

The art of misdirection: Abbott's shaky take on the VRAG-R

Vernon L. Quinsey
Queen's University, Kingston

Abstract

Abbott (2017) makes a variety of criticisms of the use of the Violence Risk Appraisal Guide-Revised (VRAG-R, Harris et al. 2015) in Washington State Sexually Violent Predator (SVP) hearings. His article uses misdirection to obfuscate and conceal two issues involved in the application of the VRAG-R in Washington State SVP cases: The nature of the decision that is to be made and the characteristics of the offenders who are involved.

DRAFT, JULY 25, 2017

I examine and provide examples of the persuasion techniques employed in Abbott's (2017) opinion piece. Misdirection is the most common of these, although misinformation and veiled conspiracy theorizing also play a role.

Background

At a time when many scientists worry about the replicability of their results (Open Science Collaboration, 2015), the many individual-level correlates of crime, such as sex and age at first arrest, are routinely found to predict reoffending across time, countries, and various populations, such as psychiatric patients and offenders incarcerated in the criminal justice system (e.g., Skeem et al., 2014; Witt, van Dom, & Fazel, 2013). So frequent and ubiquitous are these findings, that should an investigator fail to obtain them, the most parsimonious explanation is that an error has been made.

The robustness of these correlates of reoffending has encouraged their combined use in actuarial risk assessment instruments. These instruments estimate the probability that an assessee will commit at least one re-offense (or re-offense of a particular kind) within a specific time period. All assessments of risk necessarily involve the estimation of a probability (an on-line dictionary offers six definitions of risk in various contexts and all involve probabilities of adverse outcomes: <http://www.businessdictionary.com/definition/risk.html>). To illustrate: If it is asserted that an offender has a high risk of reoffending, there is no sensible answer to the question "How high?" or "What is 'high?'" without the specification of a probability. Thus, any risk assessment instrument involves a probability estimate, explicit in actuarial instruments and implicit (relying on consumers' intuitive and untutored understanding of the implied probability) in non-actuarial ones.

Calibration

Risk communication in the VRAG-R is accomplished by dividing the total range of scores into nine equal-sized bins or ranges. Each range is associated with an empirically derived probability of violent (including hands-on sexual) reoffending. This information can be used in titrating the amount of treatment or supervision an offender receives, although percentile ranks or risk ratios may also be used for these purposes in some situations (see Harris, Lowenkamp, & Hilton, 2015.).

Examinations of how faithfully the probabilities associated with each bin are replicated in new samples are termed "calibration" studies. Abbott's article focusses in large part on issues involved in calibration studies of a variety of actuarial instruments. Although calibration studies are relevant to issues of accuracy in risk communication, they are not closely linked to binary decision situations such as SVP adjudications. Because of this, Abbott's calibration discussion is an elaborate misdirection from the task at hand in SVP trials. When a binary decision is to be made, such as when an offender is classified as a sexually violent predator or not, a cut-off score or probability is established that reduces, in the case of the VRAG-R, nine

bins to two. The resulting type of decision matrix (in this case, two classes of offender by two classes of outcome) has been studied extensively in psychology for decades (e.g., Tanner & Swets, 1954).

Although not directly relevant to SVP hearings, it should be noted for the sake of completeness that Abbott's suggestion that VRAG-R bin probabilities are unstable have not been empirically supported by initial investigations--see Hertz, Rettenberger, & Eher (2016) and Olver (2017) for calibration among Austrian and Canadian sex offenders, respectively.

Base rates

Abbott's abstract states that "Clinicians find it increasingly difficult to affirm the likelihood threshold in the face of decreasing base rates...", but does not cite the evidence upon which this statement is based or in which population this decline has occurred. Although population crime rates have decreased substantially throughout North America in recent decades, remaining much higher in the U.S. than in Canada (Lalumière et al., 2005; Mishra & Lalumière, 2009; Zimring, 2006), this decrease may or may not be related to the base rate of reoffending among identified offenders—base rates of recidivism among released offenders can rise, fall, or remain the same despite a falling base rate of offending in the general population. Analyses of follow-up data from temporal cohorts in the VRAG-R data set found no decrease in the rate of sexual or violent reoffending across time (Harris et al., 2015).

Abbott spends considerable space arguing that the base rate of sexual recidivism among SVP assesses is likely to be lower than the base rate of sexual or violent reoffending in the VRAG-R construction sample. 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. The outcome of concern in SVP legislation (statute 71.09.020(17)) is narrower than the operational definition of violent re-offenses in the VRAG-R research in that the former excludes non-sexually motivated crimes of violence (with several exceptions, including murder and kidnapping). It is broader in including such crimes as sexually motivated residential burglary. More importantly, the relevant time-period in SVP legislation is the lifetime of the offender, not a specific time interval as given in the VRAG-R actuarial tables. These differences serve to make the VRAG-R 5- and 12-year probabilities underestimates of the probability of sexual reoffending among those considered for SVP status.

The VRAG-R development sample

The goal of the VRAG-R sampling strategy was to obtain a heterogeneous sample of the kind of offenders who are often assessed for risk of violent (including sexual) reoffending and to show that the VRAG-R worked similarly for the various subsamples. Because of this, very low risk offenders (despite inclusion of a sample of sex offenders referred from the community) are likely underrepresented. In contrast to the assertion of Abbott's (p.5), not all the VRAG-R

subjects came from Oak Ridge--sex offenders from a community sample and two samples from federal corrections were also included. Similarly, even the Oak Ridge sample was not composed exclusively of mentally ill offenders, a substantial number were only assessed briefly at Oak Ridge before going directly to federal or provincial corrections.

In contradiction to the statements in Abbott's Section 2.4, the controversial methods of treatment, such as "the capsule" described in this section were applied to a small number of volunteer patients who were present in the late sixties and early seventies, these offenders make up a miniscule part of the VRAG-R sample. Many of the sex offenders who were treated at Oak Ridge received a cognitive behavioural type treatment. Most of the prison sample of sex offenders received similar treatment. It would make no difference, however, whether the sex offenders received cognitive behavioural treatment or not because the evaluative literature, including a well-designed random assignment study of cognitive-behavioral treatment of sex offenders (Marques et al., 2005) does not show that treatment reduces sexual recidivism (Rice and Harris, 2013). Moreover, retrospective evaluations of cognitive behavioral programs for the sex offenders included in the VRAG-R sample (Quinsey, Khanna, & Malcolm, 1998; Rice, Harris, & Quinsey, 1991) found no treatment effect. Moreover, treatment variables were included as candidate predictors in the development of the VRAG but none of them made the cut. The assertions concerning treatment in Abbott's article are therefore essentially irrelevant.

Abbott's arguments purporting to show that the VRAG-R development sample is "special" would only work to support his thesis if they demonstrate that probability estimates based on it are lower than what would be expected in the SVP-eligible assesseees. The focus on the alleged weirdness of the VRAG-R sample is another instance of misdirection but a far more spectacular one than that involved in the discussion of calibration. It is not the VRAG-R development sample that is special, it is the SVP-eligible assesseees who are very, very special. Each year, about 1000 sex offenders are released in the state of Washington but only a few are selected for application of the SVP legislation: For example, in fiscal year 2014, only 14 cases (1.4%) were tried in court (<http://www.atg.wa.gov/sexually-violent-predators>). These "forensic superstars" are selected according to risk-related criteria such as previous sex crime and other criminal convictions, predatory characteristics (such as seeking out victims unknown to the offender), having used violence in their crime, and having refused or failed to complete sex offender treatment (www.k12.wa.us/safetycenter/Offenders/pubdocs/FAQonClassification.pdf).

Thus, when the VRAG-R is employed in Washington State SVP hearings, norms developed on a heterogeneous sample of offenders selected to be roughly representative of offenders who are ordinarily assessed for risk of sexual or violent reoffending are applied to an elite group of sex offenders who have been highly pre-selected for risk. One doesn't have to be a statistician to understand the consequences: The anticipated base rate of sexual reoffending in the Sexually Violent Predator assesseees is much higher than that indicated by the VRAG-R. Even with much less rigorous pre-selection on risk, the score-wise recidivism rates on actuarial scales are much higher than the norms for an unselected sample (Hanson, Thornton, Helmus, & Babischin, 2016). In a more directly relevant investigation, Milloy (2007) followed 135

Washington State sex offenders who were referred for civil commitment, *but not committed*, for a six-year period. Milloy concluded that in comparison to released Washington State sex offenders in general “this population of released sex offenders who were referred for civil commitment is a unique subgroup with much higher recidivism rates”.

A conspiracy against justice

Abbott (2017, p. 8, section 2.9) states that he was denied access to the VRAG-R data: “To study this effect with the VRAG-R the author requested the frequency data at all VRAG-R scores from the custodian of the data and to compute logistic regression predicted estimates [the grammar is as quoted] at each total score but the request was not granted (V.L. Quinsey, personal communication, October 20, 2015).” One must wonder why the custodian is not on the side of justice and just what nefarious secrets he is hiding within those VRAG-R bins.

In my October 20, 2015 e-mail, I replied: “Thanks for your request. Unfortunately, things are in disarray since the deaths of Marnie and Grant. It is unclear at present how long it will take to locate and render intelligible the VRAG-R data set and establish a vetting protocol.” To clarify, the e-mail refers to the sudden deaths of Marnie Rice (2015) and Grant Harris (2014), the principal authors of the VRAG-R. Those outside “the life” might wonder why one would not just copy the data and send it. There are several reasons. Because I was involved in only an advisory capacity for the VRAG-R analyses, I had to figure out how the data were organized. The data used in the final analyses had to be identified from among older overlapping data sets and the variables had to be given labels that those unfamiliar with this large data set could understand. Second, the data had to be cleaned and in some cases aggregated to ensure that there was no information that could be used to identify individuals. With the collaboration of Data Services of the Queen’s University Library, the VRAG-R data were published at the end of August 2016 (<https://qspace.library.queensu.ca/handle/1974/14829>). A link to these data were simultaneously published on the official VRAG-R Website (vrag-r.org). There is, of course, a vetting process to get access to some of the more sensitive information at the individual level.

It should be recognized that publishing forensic data that have been over thirty years in the gathering is not cost free. Because actuarial data are often the subject of partisan attack in the courts and the scientific literature more generally, a small number of individuals are encouraged to use forensic actuarial data for pecuniary advantage or to advance an ideological agenda. Moreover, there is no guarantee that potential users are competent to properly execute and interpret the analyses they undertake. When analyses appear in the literature that are incorrect or misleading, the original authors or their colleagues feel obligated to find the errors and correct the misinterpretations. This process is tiresome and frustrating—one would rather be doing new research or spending one’s (ironically labeled) “golden years” in retirement. On the other hand, it is satisfying to enable researchers to use these newly available data to move the field ahead. Because of these issues, the American Psychological Association Data Sharing Working Group (2015) advises that although investigators have an ethical responsibility to make their data available to individuals to verify the accuracy of the original analyses and to check for fraud, they do not have an ethical obligation to share their

data with others for the purposes of conducting new analyses. Rather, the association specifies that these requests should be considered bearing in mind the requestor's qualifications, and that costs of requests for data should be borne by the requestor, not the original researchers.

Allegorical Summary

To make the issues explicated in the foregoing more vivid, I have summarized them using an allegory.

Imagine that you have been captured by an evil overlord of the new corporate order and forced to participate in his hobby reality TV show—"Nouveau Serfs". There is a carnival strength test apparatus on the set. Contestants strike a plate at the bottom with a sledge hammer, attempting to send a metal ball to ring the bell at the top. The strength test apparatus has been imported from Canada and the numbers running from bottom to top are divided into nine equal sized ranges or bins. The bottom range of numbers is labelled "girly-man", the penultimate, "John Henry", and the top "Thor".

The gamble you have been forced to accept is to bet on which bin a randomly selected local yokel will score in. If you lose, you and your family will labour in the dungeons of a Trumpian tower for less than minimum wage in perpetuity. If you win, you and your family will receive free health care and coupons redeemable for plastic trinkets from popular on-line retailers.

A search of the dark web reveals an actuarial scale that has been secretly developed by Canadian carnival gamblers; it contains variables such as height, weight, and number of arms (scored zero, one, or two). Given the high stakes, you consider the scale carefully. What if the calibration of the apparatus is different for Canadian men than men in Washington State? Your mind conjures up visions of burly Canadian lumberjacks driven near mad by the incessant biting of the Northern Ontario blackfly and effete man-bunned hippies sipping kale lattes in upscale Seattle coffee houses.

Just before you place your bet, you learn that the customary crowd of local yokels has been replaced by the weight-lifting team from the University of Washington and you are now to bet whether the contestant will score in the top third or bottom two-thirds of the bins. You start wondering how many plastic baubles you can fit into your trailer.

References

- 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*, <http://dx.doi.org/10.1016/j.ijlp.2017.03.003>
- American Psychological Association (2015). *Dating sharing: Principles and considerations for policy development*. Author: Washington, DC.
- Harris, G.T., Lowenkamp, C.T., & Hilton, N.Z. (2015). Evidence for risk estimate precision: Implications for individual risk communication. *Behavioral Sciences and the Law*, *33*, 111-127.
- Harris, G.T., Rice, M.E., Quinsey, V.L., & Cormier, C. (2015). *Violent offenders: Appraising and managing risk*. Third Edition, Washington, D.C.: American Psychological Association.
- Hertz, P.G., Rettenberger, M., & 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 Sex Offenders Conference, Copenhagen.
- Lalumière, M. L., Harris, G. T., Quinsey, V. L., & Rice, M. E. (2005). *The causes of rape: Understanding individual differences in the male propensity for sexual aggression*. Washington: American Psychological Association.
- Marques, J.K. et al. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse : A Journal of Research and Treatment*, *17*, 79-107.
- Mishra, S., & Lalumière, M. L. (2009). Is the crime drop of the 1990s in Canada and the USA associated with a general decline in risky and health-related behavior? *Social Science and Medicine*, *68*, 39-48.
- Milloy, (2007). *Six-year follow-up of 135 released sex offenders recommended for commitment under Washington's Sexually Violent Predator Law, where no petition was filed*. Olympia, WA : Washington State Institute for Public Policy.
- Olver, M. (2017). Cross-validation, calibration, and risk communication using the VRAG-R. To be presented at the meeting of the Association for the Treatment of Sex Offenders Conference. Kansas City.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 1-8.

- Quinsey, V.L., Khanna, A. & Malcolm, B. (1998). A retrospective evaluation of the Regional Treatment Centre Sex Offender Treatment Program. *Journal of Interpersonal Violence, 13*, 621-644.
- Rice, M.E. & Harris, G.T. (2013). Treatment for adult sex offenders: May we reject the null hypothesis? In K. Harrison & B. Rainey (Eds.), *The Wiley-Blackwell handbook of legal and ethical aspects of sex offender treatment and management* (pp. 217-235). New York, NY: Wiley Blackwell.
- Rice, M.E., Harris, G.T., & Quinsey, V.L. (1991). Evaluation of an institution-based treatment program for child molesters. *Canadian Journal of Program Evaluation, 6*, 111-129.
- Skeem, J.L., Winter, E., et al. (2014). Offenders with mental illness have criminogenic needs, too: Toward recidivism reduction. *Law and Human Behavior, 38*, 212-224.
- Tanner, W.P. & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401-409.
- Witt, K., van Dom, R., & Fazel, S. (2013). Risk factors for violence in psychosis: Systematic review and meta-regression analysis of 110 studies. *PLoS ONE (8(2))*: e55942.
Doi:10.1371/journal.pone.0055942.
- Zimring, F. E. (2006). *The great American crime decline*. Oxford: Oxford University Press.

Acknowledgement: Maaïke Helmus, Brian Judd, and Martin Lalumiere provided helpful comments and encouragement in the preparation of this article.