Critical considerations in the development and interpretation of common risk language

Neil R. Hogan^{a,b}

^a*Integrated Threat and Risk Assessment Centre, Edmonton, AB, Canada;* ^b*University of Saskatchewan, Saskatoon, SK, Canada*

Existing risk communication procedures are marred by various well-documented problems and inconsistencies. The Council of State Governments' Justice Center (United States) developed a five-level system for risk and needs communication, to standardize these procedures and to provide a common risk language. Introduction of a common language could constitute a dramatic shift in criminal justice processes, with wide-ranging impacts. This article provides a critical review of the system and its suitability for application to various risk assessment functions. Issues discussed include: applicability to specialist and generalist offending behavior, the characteristics of suitable instruments, statistical and conceptual priorities, barriers to precision in language, and conceptual issues related to changes in risk level. A thorough understanding of each of these issues is necessary to apply the system to new contexts and populations, and facilitate straightforward and precise risk communication. Absent further elaboration of the system, many problems with risk communication will persist.

Key words: actuarial; categorical risk labels; five-level system; offender management; offender rehabilitation; risk assessment; risk communication; structured professional judgement.

In forensic psychiatry and psychology, the practice of appraising and communicating risk remains an activity of significant interest, with wide-ranging impacts. Applications of risk assessment to decision-making in various branches of the criminal justice system may impact the public in various direct and indirect ways, such as influencing citizens' safety and wellbeing, and through the practical and financial consequences (Monahan & Skeem, 2014) of resource allocation. Impacts of risk assessments on the individuals personally subjected to them are also wide-ranging. Risk assessments are applied at various stages in criminal justice processes. At early stages, they may influence case prioritization among law

enforcement agencies (Jung & Buro, 2017; Storey, Kropp, Hart, Belfrage, & Strand, 2014), or decisions granting or denying access to diversion programming among youth (Wylie, Clinkinbeard, & Hobbs, 2019). With regard to later stage criminal justice system processes, evidence suggests that risk assessment results presented by professionals may influence the nature and extent of sentences imposed by the judiciary (Jung, Ennis, Brown, & Ledi, 2015), as well as decisions regarding civil commitment under sexually violent predator legislation (Levenson & Morin, 2006) or post-sentence detention in the United States and Canada (Hanson, 2005). Of the myriad decisions potentially influenced by a

Correspondence: Neil R. Hogan, Integrated Threat and Risk Assessment Centre, Edmonton, AB, Canada. Email: neil.hogan@usask.ca

formal risk assessment, few, if any, can equal the imposition of capital punishment (Claussen-Schulz, Pearce, & Schopp, 2004) in terms of its significance to the subject of the assessment. In light of the gravity of such decisions, it is incumbent upon forensic mental health professionals to rely on valid and reliable techniques, and to convey risk information, particularly to non-expert consumers, in a prudent and deliberate manner. While this paper primarily critically analyzes a new system for communicating risk information, it will begin with a review of the foundational risk assessment concepts that informed this analysis.

The what and how of contemporary risk assessment

In the parlance of forensic psychiatry and psychology, risk information is largely defined by the techniques and practices used to produce it. Thus, to understand contemporary risk assessment technology, it can be helpful to consider the context of history. To aid in this regard, Bonta and colleagues (Bonta, 1996; Bonta & Andrews, 2007) have distinguished among generations of risk assessment practices. The first generation they described comprises unstructured professional judgements, employing idiosyncratic criteria, such as the clinical experience of the assessor. The second generation of risk assessment practices comprises structured actuarial risk instruments, developed based on statistical relationships between predictors and recidivism. These instruments are typically atheoretical, and are largely comprised of predictors that are static, or are not amenable to change. Examples include the Static-99R (Phenix, Fernandez, et al., 2016) and the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993). Third-generation practices also comprise structured instruments, but these instruments are specifically designed to include dynamic risk factors that are theoretically amenable to change. Examples include the Level of Service Inventory-Revised

(LSI-R; Andrews & Bonta, 1995) and the STABLE-2007 (Hanson, Harris, Scott, & Helmus, 2007). The fourth generation of risk assessment is the most recent, and comprises instruments incorporating both dynamic risk factors and an explicit case management component. According to Andrews, Bonta, & Wormith (2006), these instruments provide a structured mechanism for users to monitor cases and facilitate rehabilitation. Examples include the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004), the Violence Risk Scale (VRS; Wong & Gordon, 1999–2003), and the Violence Risk Scale–Sexual Offence Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003–2017).

Moving beyond this generational classification system, which is based primarily on methods of deriving item content, risk instruments may also be differentiated based on other theoretical approaches and technical considerations. One important consideration involves the manner in which test users arrive at a final appraisal or summary of risk. Some instruments, such as the LSI-R and Static-99R, combine risk ratings using predetermined algorithms. These types of instrument emphasize standardization and reliability, and developers often provide users with actuarial tables, containing recidivism information corresponding to numerical scores. In contrast, tools with a foundation in the structured professional judgement (SPJ) approach, such as the Historical–Clinical–Risk Management-20 (HCR-20^{V3}; Douglas, Hart, Webster, & Belfrage, 2013), provide no such algorithms or guidelines. SPJ guidelines typically direct users to consider and evaluate, at minimum, a predetermined number of operationalized risk factors. However, users ultimately arrive at an overall determination of risk or case prioritization based on their professional discretion. SPJ instruments are generally categorized among third- or fourth-generation instruments (Campbell, French, & Gendreau, 2009; Olver & Wong, 2019) based

on their structure and theoretically derived content, notwithstanding the technical divergence from their actuarial counterparts. It is worth noting that while actuarial and SPJ approaches clearly differ in important respects, the manner of communicating their results often belies these differences (e.g. summary statements of 'high risk').

Risk communication

Two decades ago, Heilbrun, Dvoskin, Hart, and McNiel (1999) argued that unless an assessor can convey his or her findings to end users in an appropriate manner, an otherwise defensible risk assessment may confer little benefit, and may even cause harm. Thus, risk communication is critical to the process of risk assessment. Options for risk communication take many forms, ranging from nominal or categorical descriptors, to visual depictions or illustrations (Hilton, Harris, & Rice, 2010), to statistical metrics of varying complexity. Many risk tools confer one or more nominal descriptors (e.g. low, moderate or high) to the subject's overall risk for recidivism, priority level for intervention resources and/or the severity of potential harm. Common statistical metrics applied to risk communication include: ascending risk bins or bands, percentile ranks, risk or hazard ratios and recidivism estimates.

All forms of nominal or statistical risk communication have limitations, and may be more or less desirable depending on the context and circumstances of a given assessment. With regard to categorical labels, commonly cited limitations include inconsistencies in language and definitions among tools, as well as inconsistencies in interpretation of shared terminology, even among experts (Hilton, Carter, Harris, & Sharpe, 2008). Furthermore, these labels appear unreliable across instruments, as evidenced by discrepant labels applied to the same individual (Barbaree, Langton, & Peacock, 2006; Jung, Pham, & Ennis, 2013). Scurich (2018) recently offered further criticisms of categorical risk communication,

and went as far as arguing that forensic professionals should end the practice altogether. Statistical metrics, on the other hand, are impacted by such issues as: inconsistencies in selecting and defining appropriate outcome variables, the impacts of varying base rates on criterion-referenced tools (e.g. Helmus, Hanson, Thornton, Babchishin, & Harris, 2012) and idiosyncratic interpretations of the meaning of various metrics. Such limitations have motivated efforts to develop non-arbitrary risk metrics.

The five-level system for risk communication

In recognition of the concerns noted above, the Council of State Governments' Justice Center in the United States developed a five-level system for risk communication (Hanson, Bourgon, et al., 2017). The five-level system is intended to improve the precision and reliability of risk communication by providing a common language, which should allow assessors to convey information about both the individual's probability of recidivism and his or her particular risk/need profile. The system is also intended to accommodate changes in risk, a function that some, but not all, previous systems have explicitly incorporated. The full extent to which the system has been, or will be, applied across criminal justice systems is unknown at the time of writing. However, the potential for widespread usage is evidenced by its application to the Static-99R, which, along with its variants, was identified by Neal and Grisso (2014) as the most commonly used risk assessment instrument for sexual recidivism among an international sample of forensic psychologists and psychiatrists from Canada, the United States, the United Kingdom, Australia and Europe.

Hanson, Babchishin, Helmus, Thornton, and Phenix (2017) described the five levels as follows: Level I comprises 'generally prosocial' (p. 6) persons, demonstrating few criminogenic needs, who are not at elevated risk of

recidivism compared to non-offending persons; Level II comprises persons with limited numbers of needs, who are at elevated risk compared to non-offending persons, but at lower risk than 'typical offenders' (p. 6); Level III comprises 'typical offenders' (p. 6), demonstrating an average number of needs, which require intervention; Level IV comprises individuals demonstrating many criminogenic needs, who are at higher risk than the average offender; and Level V comprises the highest risk individuals, and who are considered 'virtually certain to reoffend' (p. 6). In addition to these qualitative descriptions of an individual's applicable risk factors, each level also corresponds to expected recidivism rates, notwithstanding the developers' acknowledgement that these rates may be subject to change, with further study.

Current review

The development of a common language for risk assessment, whether it be through adoption of the five-level system or a comparable alternative, could be revolutionary in the criminal justice system. As argued by Zapf and Dror (2017), language can profoundly influence how information is perceived, and ambiguous language represents a critical source of potential bias in forensic evaluations. Increasing the reliability of both the application and interpretation of risk information could substantially improve efforts to ensure fair and equitable treatment of persons involved in criminal justice processes (Bourgon, Mugford, Hanson, & Coligado, 2018). That said, the potential for such a momentous change to a fundamental component of criminal justice processes also warrants careful consideration. A recent study conducted by Hogan and Sribney (2019) suggested that applying the five-level system to combined risk appraisals based on the Static-99R and STABLE-2007 could prompt a significant increase in the number of cases identified as requiring intervention, thereby

also potentially precipitating a corresponding increase in resource outlay. Given the potential for the adoption of the five-level system to influence public policy, correctional operations and the rights of various individuals, a thorough understanding of the system is of the utmost importance. The following review is intended to increase awareness and foster further discussion of the system through the analysis of select issues, and through the identification of practical and theoretical problems.

Generalist and specialist perspectives on risk and offending

Andrews and Bonta (Andrews & Bonta, 2010; Bonta & Andrews, 2017) have identified common risk factors that predict reoffending of all types, commonly referred to as the Central Eight. These include: criminal history, antisocial attitudes, antisocial peers, antisocial personality pattern, education/employment problems, substance misuse, family/marital problems, and lack of prosocial leisure/recreational pursuits. A considerable body of empirical research supports the relationships among these general risk factors and various types of offending. For instance, Olver, Stockdale, and Wormith's (2014) meta-analysis of the predictive validity of the Level of Service (LS) instruments, which emphasize the Central Eight risk factors, was suggestive of wide-ranging relevance. Scores on the LS instruments were associated with an omnibus outcome measure of any general offending and more narrow subcategories, including violent, non-violent and, to a lesser degree, sexual recidivism. With regard to risk for intimate partner violence (IPV), two meta-analyses, conducted by Hanson, Helmus, and Bourgon (2007) and by van der Put, Gubbels, and Assink (2019), respectively, failed to find evidence that specialized instruments were superior to generalist alternatives. Reviews of research on risk for sexual recidivism also suggest that factors associated with general criminality, such as antisocial personality features

or attitudes, are associated with specialized offending outcomes (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005). Thus, there is empirical evidence to suggest that a general proclivity toward antisocial conduct is associated with offending of all types.

On the other hand, it is also well established that certain risk factors are differentially associated with certain types of offending, and it is common to study subcategories of risk and subpopulations of offenders. For instance, meta-analyses conducted by Hanson and colleagues (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005) identified certain risk factors, such as atypical sexual interests, as being particularly relevant to sexual recidivism. Reflecting the unique characteristics of risk factors for sexual offending behavior, several specialized risk tools are currently in common use. Other empirical research (e.g. Babchishin, Hanson, & VanZuylen, 2015; Harris, Mazerolle, & Knight, 2009) suggests that there are meaningful distinctions even among subgroups of persons who have sexually offended. For instance, Babchishin and colleagues (2015) observed that, relative to persons with contact sexual offences, persons with offending patterns limited to child pornography offences demonstrated more difficulties with sexual self-regulation, but fewer traditional risk markers associated with antisociality (e.g. criminal history, supervision failures). Spurred on by such observations, researchers have undertaken efforts to produce highly specialized risk assessment technology, such as the Child Pornography Offender Risk Tool (CPORT; Seto & Eke, 2015). While increasing specialization in risk technology may motivate efforts to standardize risk communication, it can also complicate such endeavors considerably.

Conceptual issues related to the variability among types of risk, and among offending subpopulations, are important to consider because the five-level system is criterion referenced. Any subcategory of offending behavior (e.g. sexual offending, violent offending or

IPV) is, by definition, less common than the entirety of possible offending behaviors. It follows then, that observed and expected rates of specialized reoffending will necessarily diverge from the rates used to develop the five-level system. For those attempting to apply the five-level system to other forms of risk, and to populate the risk levels with specialized samples, it may not be clear what type of offending behavior should be a primary focus (e.g. general recidivism versus a delimited category of offending). One option is to prioritize consistency across tools and applications, by selecting general recidivism as the statistical criterion of choice for populating categories, and to offer specialized recidivism rates as a secondary piece of information. While this approach would compromise specificity, it would arguably increase consistency in language when applying the five-level system across the criminal justice system. Notably, to date it appears that researchers applying the five-level system to specialized tools for sexual offending (Hanson, Babchishin, et al., 2017; Olver et al., 2019) and violent offending (Davies & Helmus, 2019; Olver et al., 2019) have used specialized reoffending as the criterion of interest. A consequence of this decision is that while the categories carry the same nominal labels as the original five-level system, the criterion-based meaning of the labels has deviated from definitions presented in the original white paper.

In addition to reoffending rates, the five-level system also emphasizes the underlying risk and need profiles of offenders. Narrative descriptions focus on the number and density of criminogenic needs associated with a given level, and compare and contrast risk profiles with that of the *typical* offender. Even allowing for some fuzzy boundaries around a hypothesized prototype, some critical questions regarding the identification of a *typical* offender remain. It is reasonable to assume that a hypothetical individual/exemplar deemed representative of the entire population of offenders (i.e. persons who have engaged in

any form of criminal conduct) will differ in meaningful ways from one that might be considered representative of subpopulations. For instance, it would be reasonable to expect differences among representative examples of persons that engage in violent offending, in sexually violent offending, or in the consumption of child sexual abuse material. Consistent with the issues related to recidivism outcome variables, if the different need profiles underlying specialized types of offending (e.g. anomalous sexual arousal/interest in relation to risk for sexual offending) are used to populate risk levels, it follows that the actual meaning (i.e. latent constructs) associated with the risk levels will vary considerably across applications of the system.

If one accepts the premise that the collection of risk factors that best predicts one type of offending may be different to that which best predicts another, or accepts the premise that different instruments can or should be used for different purposes, then one must also consider the implications of such variability for the development of standardized risk language. Setting measurement error and instrument idiosyncrasies aside for a moment, if specialization is prioritized, an individual assessed with multiple tools could be assigned multiple risk level ratings. Such an individual could be deemed a Level III offender for a general recidivism, Level II for sexual recidivism and Level I for another outcome. Conversely, if stakeholders emphasize generalizability and general antisociality, then the risk levels could carry greater shared meaning. Perhaps the same individual could be deemed only a Level III offender, which by definition carries a particular risk of general recidivism, and then be assessed to pose a particular risk for violent recidivism, a particular risk for sexual recidivism, and so on. If the former approach is chosen (i.e. specialization), it is the opinion of this writer that the category labels should not be proffered without a clear articulation of the specificity of their meaning in a given context (e.g. Level III Category

for sexual recidivism). Olver and colleagues (2018) offered a related but even stronger recommendation for specificity in language, by suggesting that assessors report separate categories for different scores and subscales of the Static-99R and VRS-SO, two sexual offending related instruments. What is clear is that, like recommendations pertaining to risk communication with labels of low, moderate or high risk, it would be prudent for criminal justice stakeholders to avoid reifying labels that carry little objective meaning without further context (e.g. a specific tool and outcome of interest).

Are all tools and samples equally applicable to the five-level system?

No bricks without clay: content and constructs

In the original white paper, Hanson, Bourgon, and colleagues (2017) acknowledged that in order for the five-level system to succeed in providing a common language, researchers, clinicians and policy makers must be able to apply the system to diverse tools, samples and populations. To this end, the developers provided some guidance for populating the categories using additional instruments, which emphasized statistical metrics. However, as mentioned above, fundamental differences in approach exist among some of the most empirically supported, and widely adopted (Singh et al., 2014), risk assessment instruments. Some instrument developers (Douglas et al., 2013) and proponents of SPJ approaches, for example, have argued that actuarial data and SPJ approaches are ‘fundamentally incompatible’ (p. 11). Thus, insofar as reoffending estimates are necessary for the five-level system, it may be argued that the system cannot be applied to SPJ instruments. This is problematic at a policy level, given that SPJ instruments like the HCR-20^{V3} are currently in use in various settings, including U.S. courts (Cox et al., 2018; Vitacco, Erickson, Kurus, & Apple, 2012). While it may be possible to

modify the five-level scheme by omitting the actuarial component to apply it to SPJ tools, this would remove some of the precision intended by the developers.

While some may argue that it is inappropriate to apply the five-level system to SPJ tools on a statistical basis, a similar argument could be made about applying the system to certain actuarial tools on a conceptual basis. Hanson, Bourgon, and colleagues (2017) argued that risk levels ought to correspond to psychologically meaningful individual characteristics, and further argued that they should 'be aligned with a recognizable pattern of meaningful, distinct characteristics' (p. 3). Whether every static actuarial tool can meet these criteria is debatable. Consider the following case example, using the Static-99R, to which the five-level system has already been applied:

A 19 year old male offender, Mr. Blue, was charged with a contact sexual offence, after an incident involving a male intimate partner. Mr. Blue received a score of +4 on the Static-99R, on the basis of his age, never having lived with a lover for two years, and having offended against an unrelated, male victim. This score placed him in the Level IVa or Above Average Risk category.

Based on actuarial decision-making principles, this case may appear straightforward to some readers, particularly given that the Static-99R boasts a robust and widely representative normative database. However, concerns do arise when one attempts to apply the non-actuarial elements of the five-level system to this case. The original description of Level IV offenders (Hanson, Bourgon, et al., 2017) suggests that these individuals exhibit multiple chronic and severe criminogenic needs. However, the Static-99R is a criterion-referenced actuarial tool, comprising items selected based on a predictive relationship with recidivism. While scores are correlated with the types of criminogenic needs alluded to above, they do not provide a direct measure

of such needs. In fact, when examining the correspondence between criminogenic needs and the five-level system as applied to the Static-99R, Hanson, Babchishin and colleagues (2017) observed that the patterns 'were not entirely consistent with expectations' (p. 591). Referring to the example above, applying the standard description of the needs associated with Level IV offenders to Mr. Blue is questionable at best, given that little is known about his actual profile of needs. While the five-level system purports to facilitate the identification of specific treatment needs, the data produced by the Static-99R are arguably not fit for purpose. Particularly when one considers that direct measures of criminogenic needs exist, such as the STABLE-2007 and VRS-SO, the practice of using a tool like the Static-99R to formulate statements regarding individual needs is difficult to justify.

Adequate range

Hanson, Bourgon, and colleagues (2017) developed the five-level system to capture the range of risk propensities observed among correctional populations. The category labels demarcate conceptually meaningful groups across a spectrum of risk/need density, ranging from persons who do not differ significantly from the general population, to persons with chronic and diverse needs, for whom reoffending is essentially inevitable. It bears repeating that the developers intended for the system to be widely applicable, across various tools and unique populations. That said, the guidelines for adapting the system to new applications (e.g. populating the levels using a jurisdictional sample) are based upon the characteristics of particular samples and instruments. This raises the question of whether all samples and instruments are appropriate for the five-level system.

One critical consideration in any application of the five-level system concerns the nature of the sample. As mentioned above, the Static-99R boasts a normative database that is arguably reasonably representative of the

population of sexual offenders (Phenix, Helmus, & Hanson, 2016). As such, the sample population is statistically well suited to the five-level system, which requires anchors at the extreme ends of the risk distribution, as well as a generous range of scores in between. However, few of the more than 400 violence risk instruments identified in Singh and colleagues' (2014) survey of professionals are likely to boast such a sample. Take for example the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2010), an actuarial tool designed to assess risk for IPV recidivism, and supported by the results of meta-analytic reviews (Messing & Thaller, 2013; van der Put et al., 2019). The ODARA development sample of 589 offenders was obtained from comprehensive records of police contacts and, as such, may reasonably be considered highly representative of the population of IPV offenders in the jurisdiction. Scores on the ODARA range from 0 to 13, but the mean score among the development sample was 2.89 ($SD = 2.14$); given the relatively small number of offenders scoring above 5, the developers collapsed high scores into two categories: 5–6 and 7–13, respectively. In contrast, recent ODARA research results reported by Perrault, Hilton, and Pham (2019) described a sample with a starkly different profile of scores. While representative of the population served by an operational threat assessment unit, the Integrated Threat and Risk Assessment Centre (ITRAC), this was essentially a pre-selected high-risk/high-needs sample. The median ODARA score was 9, and no scores below 3 were observed. Put another way, every offender in Perrault and colleagues' sample scored above the mean score from the ODARA development sample. To populate the five-level system using ODARA scores, one would need to determine whether and how to integrate these discrepant samples. If nothing else, this example demonstrates that much like any norm-referenced metric of behavioral phenomena, the

boundaries of the five levels may be subject to the idiosyncrasies of sampling populations.

In addition to the sample used to populate the five-level system's categories, one must also consider basic psychometric properties of the instrument of interest. Basic measurement issues, such as floor and ceiling effects, might influence the suitability of particular tools. Using instruments that cannot adequately capture the breadth and variability of a risk domain in a given population will in turn degrade the precision and consistency of application of the five-level system.

Shifting anchors: base rates and the five-level system

One intended function of the five-level system is to facilitate comparisons among the results of diverse risk assessment instruments. Given that over 400 instruments were identified in Singh and colleagues' (2014) study of international violence risk assessment practices, it is reasonable to conclude that inconsistent risk communication practices could lead to significant confusion. That being said, in identifying an appropriate solution to this apparent problem, it is worth considering whether specific risk instruments communicate analogous information in disparate ways, or simply communicate disparate information.

Because risk is both time and context dependent (Heilbrun et al., 1999), effective risk communication requires explicit acknowledgment of each of these considerations. The five-level system was developed to convey information about risk for community recidivism, with base rates of reoffending defined by two-year follow-up periods. Not all risk instruments serve the same purpose. For instance, the Brøset Violence Checklist (BVC; see Woods & Almvik, 2002) was designed to evaluate day-to-day fluctuations in inpatient violence risk, and has demonstrated predictive validity over such short durations (Vaaler et al., 2011; Woods, Ashley, Kayto, & Heusdens, 2008). The Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink,

Nicholls, & Desmarais, 2009) was developed for forensic psychiatric settings and assesses changes in violence risk factors (and strengths) occurring over periods ranging from days to months. Various studies have supported the predictive validity of the START for inpatient violence occurring over periods of approximately one to three months (e.g. Chu, Thomas, Ogloff, & Daffern, 2011; Hogan & Olver, 2018; Wilson, Desmarais, Nicholls, Hart, & Brink, 2013). While meta-analytic evidence suggests that many general violence risk assessment tools do predict institutional violence (Campbell et al., 2009; Hogan & Ennis, 2010), a more discerning review of the literature also reveals that violence in different settings may be best predicted by different types of factors, even among the same individuals (Hogan & Olver, 2016, 2019). It is not clear whether and how the five-level system should be applied to risk assessments for short-term or highly context-dependent outcomes.

Even when applying the five-level system to community recidivism, time-scale is a critical consideration worth discussing. Insofar as recidivism base rates are used to populate the five-level system's risk categories, adherence to common time-scales is essential to ensuring consistency in meaning. Equating a recidivism rate of 5% over one year to a similar percentage over five years is equivalent to equating separate family incomes of \$100,000 divided over one and five years, respectively. Bearing this in mind, it is noteworthy that the application of the five-level system to the Static-99R used five-year sexual recidivism rates (Hanson, Babchishin, et al., 2017). Olver and colleagues (2018) prudently followed Hanson and colleagues' (2017) example in applying the five-level system to the VRS-SO, another sexual offence specific instrument. As a result, these specialized instruments do share a common language with one other, but not with more generalist applications of the system. Currently, work is underway to apply the five-level system to specialized violence risk instruments, the Violence Risk Appraisal

Guide-Revised (VRAG-R; Rice, Harris, & Lang, 2013) and VRS (Davies & Helmus, 2019; Olver et al., 2019); whether the respective research groups will converge on a common time-scale to populate the risk categories remains to be seen. It is critical that those undertaking subsequent applications of the five-level system consider such decisions, lest the problem of inconsistency in meaning afflicts the five-level system as it has other nominal systems.

The preceding discussion identified a number of potential problems related to base rates when applying the five-level system across disparate tools, contexts, time-scales and offence categories. Some readers may presume that these issues can be mitigated by controlling for these factors, clearly defining the outcome of interest and comparing like for like. However, official recidivism rates are unstable, due to a number of variable influences, including victim reporting rates, as well as policing and other criminal justice system practices, policies and resources (Zara & Farrington, 2016). In fact, empirical evidence suggests that base rates may fluctuate considerably, even when comparing risk ratings from a single tool. For instance, Helmus and colleagues (2012) conducted a meta-analytic review of base rates predicted by the Static-99R and Static-2002R among a number of samples, and concluded that the observed discrepancies were large enough that they 'could lead to meaningfully different conclusions' (p. 1164) regarding offender risk. Thus, while base rate information is a well-established means of improving human decision making (e.g. Tversky & Kahneman, 1974), and while some caveats were offered in the original white paper, the universal recidivism estimates linked to the five-level system at present could convey a greater degree of precision than is warranted.

Dynamic or changeable risk

The five-level system is intended to accommodate changes in risk. In describing the prognosis of persons assigned to each level, Hanson,

Bourgon, and colleagues (2017) indicated that an individual could transition between levels by addressing need areas. It does not appear that the system treats persons who transition into a new risk level as being distinct from persons originally assigned to the same level. This implies that certain needs that were present before a transition are essentially absent or irrelevant to risk after a transition. Hanson, Bourgon, et al. suggest that interventions should follow the risk–need–responsivity (RNR) principles of offender rehabilitation (Bonta & Andrews, 2017), but otherwise, the five-level system appears largely neutral with regard to how changes in risk and need areas may occur.

While the nature and utility of dynamic or changeable risk factors have been discussed and debated by many (Douglas & Skeem, 2005; Kraemer et al., 1997; Olver & Wong, 2016), empirical evidence for such risk factors proved elusive for many years (Serin, Lloyd, Helmus, Derkzen, & Luong, 2013). This is beginning to change, driven in large part by the pioneering work of Wong, Olver, and their colleagues, using theoretically informed dynamic risk instruments, the VRS and the VRS–SO. Changes in risk as measured by the VRS have demonstrated incremental predictive validity for violent recidivism among federally incarcerated offenders (Coupland & Olver, 2018; Lewis, Olver, & Wong, 2013) and among forensic psychiatric patients (Hogan & Olver, 2019). Similar positive results have been found with regard to sexual recidivism using the VRS–SO (Beggs & Grace, 2011; Olver, Beggs Christofferson, Grace, & Wong, 2014; Olver, Wong, Nicholaichuk, & Gordon, 2007). This body of research converges with the five-level system in that it suggests that changes in risk are both possible and relevant. However, it diverges from the five-level system based on the mechanism of change it suggests. The dynamic components of the VRS and VRS–SO are based on the stages of change model. This model assumes that change is an active

process, whereby individuals develop motivation to change, and then undertake cognitive, behavioral and experiential changes to address their problems. Using this model, need areas can be managed, but do simply cease to exist, which is a conceptually significant distinction. This more nuanced approach is broadly consistent with other models of risk, including the idea of *interactive protective factors* as described by Farrington and Ttofi (2012), which suggests that certain variables may interact with risk factors to nullify their effects, rather than render them absent, per se.

Given that the five-level system is intended to be widely applicable across tools, and given that different tools may measure change in different ways (if at all), the preceding discussion may appear immaterial to the pursuit of a common language. One could argue that treating needs in this way is consistent with generic concepts like *variable markers* and *variable risk factors* (Monahan & Skeem, 2016) and increases generalizability. On the other hand, by treating needs that are addressed through intervention as equivalent to the absence of such needs, the five-level system could be omitting important information. To illustrate this point, consider the cases of two nearly identical hypothetical offenders, charged with similar sexual offences. Is Offender A, who has no paraphilia, equivalent to Offender B, who has a paraphilia that contributed directly to the index offence, but that is currently managed with anti-libidinal medication and behavioral strategies? The five-level system speaks to needs, but does not directly account for this type of discrepancy. On a related note, in the case of Offender B, does the length of time this need has been managed matter? In their own discussion of the five-level system, Simourd and Olver (2019) offered a compelling argument against substituting rudimentary proxy variables, such as time in treatment, in place of more meaningful measures of treatment-related changes. These authors cautioned that, while appealing in many ways, simplistic conceptualizations of

interventions, dosage and changes in risk variables could distract and detract from good correctional practice, because time is not tantamount to progress. Regardless of one's position on the state of the literature on dynamic risk, producers and consumers of risk assessment information would be well advised to be mindful of the information conveyed by the five-level system.

Shared language, different theories: application to the good lives model

The five-level system is intended to provide a common risk language for criminal justice professionals. However, the system is also explicitly designed to facilitate 'the implementation of Risk–Need–Responsivity (RNR) principles' (Hanson, Bourgon, et al., 2017, p. 4), and, as noted by a peer reviewer, not all criminal justice professionals wholly and exclusively endorse the RNR model (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Proponents of the good lives model (GLM) for example, approach concepts of risk and rehabilitation from a different theoretical perspective, which emphasizes facilitating offenders' personal goals and fulfilling lives, as opposed to simply mitigating personal risk factors or deficits (Ward & Gannon, 2006). A comprehensive analysis of the five-level system from a GLM perspective is beyond the scope of this paper, but examples of potential issues are offered below.

Many professionals surveyed by McGrath and colleagues (2010) endorsed both RNR principles and the GLM. This suggests that while they may not be viewed as comprehensive, risk and need assessments informed by existing instruments are compatible with services based on the GLM. That said, it is likely that service providers following the GLM will be dissatisfied with the qualitative descriptors of the five-level system's risk levels, and particularly so when considering treatment-related changes in risk factors, strengths or protective factors. Hanson, Bourgon, and colleagues'

(2017) initial white paper does include cursory references to personal strengths in relation to the risk levels, with strengths presumed to be less prominent as risk levels increase. In the context of GLM-based interventions, however, factors such as an individual's personal goals, strengths and contexts are paramount (Ward & Gannon, 2006) and require further explication. Furthermore, the gap between the perceived needs of GLM service providers and the information communicated by the five-level system can be expected to widen each time an offender is reassessed, and as he or she progresses toward rehabilitation. As noted in the previous section, however, this concern is not limited to the GLM perspective – RNR informed assessments of offender change also pose problems for the five-level system.

Conclusions and recommendations

In light of the gravity of the decisions influenced by risk assessment, the widespread use (and misuse) of labels with arbitrary, inconsistent and imprecise definitions is troublesome. Even seemingly precise language may be interpreted differently by various parties, depending on organizational, training/education and cultural factors. For these reasons, efforts to standardize risk communication are both necessary and admirable. At the same time, it is worth acknowledging that current practices in structured risk assessment and risk communication, while imperfect, are themselves products of an evolutionary process and confer considerable benefits relative to their precursors. Prudent application of any new technique demands thorough and careful consideration of the potential consequences, both favorable and unfavorable.

The five-level system is intended to provide a framework for diverse individuals to convey and use risk information with a greater degree of precision and consistency than ever before. However, the system is not a cure-all for the limitations of risk communication. The fact that the first applications of the

system to specific tools (i.e. the Static-99R and VRS-SO) effectively produced new statistical and conceptual definitions for the risk labels demonstrates that there is further work to be done to realize the goal of common language in risk assessment. If the five-level system is widely adopted, the meaning of risk labels will continue to vary in substantive ways, depending on various contexts, instruments, outcomes and populations. While various adult offender populations may require special consideration, as discussed above, many would likely argue that youth populations require a separate system altogether. It is likely that the pursuit of the overarching goals of wide applicability, statistical precision and conceptual richness are not entirely compatible. As such, it is also unlikely that professionals can assign equal weight to each of these goals, in applying the five-level system to new circumstances. A consensus on such priorities should be sought, and will likely prove critical in determining whether the five-level system is to make a material contribution to the pursuit of a common language for risk assessment.

The problem of inconsistency in meaning is likely to be particularly evident, and arguably least defensible, in cases in which these conceptually meaningful labels are applied using second-generation instruments. This is because second-generation tools comprise atheoretical, but nonetheless predictive, risk markers. If the system is to be applied to second-generation tools in some form, it may be limited to a situation-dependent system of statistical labels. If treated like standard scores (i.e. scores based on statistical deviation around a mean), for example, the labels would be statistically defined, but essentially meaningless, independent of particular samples and tools. Such distinctions are themselves arbitrary to a certain extent, as there is little inherent meaning behind the difference between standard scores of 1 and 1.1, but they do nonetheless facilitate communication. Even if a statistically based approach is emphasized, certain actuarial instruments and samples may

be deemed unsuitable if certain pre-requisite conditions are not met. Examples of such pre-requisites include the availability of a reasonably representative sample of the entire population and a well-calibrated instrument that is sensitive to a wide range of variability in risk. If the five-level system is to retain and emphasize its current focus on psychologically meaningful constructs, then second-generation instruments may simply require another system altogether. Put simply, a defensible summary of one's needs is best informed by assessing those needs as directly as possible, rather than simply reviewing correlated risk markers.

It is well worth acknowledging that many of the aforementioned limitations of the five-level system also apply to pre-existing systems to an equal or greater extent. For example, the LS/CMI links recidivism estimates to nominal risk labels ranging from very low to very high, whereas various SPJ instruments offer risk labels ranging from low to high that have no relationship to recidivism estimates whatsoever. To be clear, this review is not intended to discredit the five-level system, or to discourage researchers, professionals or policy-makers from considering its adoption. That being said, a novel system endorsed as precise and consistent should be subjected to at least as much scrutiny as those unabashedly arbitrary systems it is intended to supplant. If not instituted carefully, a purportedly universal language may fail to solve the problem of applying dissimilar meaning to the same terms. In fact, this danger may actually increase, if users privy to problems with older systems accept the contention that the new labels represent a solution, and convey risk information 'precisely, clearly, and consistently' (Hanson, Bourgon, et al., 2017, p. 12).

It is hoped that this review may stimulate further examination and explication of the five-level system, both theoretically and empirically. One promising direction for future study would be a survey of relevant stakeholders, including forensic psychiatrists and

psychologists, as well as other criminal justice professionals, to establish a hierarchical set of priorities for the system (e.g. adherence to base rates versus density of needs; generalizability versus specificity). This may facilitate a standardized approach to the future applications. Secondly, it is recommended that researchers align work on advancing risk communication more closely with developments in risk assessment more generally. Although a broader discussion is beyond the scope of this paper, purely actuarial risk assessment (i.e. the second generation) has been criticized for potentially perpetuating systemic biases in some contexts, particularly given that one's involvement in the criminal justice system predicts, and potentially begets (Prins, 2019), future involvement. Concerns in this regard are mitigated considerably when considering evidence obtained using relatively sophisticated third- and fourth-generation instruments, which suggests that there are psychologically meaningful constructs that underlie risk at an individual level, across diverse populations (e.g. Olver, Kingston, & Sowden, 2020). As outlined earlier, related findings indicate that meaningful and measurable changes in these personal risk constructs correspond with reductions in recidivism. If the aims of the five-level system include capturing and communicating information about probabilities of recidivism, psychologically meaningful risk constructs and changes in risk, then third- and fourth-generation risk instruments should be embraced, and may be used to inform efforts to more fully accommodate change information into the system.

To maximize the potential benefits of the five-level system, purveyors and users of risk information are encouraged to make themselves aware of the issues raised in the preceding discussion, and to make note of any relevant limitations of the system as applied to their respective contexts. Even in ideal circumstances, the system does not negate one's responsibility to communicate risk information clearly in a straightforward, fulsome and deliberate

manner. At minimum, much like a label of low, moderate or high, a label of Level I, Level II or Level III should not be offered without a thorough explanation of how that label was selected, what information it provides and what information it does not provide. If it is recognized and acknowledged that the five-level system has not yet resolved the known problems with risk communication, then the system can inspire further increasingly coordinated efforts to do so, constituting a considerable step forward in the evolution of risk assessment.

Acknowledgements

The author wishes to thank Mark Olver for his review of an early version of this article, and Gabriela Corabian for her subsequent review and comments. The views expressed in this article do not necessarily represent the views of the Integrated Threat and Risk Assessment Centre.

Ethical standards

Declaration of conflicts of interest

Neil R. Hogan has declared no conflicts of interest

Ethical approval

This article does not contain any studies with human participants or animals performed by the author.

ORCID

Neil R. Hogan <http://orcid.org/0000-0002-1109-5901>

References

- Andrews, D.A., & Bonta, J. (1995). *The level of service inventory-revised*. Toronto, ON: Multi-Health Systems.
- Andrews, D.A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16(1), 39–55. doi:10.1037/a0018362

- Andrews, D.A., Bonta, J.L., & Wormith, J.S. (2004). *The level of service/case management inventory*. Toronto, ON: Multi-Health Systems.
- Andrews, D.A., Bonta, J.L., & Wormith, J.S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52, 7–27. doi:10.1177/0011128705281756
- Babchishin, K.M., Hanson, R.K., & VanZuylen, H. (2015). Online child pornography offenders are different: A meta-analysis of the characteristics of online and offline sex offenders against children. *Archives of Sexual Behavior*, 44(1), 45–66. doi:10.1007/s10508-014-0270-x
- Barbaree, H.E., Langton, C.M., & Peacock, E.J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 18(4), 423–440. doi:10.1177/107906320601800408
- Beggs, S.M., & Grace, R.C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology*, 79(2), 182–192. doi:10.1037/a0022900
- 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.
- 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.
- Bonta, J., & Andrews, D.A. (2017). *The psychology of criminal conduct* (6th ed.). New York, NY: Routledge.
- Bourgon, G., Mugford, R., Hanson, R.K., & Coligado, M. (2018). Offender risk assessment practices vary across Canada. *Canadian Journal of Criminology and Criminal Justice*, 60(2), 167–205. doi:10.3138/cjccj.2016-0024
- Campbell, M.A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36(6), 567–590. doi:10.1177/0093854809333610
- Chu, C.M., Thomas, S.D.M., Ogloff, J.R.P., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health*, 10(4), 337–345. doi:10.1080/14999013.2011.629715
- Claussen-Schulz, A.M., Pearce, M.W., & Schopp, R.F. (2004). Dangerousness, risk assessment, and capital sentencing. *Psychology, Public Policy, and Law*, 10(4), 471–491. doi:10.1037/1076-8971.10.4.471
- Coupland, R.B., & Olver, M.E. (2018). Assessing dynamic violence risk in a high-risk treated sample of violent offenders. Assessment. Advance online publication. <http://doi.org/gd6kbp>
- Cox, J., Fairfax-Columbo, J., DeMatteo, D., Vitacco, M.J., Kopkin, M.R., Parrott, C.T., & Bownes, E. (2018). An update and expansion on the role of the Violence Risk Appraisal Guide and Historical Clinical Risk Management-20 in United States case law. *Behavioral Sciences & the Law*, 36(5), 517–531. doi:10.1002/bsl.2376
- Davies, S., & Helmus, L.M. (2019, May). *The 5-level risk and needs framework and violent recidivism*. Paper presented at the Fourth North American Correctional and Criminal Justice Psychology Conference, Halifax, NS.
- Douglas, K.S., Hart, S.D., Webster, C.D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk for violence – user guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Douglas, K.S., & Skeem, J.L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11(3), 347–383. doi:10.1037/1076-8971.11.3.347
- Farrington, D.P., & Ttofi, M.M. (2012). Protective and promotive factors in the development of offending. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behavior and crime: Contributions of developmental and evaluation research to prevention and intervention* (pp. 71–88). Cambridge, MA, US: Hogrefe Publishing.
- Hanson, R.K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20(2), 212–217. doi:10.1177/0886260504267740
- 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(5), 582–597. doi:10.1037/pas0000371

- Hanson, R.K., Bourgon, G., McGrath, R.J., Kroner, D., D'Amora, D.A., Thomas, S.S., & Tavaréz, L.P. (2017). *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.
- Hanson, R.K., & Bussiere, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348–362. doi:10.1037/0022-006X.66.2.348
- 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, ON: Public Safety Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-fndrs/index-eng.aspx>
- Hanson, R.K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis (Corrections Research User Rep. No. 2007-07)*. Ottawa, ON: Public Safety Canada. Retrieved from http://www.publicsafety.gc.ca/res/cor/rep/_fl/vra_ipv_200707_e.pdf
- 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(6), 1154–1164. doi:10.1037/0022-006X.73.6.1154
- Harris, D.A., Mazerolle, P., & Knight, R.A. (2009). Understanding male sexual offending: A comparison of general and specialist theories. *Criminal Justice and Behavior*, 36(10), 1051–1069. doi:10.1177/0093854809342242
- Harris, G.T., Rice, M.E., & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20(4), 315–355. doi:10.1177/0093854893020004001
- Heilbrun, K., Dvoskin, J., Hart, S., & McNiel, D. (1999). Violence risk communication: Implications for research, policy, and practice. *Health, Risk & Society*, 1(1), 91–105. doi:10.1080/13698579908407009
- 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(9), 1148–1171. doi:10.1177/0093854812443648
- 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. doi:10.1177/0886260507309337
- 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.
- Hogan, N., & Ennis, L. (2010). Assessing risk for forensic psychiatric inpatient violence: A meta-analysis. *Open Access Journal of Forensic Psychology*, 2, 137–147. Retrieved from <https://www.oajfp.com/blank-8>
- Hogan, N.R., & Olver, M.E. (2016). Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law and Human Behavior*, 40(3), 233–243. doi:10.1037/lhb0000179
- Hogan, N.R., & Olver, M.E. (2018). A prospective examination of the predictive validity of five structured instruments for inpatient violence in a secure forensic hospital. *International Journal of Forensic Mental Health*, 17(2), 122–132. doi:10.1080/14999013.2018.1431339
- Hogan, N.R., & Olver, M.E. (2019). Static and dynamic assessment of violence risk among discharged forensic patients. *Criminal Justice and Behavior*, 46(7), 923–938. doi:10.1177/0093854819846526
- Hogan, N.R., & Sribney, C. (2019). Combining Static-99R and STABLE-2007 risk categories: An evaluation of the five-level system for risk communication. *Sexual Offender Treatment*, 14, 14. Retrieved from <http://www.sexual-offender-treatment.org/187.html>
- Jung, S., & Buro, K. (2017). Appraising risk for intimate partner violence in a police context. *Criminal Justice and Behavior*, 44(2), 240–260. doi:10.1177/0093854816667974
- Jung, S., Ennis, L., Brown, K., & Ledi, D. (2015). The association between presentence risk evaluations and sentencing outcome. *Applied Psychology in Criminal Justice*, 11(2), 111–125.
- Jung, S., Pham, A.T., & 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. doi:10.1080/14789949.2013.806567
- Kraemer, H.C., Kazdin, A.E., Offord, D.R., Kessler, R.C., Jensen, P.S., & Kupfer, D.J. (1997). Coming to terms with the terms of

- risk. *Archives of General Psychiatry*, 54(4), 337–343. doi:10.1001/archpsyc.1997.01830160065009
- Levenson, J.S., & Morin, J.W. (2006). Factors predicting selection of sexually violent predators for civil commitment. *International Journal of Offender Therapy and Comparative Criminology*, 50(6), 609–629. doi:10.1177/0306624X06287644
- Lewis, K., Olver, M.E., & Wong, S.C.P. (2013). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment*, 20(2), 150–164. doi:10.1177/1073191112441242
- McGrath, R.J., Cumming, G., Burchard, B., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American survey*. Brandon, VT: Safer Society Press.
- Messing, J.T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence*, 28(7), 1537–1558. doi:10.1177/0886260512468250
- Monahan, J., & Skeem, J.L. (2014). Risk redux: The resurgence of risk assessment in criminal sanctioning. *Federal Sentencing Reporter*, 26(3), 158–166. doi:10.1525/fsr.2014.26.3.158
- Monahan, J., & Skeem, J.L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, 12, 489–513. doi:10.1146/annurev-clinpsy-021815-092945
- Neal, T.M.S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, 41(12), 1406–1421. doi:10.1177/0093854814548449
- Olver, M.E., Beggs Christofferson, S.M., Grace, R.C., & Wong, S.C.P. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale-Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment*, 26(5), 472–499. doi:10.1177/1079063213502679
- Olver, M.E., Coupland, R.B., Lewis, K., Hogan, N.R., Higgs, T., Cortoni, F., ... Wong, S.C.P. (2019, June). *Applications of the Violence Risk Scale in dynamic risk assessment and management*. Paper presented at the Fourth North American Correctional and Criminal Justice Psychology Conference, Halifax, NS.
- Olver, M.E., Kingston, D.A., & Sowden, J.N. (2020). An examination of latent constructs of dynamic sexual violence risk and need as a function of Indigenous and nonindigenous ancestry. *Psychological Services*. Advance online publication. doi:10.1037/ser0000414
- Olver, M.E., Mundt, J.C., Thornton, D., Beggs Christofferson, S.M., Kingston, D.A., Sowden, J.N., ... Wong, S.C.P. (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. doi:10.1037/pas0000538
- Olver, M.E., Stockdale, K.C., & Wormith, J.S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, 26(1), 156–176. doi:10.1037/a0035080
- Olver, M.E., & Wong, S.C.P. (2016). Assessing treatment change in sex offenders. In L.A. Craig & M. Rettenberger (Eds.), *The Wiley-Blackwell handbook on the assessment, treatment, and theories of sexual offending (Volume: Assessment)*. Chichester, UK: Wiley.
- Olver, M.E., & Wong, S.C.P. (2019). Offender risk and need assessment: Theory, research, and applications. In D.L.L. Polaschek, A. Day, & C. R. Hollin (Eds.), *The Wiley handbook of correctional psychology*. Chichester, UK: Wiley.
- 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
- Perrault, L.P., Hilton, N.Z., & Pham, A.T. (2019, May). *Ontario Domestic Assault Risk Assessment: Predicting violent recidivism among high-risk intimate partner violence offenders*. Paper presented at the Fourth North American Correctional and Criminal Justice Psychology Conference, Halifax, NS.
- 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.
- Phenix, A., Helmus, L.M., & Hanson, R.K. (2016). *Static-99R & Static 2002R: Evaluators' workbook* (Unpublished manual). Retrieved from www.static99.org
- Prins, S.J. (2019). Criminogenic or criminalized? Testing an assumption for expanding criminogenic risk assessment. *Law and Human*

- Behavior*, 43(5), 477–490. doi:10.1037/lhb0000347
- Rice, M.E., Harris, G.T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: the Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychological Assessment*, 25(3), 951–965. doi:10.1037/a0032878
- Scurich, N. (2018). The case against categorical risk estimates. *Behavioral Sciences & the Law*, 36(5), 554–564. doi:10.1002/bsl.2382
- Serin, R.C., Lloyd, C.D., Helmus, L., Derkzen, D.M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior*, 18(1), 32–53. doi:10.1016/j.avb.2012.09.002
- Seto, M.C., & Eke, A.W. (2015). Predicting recidivism among adult male child pornography offenders: Development of the Child Pornography Offender Risk Tool (CPORT). *Law and Human Behavior*, 39(4), 416–429. doi:10.1037/lhb0000128
- Simourd, D.J., & Olver, M.E. (2019). Prescribed correctional treatment dosage: Cautions, commentary, and future directions. *Journal of Offender Rehabilitation*, 58(2), 75–91. doi:10.1080/10509674.2018.1562503
- Singh, J.P., Desmarais, S.L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., ... Otto, R.K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13(3), 193–206. doi:10.1080/14999013.2014.922141
- Storey, J.E., Kropp, P.R., Hart, S.D., Belfrage, H., & Strand, S. (2014). Assessment and management of risk for intimate partner violence by police officers using the brief spousal assault form for the evaluation of risk. *Criminal Justice and Behavior*, 41(2), 256–271. doi:10.1177/0093854813503960
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.)*, 185(4157), 1124–1131. doi:10.1126/science.185.4157.1124
- Vaaler, A.E., Iversen, V.C., Morken, G., Fløvig, J.C., Palmstierna, T., & Linaker, O.M. (2011). Short-term prediction of threatening and violent behaviour in an acute psychiatric intensive care unit based on patient and environment characteristics. *BMC Psychiatry*, 11, 44. Retrieved from doi:10.1186/1471-244X-11-44
- van der Put, C.E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, 47, 100–116. doi:10.1016/j.avb.2019.03.008
- Vitacco, M.J., Erickson, S.K., Kurus, S., & Apple, B.N. (2012). The role of the Violence Risk Appraisal Guide and Historical, Clinical, Risk-20 in US courts: A case law survey. *Psychology, Public Policy, and Law*, 18(3), 361–391. doi:10.1037/a0025834
- Ward, T., & Gannon, T.A. (2006). Rehabilitation, etiology, and self-regulation: The comprehensive good lives model of treatment for sexual offenders. *Aggression and Violent Behavior*, 11(1), 77–94. doi:10.1016/j.avb.2005.06.001
- Webster, C.D., Martin, M.-L., Brink, J.H., Nicholls, T.L., & Desmarais, S.L. (2009). *Short-Term Assessment Risk and Treatability (START)*. Coquitlam, BC: Forensic Psychiatric Services Commission.
- Wilson, C.M., Desmarais, S.L., Nicholls, T.L., Hart, S.D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law Hum Behav*, 37(6), 377–388. doi:10.1037/lhb0000025
- Wong, S.C.P., & Gordon, A. (1999–2003). *Violence risk scale*. Saskatoon, SK: Author.
- Wong, S.C.P., Olver, M.E., Nicholaichuk, T.P., & Gordon, A. (2003–2017). *The Violence Risk Scale: Sexual Offense version (VRS-SO)*. Saskatoon, Canada: Regional Psychiatric Centre and University of Saskatchewan.
- Woods, P., & Almvik, R. (2002). The Brøset Violence Checklist (BVC). *Acta Psychiatrica Scandinavica*, 106 (s412), 103–105. doi:10.1034/j.1600-0447.106.s412.22.x
- Woods, P., Ashley, C., Kayto, D., & Heusdens, C. (2008). Piloting violence and incident reporting measures on one acute mental health inpatient unit. *Issues in Mental Health Nursing*, 29(5), 455–469. doi:10.1080/01612840801981207
- Wylie, L.E., Clinkinbeard, S.S., & Hobbs, A. (2019). The application of risk-needs programming in a juvenile diversion program. *Criminal Justice and Behavior*, 46(8), 1128–1147. doi:10.1177/0093854819859045
- Zapf, P.A., & Dror, I.E. (2017). Understanding and mitigating bias in forensic evaluation: Lessons from forensic science. *International Journal of Forensic Mental Health*, 16(3), 227–238. doi:10.1080/14999013.2017.1317302
- Zara, G., & Farrington, D.P. (2016). *Criminal recidivism: Explanation, prediction and prevention*. New York, NY: Routledge.