

Combining Static-99R and STABLE-2007 risk categories: An evaluation of the five-level system for risk communication

Neil R. Hogan¹, Christine L. Sribney²

¹Integrated Threat and Risk Assessment Centre and University of Saskatchewan

²Forensic Assessment & Community Services

[Sexual Offender Treatment, Volume 14 (2019), Issue 1]

Abstract

Aim: *This study was undertaken to evaluate and quantify the potential impacts of the adoption of the Council of State Governments' five-level system for risk communication, in a community based program for the treatment of persons who have sexually offended (PSO).*

Method: *A clinical database of Static-99R and STABLE-2007 risk scores, obtained from 165 male PSO, was used to assign ordinal, categorical risk ratings, using the outgoing system and the new five-level system.*

Results: *Overall, higher ordinal risk ratings (i.e., categorical labels describing relative placement in ascending risk levels) were assigned using the new system. The differences were statistically significant, based on Wilcoxon signed-rank tests.*

Conclusions: *The results suggested that application of the new system to guide resource allocation could result in a substantial increase in resource outlay, and the number of individuals referred for interventions. Professionals may consider these results in applying the new system to their work with PSO.*

Key words: Risk assessment; risk communication; sexual offending; five-level system

The assessment of risk for sexual recidivism remains a critical component of professional practice with persons who have sexually offended. Professionals' appraisals of risk have the potential to impact important decisions (Heilbrun, Dvoskin, Hart, & McNiel, 1999) regarding sentencing, treatment, and civil commitment or preventative detention (Blais, 2015; Jung, Brown, Ennis, & Ledi, 2015). Structured methods of assessing risk for sexual and general recidivism have consistently outperformed unstructured methods in the research literature (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Meehl, 1954; Quinsey & Ambtman, 1979; Steadman & Keveles, 1972). However, structured measures constitute a heterogeneous collection of instruments. Each instrument, and each type of instrument, is associated with corresponding strengths and weaknesses.

Risk Assessment Practices

Much of the research literature on risk assessment has focused on evaluating processes of deriving overall appraisals from data. Bonta and colleagues (Bonta, 1996; Bonta & Andrews, 2007) have described risk assessment practices as evolving through successive generations based on these processes, beginning with unstructured professional judgements as the first generation. The second generation includes statistically-derived actuarial instruments, such as the Static-99R (Phenix, Fernandez, Harris, Helmus, Hanson, & Thornton, 2016). Second generation instruments rely heavily

on historical and largely static or unchangeable risk factors. While these instruments have been described as differentiating among higher and lower risk offenders with "moderate-to-large predictive accuracy" (Hanson & Morton-Bourgon, 2009, p. 6), they are often poorly suited to the identification of treatment targets. Third generation instruments are those that incorporate causal or dynamic/changeable risk factors, while maintaining the empirical basis and structured approach of second generation instruments; examples include the STABLE-2007 (Hanson, Harris, Scott, & Helmus, 2007) and the Violence Risk Scale - Sex Offender Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003). These instruments typically require higher levels of information and staff training to administer than their second generation counterparts, but may be used more directly to guide interventions. As evidence of this purported advantage, changes measured by one such instrument, the VRS-SO, have been found to predict changes in recidivism rates (Beggs & Grace, 2010; Olver, Wong, Nicholaichuk, & Gordon, 2007; Sowden & Olver, 2017). Noting that each generation of tools, and each individual tool, is associated with advantages and disadvantages, it is clear that structured instruments comprise a critical component of effective risk assessment and management practices.

Considerations in Risk Communication

While the processes of synthesizing information into appraisals of risk are certainly worthy of study, the nature of the end products, and how they are communicated to consumers, are also worthy of consideration. As argued by Heilbrun et al. (1999), the results of even high-quality risk assessments may be rendered "completely useless - or even worse than useless" (p. 94) if they are not conveyed appropriately to the end user. Examples of forms of risk communication include categorical descriptors (e.g., low, medium, high) and various statistical metrics, including percentile ranks, ascending risk bands/bins, risk ratios, and recidivism estimates. None of these forms of risk communication, whether categorical or statistical, are without limitations. Categorical rating systems and language vary among tools, similar rating systems may vary with regard to the meaning applied to common labels, and categorical risk ratings are susceptible to idiosyncratic interpretations of their meaning - even among trained forensic professionals (Hilton, Carter, Harris, & Sharpe, 2008). It is also worth noting that even highly related and well-validated instruments may produce discrepant categorical risk ratings (Jung, Pham, & Ennis, 2013). In turn, the various possible statistical and numerical indicators of risk also carry potential limitations, including varying base rates among samples, difficulties accurately defining and measuring outcome variables, and discrepancies in the definition and interpretation of their meaning. Recognition of such limitations has prompted researchers to pursue nonarbitrary metrics (Hanson, Babchishin, Helmus, & Thornton, 2013), and the developers of various risk instruments provide multiple recommended options for communicating risk. Having acknowledged the problem of arbitrariness in risk communication in general, this study focused on categorical risk labels or appraisals.

A Five Level System for Risk Communication

In part to address the problems and inconsistencies associated with idiosyncratic and arbitrary risk communication identified above, Hanson, Bourgon, McGrath, Kroner, D'Amora, Thomas, and Tavaréz (2016), on behalf of the Council of State Governments' Justice Centre, developed a five-level system for communicating about risk for offending. This system was designed with risk for general offending (including, but not limited to sexual offending) in mind, and it is intended to provide a common language for communicating the results of risk assessments and to be applicable across instruments. Briefly, the ascending levels may be summarized as follows (Hanson, Babchishin, Helmus, Thornton, & Phenix, 2017): Level I offenders are "generally prosocial" (p. 6) individuals who demonstrate few criminogenic needs, and are not at elevated risk of recidivism relative to non-offending individuals; Level II offenders have a limited number of criminogenic needs

and are at elevated risk relative to non-offending individuals, but are "lower risk than typical offenders" (p. 6); Level III offenders are "typical offenders" (p. 6), with an average number of criminogenic needs, and who require intervention to manage their risk; Level IV offenders have a large number of criminogenic needs, and are at higher risk than the average offender; and Level V offenders represent the highest risk individuals, and who are "virtually certain to reoffend" (p. 6).

While the five-level system was designed with general offending in mind, it has now been adapted and applied to risk for sexual offending, as measured by the Static-99R and Static-2002R (Hanson et al., 2017), the STABLE-2007 (Brankley, Helmus, & Hanson, 2017), and the VRS-SO (Olver et al., 2018). In recognition of the fact that documented recidivism rates among persons who have sexually offended do not appear to meet the threshold for the Level V label, the respective developers applied the labels Level IVa (i.e., above average) and Level IVb (i.e., well above average) to the two highest risk categories. Also common across the instruments, Level III includes the median score, and applies to approximately half of the offenders in the normative samples.

The Present Study

Given that the five-level system is a relatively new development, and its application to persons who have sexually offended is newer still, empirical study of its implications and impacts is needed. Indeed, the Static-99R & Static-2002R Evaluators' Workbook (Phenix, Helmus, & Hanson, 2016) contains explicit acknowledgement of the fact that the updated risk levels may not be equally well-suited to the needs of all professional and clinical decision makers in all settings, or to all professional activities (e.g., allocating treatment resources). This exploratory study was undertaken to evaluate and quantify the practical impacts and implications of the adoption of the five-level system, using the Static-99R and STABLE-2007, in a community based program for the treatment of persons who have sexually offended. In particular, existing scores were interpreted using both the former system of categorical labels and the new five-level system, to evaluate the nature of any changes and their potential implications for resource allocation.

Method

Participants and Program

Data were obtained from an archival database of clinical data, maintained by a publicly funded, community based forensic mental health clinic in Western Canada. The sample was comprised of 165 males who were referred for an assessment for treatment, on the basis of a criminal charge or conviction for a sexually-motivated offence.

The program serves as the primary community-based service-provider for probation and parole services in the region. Referrals are received from these agencies based on a mandated treatment condition and have not been selected for referral based on risk level. As a result, this sample is representative of all risk levels (i.e., the lowest risk offenders are not excluded from a referral to the clinic). All consecutive referrals for which the Static-99R could be scored based on the eligibility criteria in the coding manual (Phenix, Fernandez, Harris, Helmus, Hanson, & Thornton, 2016) between 2015 and 2018 were included. These dates reflect the program's adoption of the Static-99R and STABLE-2007 for clinical decision-making purposes. Mean age on the date of assessment was 40.0 years ($SD = 14.8$; range = 19.0 to 78.2). For 72.1% of the sample, the index offence(s) included contact or hands-on offending.

Materials

Static-99R. The Static-99R (Phenix et al., 2016) is a 10-item actuarial risk measure, designed for use among persons who have received sanctions (e.g., an arrest, charge, or conviction) for sexual offences. It is comprised of items pertaining to sexual and nonsexual offending history and offender and victim demographics that have been found to correlate with sexual recidivism in adult male sex offenders. When all 10-items are summed, a total risk score is provided that can range from -3 to 12. While the instrument previously utilized a four-level system of risk categories (i.e., Low, Moderate-Low, Moderate-High, and High), it now utilizes the five-level system, as described above. Specifically, the total score corresponds to one of the five risk categories as follows: Very Low (-3, -2), Below Average (-1, 0), Average (1, 2, 3), Above Average (4, 5) and Well Above Average (6 and higher). The instrument is widely used by forensic professionals (Archer, Buffington-Vollum, & Stredny, 2006) and meta-analytic results have supported its predictive validity (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012).

STABLE-2007. The STABLE-2007 (Hanson et al., 2007) is a tool designed to measure dynamic sexual offending risk factors (i.e., those factors that can change with time, circumstances, and intervention). This measure is comprised of 13-items, that when summed provide a total score that can range from 0 to 26. For individuals who do not have child victims (defined as at least one victim under the age of 14 years old), the emotional identification with children item is omitted, and the scale is subsequently scored out of 24 possible points. Across these items, the main areas assessed are: significant social influences, intimacy deficits, general self-regulation, sexual self-regulation, and cooperation with supervision. The STABLE-2007 can be used to identify individualized treatment targets and priorities based on empirical data. Scores on the STABLE-2007 have been found to predict sexual recidivism (e.g., Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012; Etzler, Eher, & Rettenberger, 2018; Helmus, Babchishin, & Blais, 2012), although it is worth noting that Etzler and colleagues also found that STABLE-2007 total scores did not incrementally predict sexual recidivism, controlling for Static-99 scores. STABLE-2007 scores, used in isolation, correspond to a three-level system of risk categorization: Low (0 to 3), Moderate (4 to 11) and High (12 and higher). When combined with Static-99R scores, the combined/overall appraisal is derived from the five-level system.

Procedure

Ethical and administrative reviews of this study were conducted by a Western Canadian university and the provincial health authority. As mentioned briefly above, all data were obtained from a clinical database, prepared and maintained for the purpose of program evaluation. As such, the scores reflect real-world application of the instruments. That is, ratings were completed by members of an interdisciplinary team comprised of psychologists, psychiatrists, social workers, and nurses, on the basis of clinical interviews and reviews of official documentation (e.g., criminal records). All clinicians that conducted assessments as part of their clinical work had previously received certified training on the administration of the instruments. Given that each score was produced by an individual clinician as part of routine practice, interrater reliability data were not obtained.

Data Analytic Plan

Due to the fact that the program endeavours to employ systematic and nonarbitrary methods of allocating treatment resources, the primary goal of this research was to evaluate how the distribution of treatment programming might change, should the new five-level system be used in place of the outgoing system of nominal risk categories. As such, we conducted exploratory and

descriptive analyses focusing on the distribution of nominal categories for the Static-99R and for combined Static-99R/STABLE-2007 ratings, using the outgoing and new systems. STABLE-2007 ratings were not evaluated separately from the combined risk ratings, given that the five-level system has not been applied to this tool independent of other metrics. Consistent with Jung and colleagues' (2013) study, we undertook paired comparisons between the ordinal risk levels using percentage agreement. In order to compute percentage agreement with the outgoing four-level system of the Static-99R, we collapsed Levels IVa and Level IVb of the new system, given that they are part of the same level in the original five-level system. We also conducted Wilcoxon signed-rank tests to assess differences in risk level classifications based on the Static-99R and based on the combined Static-99R/STABLE-2007 ratings. This method is appropriate for nonparametric and matched data (Howell, 2013), and has been employed in comparable research comparing risk labels applied to persons who have sexually offended (e.g., Gentry, Dulmus, & Theriot, 2005).

Results

The mean Static-99R score for the sample was 2.72 ($SD = 2.43$), and the mean STABLE-2007 score was 7.68 ($SD = 4.55$), suggesting that the overall risk profile was broadly consistent with the risk profiles of the normative samples associated with the respective tools (Brankley et al., 2017; Hanson et al., 2017).

Figures 1 and 2 depict the distribution of Static-99R nominal categories using the outgoing and new five-level systems, respectively. Based on the older system, the lowest risk category contained the largest number of cases ($n = 51$ or 31%), and the numbers of cases falling into progressively higher risk categories follow a decreasing trend. In contrast, data from the five-level system approximated a normal distribution, with the lowest category containing the smallest number of cases ($n = 8$ or 5%), and the middle category containing the largest number of cases ($n = 67$ or 41%). A comprehensive summary of risk level frequencies and percentage agreement using the Static-99R is presented in Table 1. Using paired comparisons, a total of 137 cases (84%), the majority of the sample, were assigned a higher ordinal risk level using the new five-level system, while 27 cases (16%) were assigned the same level, and no cases were assigned a lower level. Based on the paired Wilcoxon signed-rank test, the observed tendency for the new system to assign higher risk levels was statistically significant ($Z = -11.78, p < 0.001$).

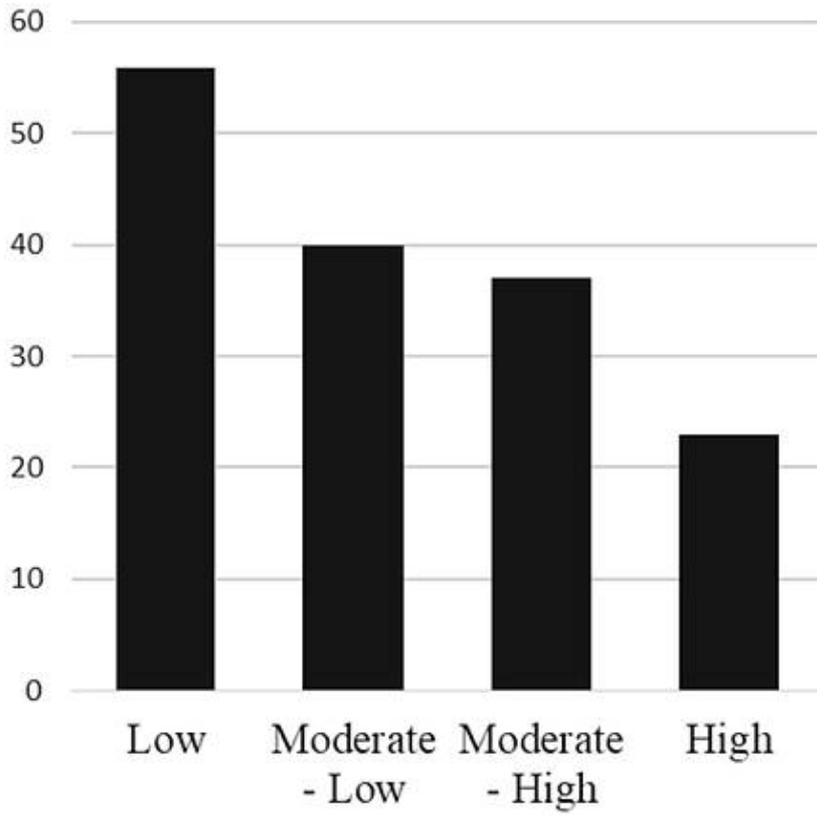


Figure 1: Histogram of risk level assignment based on Static-99R scores using the outgoing system of risk labels

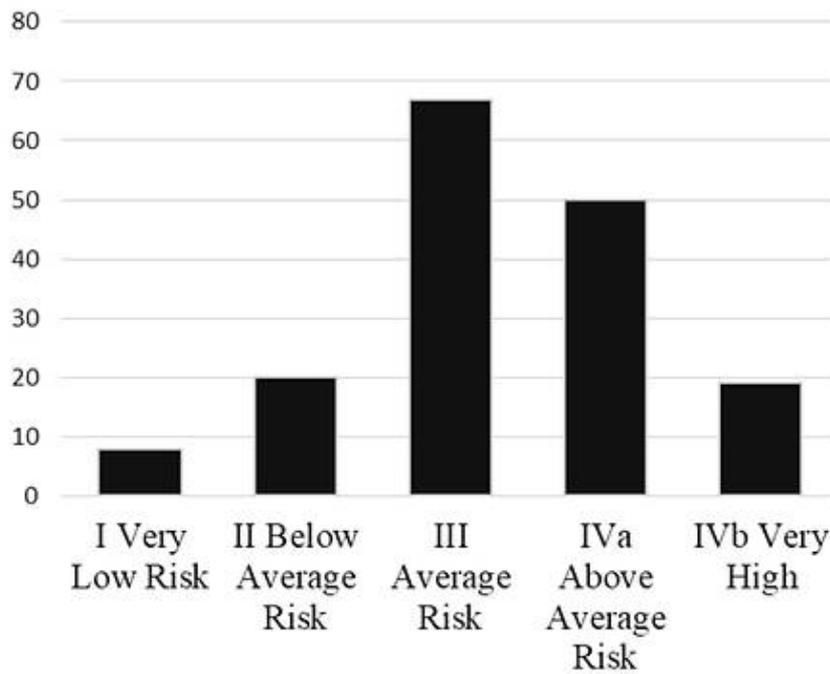


Figure 2: Histogram of risk level assignment based on Static-99R scores using the five-level system of risk labels

Table 1: Percentage agreement between ordinal risk categories assigned by two systems for Static-99R risk ratings (N = 164)

Outgoing system	New five-level system				Percentage agreement
	Level I	Level II	Level III	Level IVa/b	
Low	8	20	23	0	15.7
Moderate-Low	0	0	44	0	0.0
Moderate-High	0	0	0	50	0.0
High	0	0	0	19	100.0
Percentage agreement	100.0	0.0	0.0	27.5	16.5

Moving to the combined Static-99R and STABLE-2007 combined risk ratings, Figure 3 provides a visual representation of the distribution of combined nominal risk categories using the outgoing system. Figure 4 provides a visual representation of the distribution of combined nominal risk categories using the new five-level system. As with Static-99R scores, using the older system the lowest risk category contained the largest number of cases ($n = 56$ or 34%), and the number of cases falling into progressively higher risk categories follow a decreasing trend. Data from the five-level system approximated a normal distribution, with the lowest category containing the smallest number of cases ($n = 7$ or 4%), and the middle category containing the largest number of cases ($n = 64$ or 39%). A comprehensive summary of risk level frequencies and percentage agreement is presented in Table 2. Using paired comparisons, a total of 140 cases (85%), the majority of the sample, were assigned a higher risk level using the new five-level system, while 24 cases (15%) were assigned the same level, and only 1 case was assigned a lower level. Based on the paired Wilcoxon signed-rank test, the observed tendency for the new system to assign higher risk levels was statistically significant ($Z = -10.91$, $p < 0.001$).

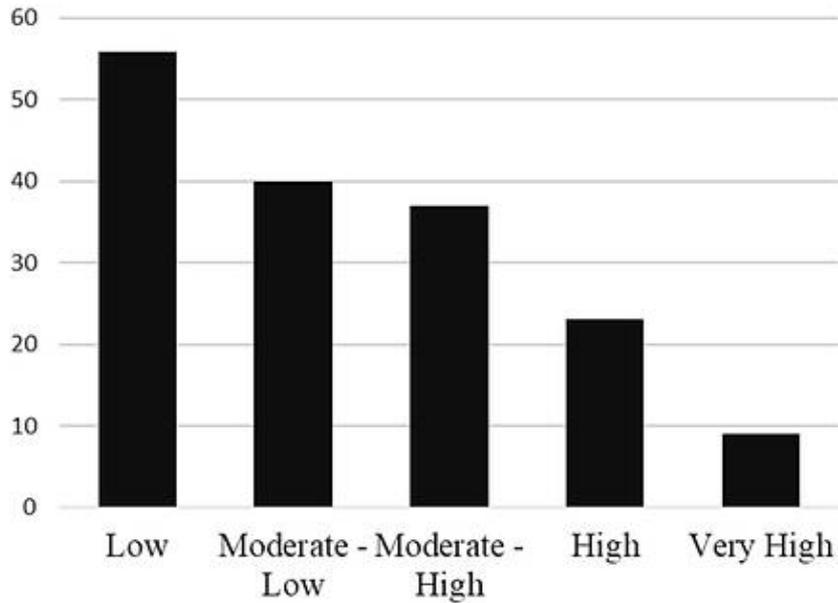


Figure 3: Histogram of risk level assignment based on combined Static-99R/STABLE-2007 scores using the outgoing system of risk labels

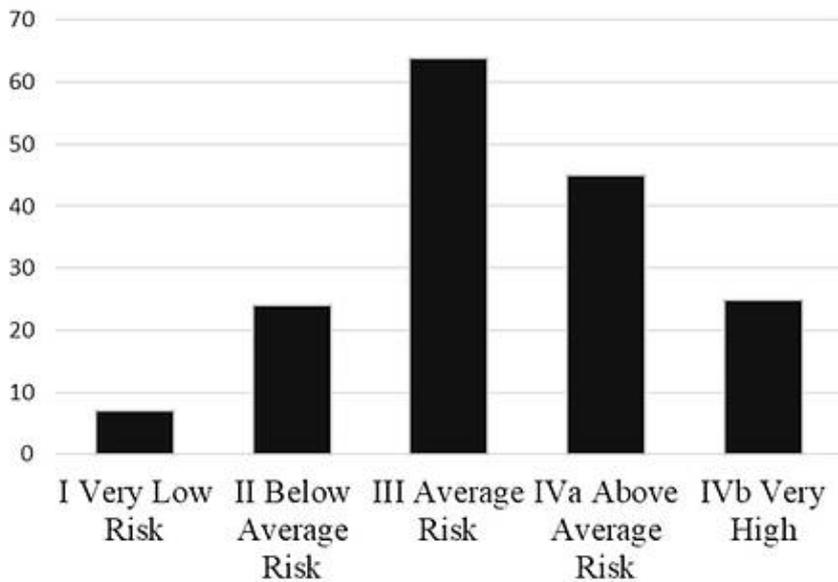


Figure 4: Histogram of risk level assignment based on combined Static-99R/STABLE-2007 scores using the five-level system of risk labels

Table 2: Percentage agreement between ordinal risk categories assigned by two systems for combined Static-99R/STABLE-2007 risk ratings (N = 165)

Outgoing system	New five-level system					Percentage agreement
	Level I	Level II	Level III	Level IVa	Level IVb	

Low	7	24	25	0	0	12.5
Moderate-Low	0	0	37	2	1	0.0
Moderate-High	0	0	1	36	0	2.7
High	0	0	1	7	15	34.8
Very High	0	0	0	0	9	100.0
Percentage agreement	100.0	0.0	1.5	12.7	43.3	14.5

Discussion

Various authors have reported on limitations and inconsistencies associated with existing categorical risk ratings, which have been described as "fairly arbitrary and ill-defined" (Helmus, 2018, p. 3). The current writers believe that the five-level system, which is intended to provide a common language for risk communication, represents a significant advancement towards addressing such concerns. That said, we believe that it would be prudent for professionals adopting the new system to be as informed as possible of its implications for their practice. As such, this study examined the practical implications of the adoption of the new five-level system for risk communication, as applied to Static-99R and combined Static-99R/STABLE-2007 risk ratings. Using a single set of raw risk scores obtained from a clinical database, risk level assignments from both systems were compared. The results suggested that, overall, the new system resulted in the assignment of higher ordinal risk level ratings, when compared to the outgoing system of risk categories.

This study is at least somewhat unique in that, unlike studies of risk appraisals derived from different risk tools (e.g., Gentry et al., 2005; Jung et al.; 2013) or of risk appraisals derived from different raters using the same tool (Howell, Christofferson, & Olver, 2017; Murrie, Boccaccini, Guarnera, & Rufino, 2013), it compared categorical risk appraisals derived from a single set of scores. Thus, many of the proposed explanations for discrepancies in these other research paradigms, such as simple error, structural differences among tools, or rater bias, are not applicable. Any and all observed differences among risk ratings are directly attributable to the adoption of the new system. Another unique aspect of the issues explored in this study is the fact that discrepancies between the two systems are not inherently problematic, per se. The five-level system constitutes a new approach to assigning risk levels, and is intentionally different from prior approaches. The nature and extent of the impacts of this advancement, both positive and negative, largely remain to be determined. That being said, Etzler and colleagues (2018) observed that STABLE-2007 risk categories, but not total scores, demonstrated incremental predictive validity, controlling for Static-99 risk categories. By employing similar paradigms, future studies may assess whether new risk categories based on the five-level system serve to maintain, enhance, or degrade the predictive validity of prior systems.

As mentioned previously, the developers of some tools, such as the Static-99R (Phenix et al., 2016), have explicitly acknowledged that the new system may not be appropriate in all settings, for all purposes (e.g., allocating finite intervention resources to a predetermined percentage/number of offenders). Nonetheless, for those professionals wishing to determine whether, and how, to adopt the new system in their work with persons who have sexually offended, the results from this study may provide some important points for consideration. The results are particularly relevant to situations in which resources are allocated on the basis of risk level, consistent with the Risk-Need-Responsivity model (RNR; Andrews, Bonta, & Hoge, 1990).

The authors wish to acknowledge that some aspects of the decision to use ordinal risk ratings in this study, and the manner in which we defined and compared the categories, could be debated. After all, the levels constituting the new system were not designed to correspond directly to the old. Similarly, one could reasonably argue that the manner in which institutions apply categorical risk labels to their services will vary, and (hopefully) adapt with changing technology. That said, some of the discrepancies that our methods have identified are less debatable. Research suggests that interventions offered to persons who have sexually offended are most effective when they follow the RNR model (Hanson, Bourgon, Helmus, & Hodgson, 2009), by ensuring that, among other considerations, professionals provide "intensive interventions to higher risk offenders and little or no service to low risk offenders" (p. 871). Given that Hanson and colleagues' landmark meta-analysis was published prior to the development of the five-level system, it follows that persons assigned a risk rating of Low by the Static-99R, or by a combination of the Static-99R and STABLE-2007, might reasonably have been offered relatively few services, or been excluded from treatment or correctional interventions, with empirical justification. However, our data indicate that approximately 45% of those individuals assigned a Low risk rating (the lowest of five categories) using the old system would now be assigned a Level III rating (the third of five possible categories); based on the Static-99R & Static-2002R Evaluators' Workbook (Phenix et al., 2016), Level III offenders are considered to "have criminogenic needs in several areas, and require meaningful investments in structured programming to decrease their recidivism risk" (p. 4). While we consider these discrepancies to be clinically significant in their own right, our data do not shed light on the question of which of the discrepant labels are more valid or useful.

Overall, it appears that the adoption of the new system has at least the potential to contribute to considerable increases in resource outlay, and in the number of individuals recommended to receive formal programming, notwithstanding variability in approaches among services, settings, and professionals. The potential impact on services for the lowest risk individuals is particularly notable given that research suggests that providing excessive services to low risk offenders may actually increase recidivism (e.g., Andrews et al., 1990; Lovins, Lowenkamp, & Latessa, 2009). While the reasons for this phenomenon are not known with certainty, potential contributing factors identified in the literature include disruption of opportunities for prosocial activities and relationships, and opportunities for criminogenic social learning created by exposure to higher risk individuals (Lovins et al., 2009). On a related note, Bourgon, Mugford, Hanson, and Coligado (2018) recently argued that discrepancies between risk assessment tools could result in unfair and inequitable treatment of individuals, thereby undermining a "cornerstone" (p. 176) of the criminal justice system. Similarly, part of Hanson and colleagues' (2016) justification for the development of the five-level system was the observation that "the same person can be described by different categories across different assessment instruments" (p. 3). While adoption of the new system may represent a necessary step towards increased precision in risk communication, professionals should be aware that the adoption of the new system of risk categories could result in discrepant risk ratings, and discrepant treatment recommendations, for two individuals receiving the same score from the same instruments. It is also worth noting that while the preceding discussion focuses on treatment resources, it is expected that similar issues may be observed in any other domain in which risk appraisals are used to inform criminal justice processes, such as sentencing or detention decisions (Blais, 2015; Jung et al., 2015). Thus, as the literature develops, incorporation of the new system should be undertaken with care.

In the absence of research specifically examining outcomes (e.g., recidivism) related to use of the five-level system to determine level of service, professionals are left to weigh other factors in determining whether to adopt the new model. One argument in favor of the new system is that it ostensibly provides clearer practical direction than alternative systems. In particular, category assignments based on the five-level system are, by definition, intended to reflect the underlying

"density of criminogenic needs" (Brankley et al., 2017, p. 6) of the offender, and therefore may have more direct implications for treatment than alternatives. While theoretically informed and nonarbitrary risk categories are certainly desirable, care should be taken to ensure that appropriate tools are used to assign such categories. It is tempting to make inferences about *why* a second generation instrument, such as the Static-99R, predicts recidivism, but any such inferences are speculative (Hilton, Harris, & Rice, 2010). In light of their basis in both theory and empirical data, third generation instruments, such as the STABLE-2007 or the VRS-SO, are likely best suited to the five-level system, at least insofar as they inform conclusions about densities of criminogenic needs. Another consideration is whether, and how, the five-level system will apply to risk for sexual recidivism specifically. The system was designed with general offending (i.e., including, but limited to sexual offending) in mind, and while many of the risk factors that predict general offending also predict sexual offending (Hanson & Morton-Bourgon, 2009), other considerations, such as deviant sexual interests, apply differentially to sexual offending risk (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005). As such, the generic descriptors associated with each level may not apply in an equivalent manner across subtypes of offending.

Some study limitations are worth noting. As an archival study, the data was limited to that which could be gleaned from the available files. In particular, this resulted in limited information with regard to demographic information; that said, the results of the study highlighted matters that are likely notable regardless of the idiosyncrasies of the particular sample. The study is also relatively limited in scope, in that it focuses on a narrow issue, using two commonly used instruments. We believe that it would be worthwhile to explore this issue with other tools, such as the VRS-SO, that have adopted the five-level system. Of particular interest for further study, would be the impacts of the new system on studies of treatment/intervention outcomes (e.g., re-offending rates) amongst programs that adopt the system to allocate resources. Given that the system is a relatively new development, we expect that such research will take time. In the meantime, it is hoped that the current results provide useful information for professionals working towards reducing sexual offending.

Conclusion

The five-level system for risk communication (Hanson et al., 2016) was developed to mitigate the negative consequences of arbitrary risk communication. It has now been adapted and incorporated into existing risk tools for sexual offending behaviour. Results of the current study suggested that this system, as adapted for use with the Static-99R and STABLE-2007, has the potential to increase ordinal/categorical risk ratings, and thereby increase the level of service recommended for populations of persons who have sexually offended. Thus, care should be taken in applying the new system to resource allocation, and consistent with Phenix and colleagues' (2016) recommendations, professionals may consider metrics other than the risk categories for specific resource allocation decisions. That said, if applied in an informed and judicious manner, we agree with Helmus' (2018) assertion that this system has the potential to constitute a significant advancement in risk communication.

Acknowledgements

The authors wish to acknowledge Anne Walley from MacEwan University for her assistance with data collection. An acknowledgement is also due to Dr. Gabriela Corabian for her review of and helpful comments on an early version of this article. The authors take responsibility for the integrity of the data, the accuracy of the data analyses, and have made every effort to avoid inflating statistically significant results.

References

1. Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17, 19-52.
2. Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 84-94.
3. Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment*, 22(2), 234-251.
4. Blais, J. (2015). Preventative detention decisions: Reliance on expert assessments and evidence of partisan allegiance within the Canadian context. *Behavioral Sciences & the Law*, 33(1), 74-91.
5. Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.
6. Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation (Corrections Research User Report No. 2007-06). Ottawa, ON: Public Safety Canada.
7. Bourgon, G., Mugford, R., Hanson, R., & Coligado, M. (2018). Offender risk assessment practices vary across Canada. *Canadian Journal of Criminology and Criminal Justice*, 60(2), 167-205.
8. Brankley, A. E., Helmus, L. M., & Hanson, R. K. (2017). *STABLE-2007 Evaluator Workbook - Revised 2017*. Ottawa, ON.
9. Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment*, 24, 5-28. doi:10.1177/1079063211403164
10. Etzler, S., Rettenberger, M., & Eher, R. (2018). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the Stable-2007. *Assessment*. Advance online publication. doi:10.1177/1073191118754705
11. Gentry, A. L., Dulmus, C. N., & Theriot, M. T. (2005). Comparing sex offender risk classification using the Static-99 and LSI-R assessment instruments. *Research on Social Work Practice*, 15(6), 557-563.
12. 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(1), 19-30.
13. Hanson, R. K., Babchishin, K. M., Helmus, L. M., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse: A Journal of Research and Treatment*, 25(5), 482-515.
14. Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment*, 29, 582-597. [dx.doi.org/10.1037/pas0000371](https://doi.org/10.1037/pas0000371)
15. Hanson, R. K., Bourgon, G., Helmus, L. M., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, 36(9), 865-891.
16. Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavaréz, L. P. (2016). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. Washington, DC: Justice Center Council of State Governments.

17. Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
18. Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. M. (2007). Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (User Report No. 2007-05). Ottawa, Ontario: Public Safety Canada. Retrieved from www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/index-eng.aspx
19. 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, 1154-1164.
20. 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.
21. Heilbrun, K., Dvoskin, J., Hart, S., & McNeil, D. (1999). Violence risk communication: Implications for research, policy, and practice. *Health, Risk & Society*, 1(1), 91-105.
22. Helmus, L. M. (2018). Sex offender risk assessment: Where are we where are we going? *Current Psychiatry Reports*, 20(6), 46. <https://doi.org/10.1007/s11920-018-0909-8>.
23. Helmus, L. M., Babchishin, K.M., & Blais, J. (2012). Predictive accuracy of dynamic risk factors for Aboriginal and Non-Aboriginal sex offenders: An exploratory comparison using STABLE-2007. *International Journal of Offender Therapy and Comparative Criminology*, 56, 856-876. doi:10.1177/0306624X11414693
24. Helmus, L. M., 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. doi:10.1177/0093854812443648
25. Hilton, Z. N., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, 23(2), 171-188. dx.doi.org/10.1177/0886260507309337
26. Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). Risk assessment for domestically violent men: Tools for criminal justice, offender intervention, and victim services. Washington, DC: American Psychological Association.
27. Howell, D. C. (2013). *Statistical methods for psychology* (5th ed.). Belmont, CA: Wadsworth.
28. Howell, M. J., Christofferson, S. M. B., & Olver, M. E. (2017). Evaluating the inter-rater reliability of the Violence Risk Scale-Sexual Offense Version (VRS-SO) in a community-based treatment setting. *Sexual Offender Treatment*, 12(1), 1-14. Retrieved from www.sexual-offender-treatment.org/160.html
29. Jung, S., Brown, K., Ennis, L., & Ledi, D. (2015). The association between presentence risk evaluations and sentencing outcome. *Applied Psychology in Criminal Justice*, 11(2), 111-125.
30. Jung, S., Pham, A., & Ennis, L. (2013). Measuring the disparity of categorical risk among various sex offender risk assessment measures. *Journal of Forensic Psychiatry & Psychology*, 24(3), 353-370. dx.doi.org/10.1080/14789949.2013.806567
31. Lovins, B., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse? *The Prison Journal*, 89(3), 344-357.
32. Meehl, P. E. (1954). *Clinical vs. statistical prediction*. Minneapolis, MN: University of Minnesota Press.
33. 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(10), 1889-1897. <https://doi.org/10.1177/0956797613481812>

34. Olver, M. E., Mundt, J. C., Thornton, D., Beggs Christofferson, S. M., Kingston, D. A., Sowden, J. N., ... & Wong, S. C. (2018). Using the Violence Risk Scale-Sexual Offense version in sexual violence risk assessments: Updated risk categories and recidivism estimates from a multisite sample of treated sexual offenders. *Psychological Assessment*, 30(7), 941-955. [dx.doi.org/10.1037/pas0000538](https://doi.org/10.1037/pas0000538)
35. Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale- Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, 19(3), 318-329. doi:10.1037/1040-3590.19.3.318
36. Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, L. M., Hanson, R. K., & Thornton, D. (2016). Static-99R coding rules Revised - 2016. Ottawa, ON: Department of the Solicitor General of Canada.
37. Phenix, A., Helmus, L. M., & Hanson, R. K. (2016). Static-99R & Static 2002R: Evaluators' workbook. [Unpublished manual]. Retrieved from www.static99.org
38. Quinsey, V. L., & Ambtman, R. (1979). Variables affecting psychiatrists' and teachers' assessments of the dangerousness of mentally ill offenders. *Journal of Consulting and Clinical Psychology*, 47(2), 353-362. [dx.doi.org/10.1037/0022-006X.47.2.353](https://doi.org/10.1037/0022-006X.47.2.353)
39. Sowden, J. N., & Olver, M. E. (2017). Use of the Violence Risk Scale-Sexual Offender Version and the Stable 2007 to assess dynamic sexual violence risk in a sample of treated sexual offenders. *Psychological Assessment*, 29(3), 293-303. doi:10.1037/pas0000345
40. Steadman, H., & Keveles, G. (1972). The community adjustment and criminal activity of the Baxstrom patients: 1966-1970. *The American Journal of Psychiatry*, 129, 304-310. <https://doi.org/10.1176/ajp.129.3.304>
41. Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). The Violence Risk Scale-Sexual Offender version (VRS-SO). Saskatoon, SK: Regional Psychiatric Centre and University of Saskatchewan.

Author address

Neil R. Hogan

Integrated Threat and Risk Assessment Centre

ALERT West Campus

Edmonton, Alberta, Canada

neil.hogan@usask.ca