

Sexual offender risk assessment strategies: is there a convergence of opinion yet?

Douglas P. Boer, Ph.D.
Department of Psychology, University of Waikato

[Sexual Offender Treatment, Volume 1 (2006), Issue 2]

Abstract

While there is no convergence of opinion in the academic literature, many clinicians have already adopted a convergent approach to risk assessment, whether through deliberate adoption of complementary instruments or through informal clinical modification of risk assessment instrument findings. The present paper suggests that a convergent approach to risk assessment may be both the most responsible and most appropriate approach at this time, given the lack of research to substantiate a strong superiority of one type of instrument (actuarial risk tests versus structured clinical guidelines) over the other as well as the complementary nature of these types of tests in terms of how risk is conceptualized and analyzed. The research in a convergent approach is meager and this issue deserves some attention to ensure that all clinicians doing risk assessments of sexual offenders are providing the most information possible to decision makers to protect the rights, safety, and security of our clients (the correctional and justice systems, the public, and the offenders).

Key words: Risk assessment, sexual offenders, actuarial risk assessment, clinical risk assessment

The brief answer to the question posed in the title of this article is no. The research literature remains fragmented and oppositional, with test authors from the actuarial risk test (ART) and the structured (or professional) clinical guideline (SCG) variety rarely admitting in print that their instruments do not work as well as instruments from the opposing camp. In sum, there has been little official change in the state of the ART versus SCG debate since Douglas, Cox, and Webster (1999) determined that there was no clear front-runner between these two types of instruments. If published today, that same article would be little changed except for additional references in the opposing sections.

Seven years since the seminal paper by Douglas et al, (1999), the issue of what type of instrument works best in the determination of sex offender risk remains unresolved. There are some strong opinions on either side, and the status of this debate is perhaps best reflected in introductory section of the Risk for Sexual Violence Protocol (RSVP; Hart, Kropp, and Laws, 2003). These authors cogently argue that the two approaches have their own respective research basis and these research bases are flawed in specific ways. These authors were primarily looking at the Sexual Violence Risk 20 (SVR-20; Boer, Hart, Kropp, and Webster, 1997) as an example of a SCG and a variety of ARTs used with sex offenders.

Hart, et al., (2003) suggested that much of the research conducted on the SVR-20 to date is flawed due to the use of institutional files in most of the studies (as opposed to actual offender interviews in conjunction with file review), lack of systematic inter-rater reliability in some of the studies, and the simple lack of enough studies to provide a basis from which generalizations of the usefulness of the SVR-20 as a SCG approach. The ART research basis for a wide variety of actuarial tests used with

sex offenders was then criticized by these same authors, for reasons including: the lack of a published manual for administration, scoring and test interpretation; lack of inter-rater reliability; short-comings in the test construction samples; lack of replication in other samples (although this criticism is not clearly relevant to all actuarial tests, e.g., the STATIC-99 by Hanson and Thornton, 1999, and the Sex Offender Risk Appraisal Guide, or SORAG by Quinsey, Harris, Rice, and Cormier, 1998). The last criticism is particularly critical, according to Hart, et al., (2003) as the predictive accuracy of ARTs may be over-estimated due to the mathematical procedures used to construct the scales, and the shrinkage of their predictive accuracy has not been directly measured using direct replication. Basically, Hart, et al., (2003) argue that the accuracy with which these tests make specific predictions of sexual violence has never been evaluated (p. 9). More simply put, the predictive accuracy of the ARTs for sexual violence has not been thoroughly evaluated as even the construction sample would have changed unpredictably after the scale was developed (i.e., after the statistically-determined optimal follow-up periods) and the tests are not re-developed, but simply re-validated, on independent samples. The latter criticism suggests that other variables originally tested on the construction sample could outperform the variables in the ART, but this is never tested in a simple revalidation study. If this was done, it is quite possible that the final variables in an ART would differ from sample to sample making it difficult to adopt any specific measure as it would likely change with each new application. Of course, selling the method of scale derivation may be better than very different jurisdictions simply adopting an ART developed elsewhere as is commonly done around the world (with or without validation exercises done after implementation), but deriving ARTs is a slow and arduous process best left to the experts. However, endorsing adoption of an ART without initial local validation would seem unethical at best, and dangerous to the public good in that jurisdiction at worst.

While the debate of what works better ARTs or SCGs is of some academic and ethical interest, it is my opinion that this debate is of a relatively unimportant nature to clinical practice if the derivation issues are put to rest. That is, if the ART is locally validated, then the ART in question is useable and whether it works better (or not as well as) a relevant SCG is a research question of limited interest to clinical practice. I would like to further propose that, until the debate dust is settled that it is best that we use a convergent approach in all sex offender risk assessments. This suggestion is not being made because I feel that the SVR-20 (currently under revision), the first of the SCGs for sexual violence, is at risk of being shown to be less statistically predictive of sexually violent recidivism than any of the ARTs constructed for this purpose. In fact, the data to date are indicating that the opposite is true. Nor am I concerned that subsequent SCGs, such as the RSVP will replace the SVR-20 as the new standard for SCGs for sex offender risk assessment. However, I would note that if the additional few new items in the RSVP add significant power over and above the SVR-20 in terms of predictive accuracy, I would be an advocate of the newer test. Meanwhile, the SVR-20 remains in common use around the world, translated into numerous languages. New studies continue to show that the instrument works well, but these same studies (e.g., de Vogel, van Beek, Mead, and de Ruiter, 2002) show that certain ARTs also work well and lend further support for their use.

This brief paper examines the usefulness of considering laying down the research gauntlets when one is not doing research, but is doing the practice of risk assessment, where the goal of practice is public safety, and not test advocacy. This paper is, admittedly, an editorial of sorts and I will voice my opinions to spark further debate on this and related topics. My reason for doing this paper is simple: there has not been much by way of reasonable clinical practice suggestions coming out from the test authors in the test manuals (other than to use their test), the literature on test validity, or even the meta-analyses of predictive variables. Thus, we need to have a discussion on what is best practice in risk assessment practice.

Before I address my first issue regarding the ART versus SCG debate (i.e., the debate being of a relatively unimportant nature in terms of practical utility), I must admit that I am perplexed by specific test allegiance or allegiance to one type of test or another in general. Quinsey and his colleagues

(1998), all of whom I hold in high esteem personally, have made indefensible claims about the virtues of ARTs compared to any non-ART approach (the latter including, presumably, SCGs). The silliest of these claims is the oft-cited statement: actuarial methods are too good and clinical judgment too poor to risk contaminating the former with the latter (p. 171, Quinsey, et al, 1998). Why a group of generally careful social scientists would ever say such a statement like this is beyond scientific reason and appears to be simply a matter of belief on their part. Alternatively, I would like to suggest that we know very little about risk assessment and that it is presumptuous to suggest otherwise. Taking a strong stance and voicing such polemics against other types of risk assessment strategies, only insults the many clinicians who use their clinical knowledge quite accurately in their practice of conducting risk assessments. If unstructured clinical acumen was worthy of criticism in the past (e.g., Litwack, 2001), it would seem to be less so given the meta-analytic findings of Hanson and Morton-Bourgon (2004) who showed that unstructured clinical opinion was predictive of sexual recidivism. However these same authors found that risk levels determined by unstructured clinical opinion were not as predictive as those determined by ARTs, and the latter risk levels not as predictive as those determined by the SVR-20 (respective d levels of .40, .61, and .77). Therefore, one could ask, why do we need a convergent approach for sex offender risk assessment, when the SVR-20 clearly outperforms the ARTs and unstructured opinion? This is a question more easily asked than answered.

First, the SVR-20 and all other SCGs look at risk from a different perspective than do the ARTs. ARTs only look at risk in terms of likelihood in discrete, optimized intervals, while SCGs look at risk in terms of lethality, victim specificity, frequency, offender-unique factors, and (non-numerical estimates of) likelihood. Second, these two types of instruments are evaluated and constructed very differently. ARTs are derived post-analysis from populations of known recidivists. The best predictors and the best combination of predictors are those that make it into the actuarial equation that best describe the recidivism data. SCGs are derived pre-analysis from the literature, clinical and research (and meta-analytic). SCGs are then validated via follow-up studies, and via comparison with relevant ARTs. There are numerous variables in the SCGs that are not present in ARTs. The STATIC-99, for example, does not include many of the variables that are present in the SVR-20 (e.g., psychopathy, sexual deviancy) and these would seem key to risk assessment given the findings of past and recent meta-analyses (e.g., Hanson and Morton-Bourgon, 2004, Craig, Browne, Stringer, and Beech, 2005). Third, the research and validation literature is lacking in independent appraisal of ARTs and SCGs. It is best, in my opinion, that such research be conducted by a third party (i.e., not a test author) for best confidence in the data basically because it is rare to find studies by test authors that put their own instrument's ability to predict recidivism below that of other's instruments. Hanson and Morton-Bourgon's (2004) recent meta-analysis provides a very nice appraisal of the current risk assessment instruments available to the professionals doing this work and is probably a better source of information in this area than most test-specific or test-comparison articles on risk assessment measures in the current literature. The problem that arises when one checks the reference section of Hanson and Morton-Bourgon's article is that the independence of validation studies is clearly problematic. In addition, the number of validation articles are relatively few for some tests (including the SVR-20). Few tests of either the ART or SCG variety can boast the validation samples that bolster the use of the STATIC-99 (Hanson and Thornton, 1999). While the d is less than that for the SVR-20 (Boer, et al., 1997), if the sample size is great, perhaps this difference in d level is unimportant? In summary, why do I recommend a convergent approach? I suggest that a convergent approach is best at this time due to a lack of convincing data, a lack of cooperative research between ART and SCG authors, and a lack of recognition of the real task for risk assessment report writers protection of client's rights. In addition, these types of instruments are composed of different variables and not using both types in a risk assessment means that some important variables are not being considered (or at least not being given equal consideration). To use the convergent approach appropriately, one must use an appropriate ART in doing a risk assessment. Such

instruments provide a risk baseline to ensure that the clinician doing the risk assessment is given some actuarial data to ground his or her clinical prognostications. Also, the use of an appropriate SCG in the convergent approach will further ground the clinician and ensure that he or she does not run off and use variables that are unrelated to risk when discussing the individual client. In addition, the risk levels provided by two types of risk instruments can be reported separately and the clinician can use logical decision rules to convey the overall convergent risk level to the reader. These decision rules would be spelled out in the report for example, the overall risk level is low/moderate/high given his/her STATIC-99 score (which was in the low/moderate/high category) and the fact that he/she was found to be low/moderate/high on the SVR-20. Unique client factors, including concerns about psychopathy or sexual deviance, can be emphasized as reasons why the convergent (summary) risk estimate is higher than that suggested by the ART and other mitigating factors (e.g., treatment completion) as to why the convergent risk estimate is lower than that suggested by the ART being used.

References

1. Boer, D.P., Hart, S.J., Kropp, P.R., & Webster, C.D. (1997). *Manual for the Sexual Violence Risk - 20: Professional Guidelines for Assessing Risk of Sexual Violence*. Vancouver, B.C.: The Institute Against Family Violence.
2. Craig, L.A., Browne, K.D., Stringer, I., & Beech, A. (2005). Sexual Recidivism: a review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, 11, 65-84.
3. Douglas, K.S., Cox, D.N., & Webster (1999). Empirically validated violence risk assessment. *Legal and Criminological Psychology*, 4, 149-184.
4. Hanson, R.K., & Morton-Bourgon, K.E. (2004). Predictors of sexual recidivism: an updated meta-analysis. (Research Rep. No. 2004-02). Ottawa, Canada: Public Safety and Emergency Preparedness Canada.
5. Hanson, R.K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessment for sexual offenders*. Ottawa: Solicitor General of Canada (Corrections Research User Report 199-02).
6. Hart, S., Kropp, P.R., & Laws, D.R.; with Klaver, J., Logan, C, & Watt, K.A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Vancouver, B.C.: The Institute Against Family Violence.
7. Litwack, T.R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law*, 7, 409-443.
8. Quinsey, V.L., Harris, G.T., Rice, M.E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk*. Washington, D.C.: American Psychological Association.
9. de Vogel, V., de Ruiter, van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, 28, 235-251.

Author address

Douglas P. Boer, Ph.D.
 Associate Professor and Director of Clinical Psychology
 Department of Psychology
 The University of Waikato
 Private Bag 3105
 Hamilton, New Zealand
 E-mail: drdoug@waikato.ac.nz