for the sake of discussion that the individual being assessed in the state of Washington is a member of a population of sexual offenders who demonstrate the same bin-wise frequency of violent recidivists as the Clearwater sample. Further assume, that five-year violent recidivism rate reported in the VRAG-R standardization sample reliably evaluates the commission of predatory acts of sexual violence under Washington law. Based on these assumptions, the forensic examiner might conclude the likely threshold is satisfied because the individual fell into the VRAG-R risk bin 8 where the standardization sample reoffended violently at the rate of 58% within five years after release in the community. But, if the forensic examiner divided the probability estimate of .58 from the VRAG-R standardization at risk bin 8 by the corresponding E/O index of 1.3, it would reveal a five-year violent recidivism rate of 44.6%. The violent recidivism rate of 44.6% is below the likely threshold and reveals the expert arrived at a false positive conclusion. Therefore, in my opinion, it is critically important for forensic evaluators who rely upon studies reporting the E/O index to be cognizant that the absence of statistically significant differences between the expected and observed number of violent recidivists among two groups of offenders does not mean the findings do not have practical significance in formulating forensic opinions about an individual.

It is likely that continued research with the VRAG-R and other samples of sexual offenders will establish the calibration of the VRAG-R violent recidivism rates to some degree. This alone, in my view, is insufficient to claim the VRAG-R is an appropriate instrument to use in SVP risk assessments for two reasons. One, as described previously, calibration established by the E/O index may still produce forensically meaningful results where the VRAG-R score-wise probabilities may overestimate or underestimate violent recidivism when assessing individuals against the likely threshold. In my view, forensic evaluators should take this potential source of error into account when rendering risk assessment conclusions and qualify opinions as necessary. Two, the developers of the VRAG-R have not overcome the burden of proving that the violent recidivism estimates at each VRAG-R risk bin are an accurate fit to the legally defined threshold of sexual reoffense under SVP statutes, which I addressed in detail (Abbott, 2017).

Red Herring

Quinsey (2017) and I agree on another point when Quinsey states, "Of course, the VRAG-R was not designed for SVP hearings and the base rates of SVP populations may indeed differ from those of the VRAG-R construction sample, the question is whether they are the best available estimate (because an estimate must be made) and whether they are likely to be higher or lower." We diverge as to the direction of the difference.

I lay out a principled statistical and conceptual presentation in sections 2.5 - 2.9 of my article (Abbott, 2017) that explains how the VRAG-R misestimates risk as defined by the likely threshold in three areas, including base rates, counting all nonsexual violent offenses as sexually motivated, and the inclusion of offenders who did not commit sexual offenses at all or who did not have a sexual offense for their most recent criminal conviction. Quinsey (2017) only addresses the base rate issue but, in doing so, he relies upon information irrelevant to my analysis when asserting the base rates of sexual recidivism among the SVP population exceeds the VRAG-R base rate of violent

recidivism.

Quinsey (2017) does not cite sources of actual data that SVP sexual recidivism base rates over five years are greater than the 32% five-year base rate of violent recidivism observed in the VRAG-R standardization sample. Rather, his position relies on what I admittedly found to be a confusing comparison of the statutory definitions of sexually violent offenses in Washington State (Revised Code of Washington 71.09.020(17)) with the violent recidivism criterion used by the VRAG-R. Quinsey (2017) offers not a single authority for his preposterous proposition that the statutory definitions of sexual recidivism as defined by SVP laws in the United States (or Washington state, in particular) produce a five-year base rate of sexual recidivism greater than the five-year violent reoffense base rate observed in the VRAG-R standardization sample. In fact, his assertion is directly refuted by a study he referenced.

Quinsey (2017) cites a Washington State SVP study (Milloy, 2007) to support his suggestion that SVPs have sexual recidivism base rates that are greater than the violent recidivism base rate observed in the VRAG-R standardization sample. It is noteworthy that Quinsey (2017) omitted mention of the sexual recidivism base rate observed in the study when making his contention. Milloy (2007) examined 135 individuals who were found by psychologists to meet the legal definition of an SVP, but the state decided not to file petitions for civil confinement. The sexual recidivism base rate was 23% over six years. To make a fair comparison to the VRAG-R five-year violent recidivism rate, I simply divided the reported base rate from Milloy (2007) by 6 and then multiplied the product by 5 to obtain an estimated five-year sexual recidivism base rate of 19.2%. The estimated five-year sexual recidivism base rate is 40% less than the 32% observed violent recidivism rate over five years among the VRAG-R standardization sample. This observation is consistent with what I described in sections 2.7 and 2.8 of Abbott (2017), where I point out that the VRAG-R violent recidivism rates treat all nonsexual violent offenses as sexually motivated when assessing the likelihood to commit sexual reoffense under SVP laws. This is an unknown source of error that effectively inflates the probability for reoffending sexually and this will contribute to forensic evaluators arriving at false positive conclusions when opining that individuals meet the likely threshold. Since the publication of Abbott (2017), the study by Olver and Sewall (2018) presents data that illustrates this issue.

Table 2 compares the difference in the VRAG-R bin-wise probabilities of reoffense for violent offenses (nonsexual violent and violent) and sexual offenses only among the individuals comprising the Clearwater sample. Risk bin 4 is the only score at which the sexual-only and violent reoffense rates match. In all other risk bins, which contained recidivists, the sexual only reoffense rates were 9.9% to 100% less than the corresponding violent recidivism rates. The data in Table 2 clearly demonstrates, in my opinion, the strong tendency for the VRAG-R violent recidivism rates to inflate the probability of sexual recidivism as legally defined by SVP laws. Reliance on the VRAG-R violent recidivism rate may lead forensic evaluators to falsely conclude that an individual meets the likely threshold. Critics might counter that violent recidivism rates best estimate sexual recidivism rates due to sexually motivated crimes that are charged or adjudicated as nonsexual violent crimes. While this issue has some effect in suppressing the sexual recidivism rate, it is unknown as to the magnitude of its effect since the VRAG-R developers apparently did not identify the extent to which nonsexual violent crimes that were sexually motivated (Harris et al., 2015). For a more detailed discussion about this issue related to the VRAG-R, the interested reader is referred to Abbott (2017).

**Table 2: Clearwater 5-year reoffense rates:
Comparison of Violent to Sexual Only**

VRAG-R Risk Bin	5-year Violent %	5-Year Sexual %	Difference %
1	0	0	-
2	0	0	-
3	5.6	0	-100.6
4	6.7	6.7	0.0
5	22.2	20.0	-9.9
6	22.2	14.8	-33.3
7	29.2	8.3	-71.6
8	44.7	21.1	-52.8
9	55.0	27.5	-50.0

Inspection of other relevant literature exposes Quinsey's fallacious argument that SVP sexual recidivism base rates exceed the violent reoffense base rate reported by the VRAG-R. DeClue and Rice (2016) and DeClue and Zavodny (2014) report findings from studies of sexual recidivism among contemporary SVP offender samples. These data demonstrate five-year sexual recidivism base rates below 10%. Based on the mathematical properties of actuarial tables (Donaldson & Wollert, 2008), the five-year violent recidivism rate of 32% from the VRAG-R would likely overestimate sexual recidivism by approximately 70% to 80% among contemporary SVP offenders. It is reasonable to conclude that the higher base rate observed in the VRAG-R standardization sample when compared to the sexual recidivism base rates from SVP samples would lead to a high rate of false positive conclusions that individuals meet the likely threshold.

In further avoiding current empirical evidence contrary to his opinion, Quinsey (2017) tries to persuade the reader to believe that the Washington State SVP population demonstrates a sexual recidivism base rate that exceeds that of the VRAG-R because the law requires prediction of risk over the lifetime of the offenders. The validity of the notion that the rate of sexual reoffense over the lifetime escalates is simply speculative because empirical evidence lacks regarding the observed sexual recidivism among sexual offenders over their lifetimes. Research has demonstrated that sexual reoffense rates decrease or plateau over time associated with such factors as offender types and social context, time free in the community, and advancing age (Lussier & McCuish, 2016; Rettenberger, Briken, Turner, & Eher, 2015; Hanson, Harris, Helmus, & Thornton, 2014; Helmus, Thornton, Hanson, & Babchishin, 2012; Barbaree & Blanchard, 2008) and these data contravene Quinsey's assertion of an ever-increasing rate of sexual recidivism over the lifetime.

In addition, the idea of lifetime risk appears to have no basis in law and, thus, asserting it amounts to the inappropriate imposition of an extra-statutory condition when assessing the likely threshold. The Washington SVP statute or decisional law does not specify that risk for predatory acts of sexual violence must be assessed over the lifetime of the individual. Quinsey (2017) provides no authority in his response that SVP laws require the prediction of risk over the lifetime of the offender. In fact, I am not aware of statutory or case law in other SVP jurisdictions in the United States that specify the likely threshold must encompass the remaining lifetime of the offender. The notion of lifetime risk is also contrary to the United Supreme Court's ruling on the constitutionality of SVP laws where the

court found it permissible to identify and incapacitate those sexual offenders who *currently* exhibit a mental condition that makes them sexually dangerous (Kansas v. Crane, 2002; Kansas v. Hendricks, 1997).

Appeal to Authority

Quinsey (2017) references State of Washington governmental sources (King County Sexual Assault Resource Center, 2017; Washington State Attorney General, 2017) to support the notion that the SVP population is unlike the VRAG-R standardization sample and, therefore, SVP's reoffend sexually at a higher rate than violent recidivism rate observed among the VRAG-R standardization sample. Inspection of the source documents reveals that Quinsey (2017) took great liberty with the information to support his conclusion.

Quinsey (2017) references that 1.4% of all sexual offenders released from Washington prisons were tried as SVP's in 2014 (Washington State Attorney General, 2017). I have practiced in the state of Washington conducting SVP evaluations for many years and it escapes me how the proportion of cases that go to trial makes the SVP population "very, very special," as Quinsey (2017) contends. What Quinsey (2017) fails to recognize is that the rate of trials per year is driven by available human resources in prosecutor and defense attorney offices and availability of courts to conduct trials rather than representing some unique features causing the sexual recidivism base rate for Washington SVPs to exceed the violent recidivism base rate from the VRAG-R standardization sample. Quinsey (2017) fails to provide data in support of or to explain how the designation of "very, very special" produces sexual recidivism base rates among Washington SVPs that exceed the VRAG violent reoffense rates.

Quinsey (2017) reports information from a governmental website (King County Sexual Assault Resource Center, 2017) that he describes as characteristics for identifying sexual offenders who were evaluated for civil confinement as SVPs. He further argues these features are consistent with SVPs, in Washington at least, having higher sexual recidivism base rates than the violent reoffense base rate observed in the VRAG-R standardization sample. Once again, he offers no data from the source to support his opinion (because none is provided). The King County Sexual Assault Resource Center (2017) provides examples of characteristics that define the three tiers of sexual offender registration in Washington. It does not state the characteristics are used to preselect individuals for SVP consideration and no sexual recidivism data are provided that would support Quinsey's (2017) contention that sexual recidivism base rates among Washington state SVPs exceed the VRAG-R's violent reoffense base rate.

Straw man arguments

In my article, I addressed a generally accepted standard for using criterion referenced tests, like the VRAG-R, where the validity of the sexual recidivism outcome is dependent upon the presumption that the individual is fungible with the standardization sample. In section 2.1 - 2.4 (Abbott, 2017), I discussed four areas that raise serious questions as to whether the assumption of mutual exchangeability is met when using the VRAG-R with SVPs. I appreciate Quinsey (2017) clarifying the make-up of the standardization sample; however, this information by no means makes the VRAG-R standardization sample a good fit for individuals being considered for civil confinement as SVPs.

Quinsey (2017) misdirects criticism toward me related to those members of the VRAG-R standardization who did or did not receive cognitive behavioral sexual offender treatment. He

appears to argue that cognitive behavioral treatment should not be considered when deciding the extent of fit of the VRAG-R standardization sample to an individual being assessed because of the proven ineffectiveness of treatment in reducing sexual recidivism. I do not disagree with Quinsey about the ineffectiveness of sexual offender treatment in reducing sexual recidivism risk over untreated sexual offenders, but his point is simply irrelevant to my analysis of the VRAG-R.

Conclusion

The criticisms that Quinsey (2017) levels against my analysis of and conclusions about the VRAG-R when evaluating the likely threshold in SVP psychological evaluations and testimony entails inconsistent arguments and questionable interpretation of source material. Quinsey's response offers no substantive information or data that would lead me to alter my analysis, findings, conclusions, or recommendations (Abbott, 2017). The VRAG-R violent recidivism probability estimates lack sufficient fit to the legally defined likely threshold and an individual being assessed for civil confinement as an SVP would not be considered fungible with the VRAG-R standardization sample. Without satisfying these fundamental premises for the application of actuarial risk data, forensic evaluators and triers of fact would lack necessary confidence that the VRAG-R bin-wise probabilities for violent recidivism are sufficient, relevant, or probative to evaluate whether the individual being assessed meets the SVP likely threshold. Moreover, the VRAG-R violent recidivism rates are likely to inflate the probability of sexual recidivism and this would result in an unknown magnitude of false positive conclusions that individuals meet the likely threshold. For these reasons, I continue to conclude that the VRAG-R, at its current state of development, is not an appropriate actuarial measure to use in SVP civil confinement risk assessments. Interested readers are referred to section 3.1 of Abbott (2017) for a discussion about research efforts that would be necessary to complete for the VRAG-R to have a sufficient fit for assessing individuals undergoing legal proceedings in the USA as SVPs. In the meantime, other actuarial instruments such as the Static-99R (Hanson, Thornton, Helmus, & Babchishin, 2016) and VRS-SO (Wong, Olver, Nicholaichuk, & Gordon, 2003-2017) are better suited to address the likely threshold as defined by SVP laws.

Acknowledgement

The author thanks Thomas K. Zander and Ian Coyle for their helpful comments in preparing this response.

References

1. Abbott, B. R. (2017). Sexually violent predator risk assessments with the Violence Risk Appraisal Guide-Revised: A shaky practice. *International Journal of Law and Psychiatry*. Online First. doi: 10.1016/j.ijlp.2017.03.003.
2. Abbott, B. R. (2013). The Utility of Assessing "External Risk Factors" When Selecting Static-99R Reference Groups, *Open Access Journal of Forensic Psychology*, 5, 89-118.
3. American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
4. American Psychological Association. (2013). *Specialty guidelines for forensic psychology*. Washington, D.C.: American Psychological Association.
5. American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, D.C.: American Psychological Association.

6. Barbaree, H. E., & Blanchard, R. (2008). Sexual deviancy over the lifespan. In, D.R. Laws & W.T. O'Donohue (Eds.), *Sexual deviance: Theory, Assessment, and Treatment* (pp. 37-60), New York: Guildford Press.
7. DeClue, G. & Rice, A., (2016). Florida's released "Sexually Violent Predators" are not "High Risk," *Open Access Journal of Forensic Psychology*, 8, 22-51.
8. DeClue, G., & Zavodny, D. L. (2014). Forensic use of the Static-99R: Part 4. Risk Communication. *Journal of Threat Assessment and Management*, 1(3), 145-161.
9. Donaldson, T. S. & Abbott, B. R. (2011). Prediction in the Individual Case: An Explanation and Application of Its Use with the Static-99R in Sexually Violent Predator Risk Assessments. *American Journal of Forensic Psychology*, 29(1), 5-35.
10. Donaldson, T.S., Abbott, B.R., & Michie, C. (2012). Problems with the Static-99R prediction estimates and confidence intervals. *Open Access Journal of Forensic Psychology*, 4, 1-23.
11. Donaldson, T. & Wollert, R. (2008). A Mathematical Proof and Example That Bayes's Theorem Is Fundamental to Actuarial Estimates of Sexual Recidivism Risk. *Sexual Abuse: A Journal of Research and Treatment*, 20(2), 206-217.
12. Duwe, G. & Freske, P. (2012). Using logistic regression modeling to predict sex offense recidivism: The Minnesota sex offender screening tool-3 (mnsost-3). *Sexual Abuse: A Journal of Research and Treatment*, Online First, doi: 10.1177/1079063211429470.
13. Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (2003). Minnesota sex offender screening tool-revised (Mn-SOST-R): Development, performance, and recommended risk level cut scores. Unpublished manual.
14. Gregório-Hertz, P., Etzler, S., Rettenberger, R., & Eher, R. (2016). Cross-Validation of the Revised Version of the Violence Risk Appraisal Guide (VRAG-R) in a Sample of Individuals Convicted of Sexual Offenses. *Sexual Abuse*. Advanced online publication. doi: 10.1177/1079063219841901
15. Gregório-Hertz, P., Eher, R., & Rettenberger, M. (2017). An independent validation of the VRAG-R in a sample of sexual offenders. Submitted for publication.
16. Gregório-Hertz, P, Rettenberger, R., & Eher, R. (2016). A cross validation of the VRAG-R using a sexual offender sample from Austria. Presented at the International Association for the Treatment of Sexual Abusers, Copenhagen.
17. Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales. *Criminal Justice and Behavior*, 44(1), 26-39, doi: 10.1177/0093854816683956.
18. 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(15), 792-813. doi: 10.1177/0886260514526062.
19. Hanson, R. K., & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136. doi: 10.1023/A:1005482921333.
20. 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: Journal of Research and Treatment*, 28(3), 218-252. doi:10: 1177/1079063215574710.
21. Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.
22. Helmus, L. M. & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, 44(1), 8-25. doi: 10.1177/00938554816678898.
23. Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Assessing risk of older sex offenders: Developing the Static-99R and Static-2002R. *Sexual Abuse: A Journal of Research and Treatment*, 24(1). 64-101. DOI: 10.1177/1079063211409951
24. Howard, P. D. (2017). The effect of sample heterogeneity and risk categorization on area under the curve predictive validity metric. *Criminal Justice and Behavior*, 44(1), 103-120. doi:

- 10.1177/00938554816678899.
25. Jackson, R., Travia, T., & Schneider, J. (2009). Annual survey of sex offender civil commitment programs. Sexual Offender Civil Commitment Network Research Committee 2008. Retrieved October 25, 2009, from http://www.soccpn.org/images/2008_20Annual_20Survey_20of_20Programs.pdf.
 26. Janus, E. S. & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility, and accountability. *American Criminal Law Review*, 40(1143), 1-59.
 27. *Kansas v. Crane* (2002), 534 U.S. 407.
 28. *Kansas v. Hendricks* (1997), 521 U.S. 346.
 29. King County Sexual Assault Resource Center (2017). Classification of sex offenders: Frequently asked questions. (Retrieved October 1, 2017, from <http://www.k12.wa.us/safetycenter/Offenders/pubdocs/FAQonClassification.pdf>).
 30. Lussier, P. & McCuish, E. (2016). Desistance from crime without reintegration: A longitudinal study of the social context and life course path to desistance in a sample of adults convicted of a sex crime. *International Journal of Offender Therapy and Comparative Criminology*, 60(15), 1791-1812. doi: 0.1177/0306624X16668179.
 31. Milloy, C. (2007, June). Six-year follow up of 135 released sex offenders recommended for commitment under Washington's sexually violent predator law, Washington State Institute for Public Policy.
 32. Olver, M. E. (2017, October 27). Predictive properties of the VRAG-R scores in a treated sexual offender sample. Presentation at the 36th annual association of the treatment of sexual abusers, Kansas City, MO.
 33. Olver, M. & Seawall, L. A. (2018). Cross-validation of the discrimination and calibration properties of the VRAG-R in a treated sexual offender sample. *Criminal Justice and Behavior*, 45(6), 741-761.
 34. Quinsey, V. (2017). The art of misdirection: Abbott's shaky take on the VRAG-R. *Sexual Offender Treatment*, 12(2), 1-6.
 35. Rettenberger, M. (2017, October 27). Cross-validation of the German version of the VRAG-R. Presentation at the 36th annual association of the treatment of sexual abusers, Kansas City, MO.
 36. Rettenberger, M., Briken, P., Turner, D., & Eher, R. (2015). Sex offender recidivism among a population-based prison sample. *International Journal of Offender Therapy and Comparative Criminology*, 59(4), 424-444. doi: 10.1177/0306624X13516732
 37. Rice, M., Harris, G.T., Lang, C. (2013). Validation of and revision of the VRAG and SORAG: The violence risk appraisal guide-revised (VRAG-R). *Psychological Assessment*, 25(3), 951-965. doi: 10.1097/a0032878.
 38. Scurich, N. (2018). The case against categorical risk estimates. *Behavioral Sciences and the Law*, 36, 554-556. doi: 10.1002/bsl.2382.
 39. Scurich, N., & Krauss, D. (2014). The presumption of dangerousness in sexual violent predator commitment hearings. *Law, Probability, and Risk*, 13, 91-104. Doi: 10.1093/lpr/mgt015.
 40. Singh, J. P. (2013). Predictive validity performance indicators in violent risk assessment: A methodological primer. *Behavioral Sciences and the Law*, 31, 8-22. Doi: 10.1002/bsl.2052.
 41. Washington State Attorney General. (2017). Sexual violent predators. (Retrieved October 1, 2017, from <http://www.atg.wa.gov/sexually-violent-predators>).
 42. Woodworth, G. G., & Kadane, J. B. (2004). Expert testimony supports post-sentence civil incarceration of violent sexual offenders. *Law, Probability, and Risk*, 3, 221-241.
 43. doi: 10.1093/lawprj/3.3-4.221.
 44. Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003-2017). The Violence Risk Scale-Sexual Offense version (VRS-SO). Saskatoon, Canada: Regional Psychiatric Centre,

University of Saskatchewan.

Author address

Brian R. Abbott, Ph.D.

111 N. Market Street

San Jose, CA 95136

USA

brian@dr-abbott.net