Risk Prediction and Sex Offending

Jacinta R. Cording[a], Tony Ward[b], and Sarah M. Beggs Christofferson[a]

[a] Department of Psychology, University of Canterbury, Christchurch, New Zealand; [b] School of Psychology, Victoria University of Wellington, Wellington, New Zealand

Address correspondence to:

Jacinta Cording
c/o Sarah Christofferson
University of Canterbury
Department of Psychology
Private Bag 4800
Christchurch, New Zealand
Email: jacinta.cording@gmail.com

Risk Prediction and Sex Offending

Accurately predicting rates of future sexual offending for a given individual is the 'holy grail' for those working within the field of sexual violence. This is in large part due to the important functions risk prediction plays in informing harm reduction efforts and in providing justification for the continued restriction of personal liberty inherent in the custodial management of offenders. As such, an increasing amount of research has focussed on improving the methods and accuracy of risk prediction for sexual offending, and the applied use of risk assessments in the management and treatment of offenders is continuing to grow.

Predictions of risk generated by risk assessment methods or tools are currently used to inform decision-making in a wide variety of areas within the justice system, including sentencing, custodial placement and security rating, release planning, parole conditions, and preventative detention or extended supervision of released offenders. One of the primary considerations for clinicians and other decision-makers in each of these areas is the balance between minimising risk to the public versus supporting self-improvement and enhancing well-being for offenders (Ward, 2013); risk prediction provides one way in which to determine where this balance should lie for a given individual, dependent on the amount of risk that they potentially pose to the public if unmanaged. Additionally, risk assessment can provide clinicians with guidance on case formulation and treatment targets by highlighting particular areas of need for particular offenders, and an assessment of overall levels of risk inform decisions regarding the intensity of treatment that should be provided (as per the risk principle of Andrews and Bonta's Risk-Need-Responsivity (RNR) model; Andrews & Bonta, 2016).

Because of the extensive use of risk assessment within the management and treatment of sexual offenders, the accuracy of predictions is of utmost importance. The over-prediction of risk (i.e. predicting higher rates of offending than are actually observed) is likely to have negative implications for offender rights as well as resource management, including unjustified extended custodial sentences, overly restrictive parole conditions, and unnecessary treatment provision. Conversely, the under-prediction of risk (i.e. predicting lower rates of offending than are actually observed) could result in the inadequate provision of treatment to meet individual need, and/or the failure to provide the necessary level of supervision for high-risk offenders, thereby threatening the safety of the community.

Predicting the risk of sexual offending is therefore a key task for clinicians and other individuals involved in the management and treatment of offenders. However, despite extensive advancements in the accuracy and utility of risk assessment within the forensic field, there remain a number of theoretical and practical issues related to the prediction of sexual offending that have implications for the application of these predictions. The purpose of this chapter is to provide a critical overview of these issues and to discuss possible solutions moving forward. We begin with a discussion of the development of risk assessment approaches over time, outlining the "four generations" of risk assessment, including a discussion of approaches based on structured professional judgement. We then discuss conceptualizations of construct validity and their applications to the development of risk assessments; this includes a discussion of issues related to the indistinct nature of dynamic risk factors, and an over-reliance on data and the hypothetico-deductive method of science to develop risk assessments. We also discuss the different applications that risk assessments are used for, and the competing requirements of the assessments as a result of this dual application. Next, we discuss the implications that a data-driven method of developing risk assessments has for the generalizability and availability of such tools for different sub-groups

of sexual offenders. We then finish with a discussion of approaches to the incorporation of change information into predictions of risk.

The development of risk assessment tools

Although consideration of the potential for further offending has always been one of the key considerations in offender management and treatment decisions, it is only relatively recently that specific tools have been developed for the quantifiable prediction of specific types of antisocial behavior. The development of risk assessment approaches has previously been described as having progressed in four distinct generations (Andrews, Bonta, & Wormith, 2006). Although this was initially outlined in relation to the assessment of risk for general offenders, risk assessment approaches for the prediction of sexual offending has largely followed the same trajectory.

First-generation: Unstructured clinical judgement

First generation risk assessments refer to largely unstructured professional judgement, in which the clinician or other decision-maker forms a subjective opinion on risk level, based on their knowledge and experience. This approach to sex offending risk assessment is now discouraged as research has shown that unstructured judgement typically has poor predictive accuracy (Hanson & Morton-Bourgon, 2009) and is often biased towards over-estimating risk (Craig, Browne, Stringer, & Beech, 2004). This bias is suggested to be due to a failure to consider the relatively low base rate of sexual recidivism in judgements of risk. In a meta-analysis including 28,757 convicted sexual offenders from 100 samples, the observed rate of sexual recidivism was 11.5% over an average follow-up period of just under 6 years (Hanson & Morton-Bourgon, 2009). These base rates do not substantially increase even given a longer follow-up period; one study reported sexual recidivism rates of 26% for adult sexual offending and 32% for child sexual offending over a follow-up period of 25 years (Prentky,

Lee, Knight, & Cerce, 1997), and a more recent study reported sexual recidivism rates of 11.2% for child sexual offending and 13.5% for adult sexual offending with an average follow-up of 15 years (Vess & Skelton, 2010). Low base rates of reoffending make over-estimation of risk more likely, in part leading to unnecessarily restrictive management and supervision decisions, and an over-allocation of treatment resources. However, although base rates of official sexual recidivism are relatively low, it is important to note that official recidivism rates are likely to under-represent the true rates of recidivism. This is because of the likelihood of undetected offences that are not captured in official records, and because of the impacts of judicial processes (such as plea-bargaining) that can obscure the true nature of the crime committed. This under-representation is something that should be considered when assessing the implications of official recidivism rates.

Second-generation: Structured actuarial tools

The next step in the progression of risk assessment approaches was the development of structured actuarial tools (Andrews et al., 2006). These tools typically include a pre-determined set of risk factors (i.e. characteristics shown to bear a statistical link with recidivism) that are rated according to a standardized scoring framework. Ratings of risk are then combined to derive a total risk score and/or risk band that corresponds to empirically-derived estimates of risk. This approach to risk assessment therefore provides a way to quantify expected recidivism rates that avoids issues with the subjectivity and bias inherent in unstructured clinical judgement. Structured actuarial tools were initially comprised of static risk factors only i.e. those not able to be changed through treatment, such as previous sexual offenses or age at first sexual offence. Some key static risk factors for sexual offending include prior sexual offending, having male or stranger victims, and a history of treatment dropout (Hanson & Bussiere, 1998). Tools comprising static factors only include the Static-99 (Hanson & Thornton, 1999), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey,

Harris, Rice, & Cormier, 1998), and the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997).

The increase in predictive validity provided by second generation tools was demonstrated by Hanson and Morton-Bourgon's (2009) meta-analysis, in which empirical actuarial tools were found to be more accurate in predicting sexual recidivism than unstructured professional judgement (d = 0.67 and 0.42, respectively). These effect sizes are indicative of the size of the mean difference in scores between the recidivists and the non-recidivists on a given measure, and can be interpreted as a medium effect for empirical actuarial tools and a small effect for unstructured professional judgement. Predictive accuracy can also be assessed using Area under the Curve (AUC) values. AUC values represent the probability that a randomly-selected recidivist will have a higher score on a given measure than a randomly-selected non-recidivist; a score of 0.5 means that the measure does no better than chance at predicting recidivism, whereas a score of 1 means that a measure perfectly predicts recidivism. The AUC value therefore gives an indication of the accuracy of a measure in terms of its rate of true positives versus false positives. Effect sizes of d = 0.67 and 0.42 convert to AUC values of 0.68 and 0.62, respectively. This means that in Hanson and Morton-Bourgon's study, empirical actuarial tools were found to have a 68% probability of assigning a recidivist a higher score than a non-recidivist, whereas unstructured clinical judgement had a 62% probability of doing the same.

In addition to improved predictive accuracy, risk assessments based on static risk factors are easily administered and scored, cost-effective, and enable the efficient screening of large numbers of individuals at a time. They can therefore be utilized relatively easy to inform sentencing and parole decisions (which must be undertaken for large numbers of offenders each year), and they can also contribute to the efficient allocation of individuals to appropriate treatment options or levels of intervention.

One of the major issues with risk assessments based on static measures, however, is their lack of ability to identify potential treatment targets, and their failure to take into account any of the environmental or situational factors that may influence offending. Static risk factors are often based on demographic or historical factors, and are therefore unchangeable, and are therefore unable to reflect changes in risk that occur due to variations in situations or external influences over time, including treatment. It is also important to note that while static factors may be useful in predicting long-term recidivism rates for an aggregate of offenders, because static factors are poor indicators of change, they are of little help to predicting *when* an offender will reoffend (Hanson & Harris, 2001). For these reasons, and because of a growing body of literature demonstrating the empirical validity of dynamic factors in predicting risk (e.g., Hanson & Harris, 2000; Hudson, Wales, Bakker, & Ward, 2002), there was a subsequent shift to the inclusion of dynamic factors in risk assessment.

Third-generation: Risk/Need assessments

Andrews and colleagues (2006) define third generation risk assessments as being distinct from second generation tools in their systematic and objective consideration of individual needs, typically through the incorporation of dynamic risk factors. Dynamic risk factors are factors linked with offending that are amenable to change; key dynamic risk factors for sexual offending include substance abuse problems, pro-offending attitudes, and deviant sexual interest (Hanson & Harris, 2000). This provides an advantage over the second generation tools in that assessments are theoretically able to capture changes in risk over time or in response to direct interventions, and they are also able to be used to inform the selection of appropriate treatment targets. Examples of commonly-used third-generation risk assessments for sexual offending include the STABLE 2007 and the ACUTE 2007 (Hanson, Harris, Scott, & Helmus, 2007), and the Violence Risk Scale - Sexual Offender Version

(VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003). The sensitivity of third-generation measures to changes in risk-related factors over time suggests that they may be able to improve upon the predictive accuracy of static risk assessments, particularly for individuals who have completed treatment or for whom circumstances have substantially changed over time. Indeed, a number of studies have been able to show that assessments incorporating dynamic risk factors are able to provide incremental predictive validity beyond assessments incorporating static factors alone (Allan, Grace, Rutherford, & Hudson, 2007; Beggs & Grace, 2010; Craissati & Beech, 2003; Hanson et al., 2007). The size of this difference varies across studies and dynamic risk measures, however one study found an increase of 0.08 in AUC value when using post-treatment dynamic risk assessment as opposed to static risk assessments (Beggs & Grace, 2010). They found that the assessment incorporating dynamic risk measures gave a higher risk score to recidivists than non-recidivists 80% of the time, as opposed to 72% of the time for the static risk measure. This increase in predictive accuracy lends support to the idea that dynamic risk factors are tapping into a distinct facet of risk that is not being captured by historical or other static factors alone. Because third-generation assessments are increasingly being utilized as one of the more common approaches to risk assessment within forensic services, further discussion of issues specific to these tools (including incorporating treatment change into risk assessment) is found at a later stage of this chapter.

Structured professional judgement

Before moving on to fourth-generation risk assessment tools, it is important to briefly discuss risk assessment approaches based on structured professional judgement (SPJ). Approaches based on SPJ use empirically-derived frameworks to guide the assessment of risk in a structured, but flexible, manner. Specific SPJ measures for sexual offending include the Sexual Violence Risk–20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997) and its evolved

version, the Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003). Although Andrews and colleagues (2006) considered SPJ tools to fall within the first generation of risk assessment approaches, we argue that it is more appropriate to consider these tools as an alternative third generation approach. This is because although clinical judgement plays a primary role in this approach to risk assessment, in this case clinical judgement is applied in a guided manner to a pre-determined set of risk domains; examples from the RSVP include history of sexual violence, psychological adjustment, social adjustment, and mental disorder. This means that there are formal systems in place that reduce the level of subjectivity inherent in the unstructured clinical judgement approach.

Instead, SPJ tools arguably allow for a greater utilization of the unique knowledge of the offender and their circumstances that is held by the clinician to inform the assessment of treatment progress and risk, particularly in areas that are not fully captured by other dynamic tools. Such an approach acknowledges the complexity of risk assessment in the real world, in which a variety of factors, both psychological and external, can influence an individual's behavior at any given point in time. This complexity may be difficult to incorporate into a fully actuarial tool, which must balance breadth of variables covered with the practicalities of the time and resources required to complete the measure. Approaches that allow for structured professional judgement therefore potentially provide a useful means of assessing real-world risk because of their ability to reflect change in the attitudes and behaviors of the individual across a wide range of areas, whilst still being based on a credible empirical foundation. Support has been found for the interrater reliability of SPJ approaches (e.g., intraclass correlation coefficients (ICC) in the "fair" to "excellent" ranges for the RSVP across the four studies overviewed by Judge, Quayle, O'Rourke, Russell, & Darjee [2014]). Structured professional judgement tools also potentially allow for the development of risk

predictions for populations for which there are no specific actuarial tools, due to a lack of available empirical information (these kinds of populations are discussed further below).

Despite the advantages of the SPJ approach, it is important that its limitations in terms of predictive accuracy are noted. Although a large meta-analysis found that SPJ tools were predictive of sexual recidivism, the predictive accuracy of this approach was lower than that obtained by empirical actuarial tools (d = 0.46 and 0.67, respectively; Hanson & Morton-Bourgon, 2009). These effect sizes relate to small differences between mean scores on SPJ tools for recidivists compared with non-recidivists, and moderate differences for empirical actuarial tools. The effect sizes can also be interpreted as AUC values, with an AUC of 0.63 for SPJ tools and 0.68 for empirical actuarial tools. Thus, although SPJ tools may be a useful inclusion to the overall approach to evaluating risk and treatment outcomes, it is important that this is augmented by information obtained from actuarial tools as part of a wider assessment of risk. Additionally, some research suggests that professional judgement may be more accurately applied to risk assessment only when adjusting risk levels downwards on the basis of additional information, rather than to increase risk level (Wormith, Hogg, & Guzzo, 2012).

Fourth-generation: Case management

Due to some concerns that risk assessments were being administered for individuals but that the results of these assessments were not then being used to inform case management and decision-making (Andrews & Bonta, 2016), fourth-generation tools were developed to more explicitly highlight the necessary links between assessment and case management. The tools include measurement of common risk-related factors to assess risk level, but they also include measurement of specific responsivity needs, planning of treatment targets and intervention, and recording of treatment progress. There are currently no fourth-generation tools that have

been developed specifically for sexual offenders, although there has been some support for the ability for existing tools to predict sexual recidivism. Wormith and colleagues (2012) examined the predictive accuracy of the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) with a sample of 1,905 sex offenders and 24,545 non-sexual offenders (i.e., individuals who had a history of offending that did not include a sexual offence). They found that the LS/CMI was significantly predictive of sexual reoffending for both sexual offenders (AUC = .77) and non-sexual offenders (AUC = 0.75), indicating that fourth-generation tools may provide a promising direction for the development of future sexual offender risk assessments. Indeed, given previous findings that predictive accuracy is higher for measures when predicting outcomes that they were specifically designed for (Hanson & Morton-Bourgon, 2009), it is important to assess whether fourth-generation tools developed specifically for the prediction of sexual offending may provide even greater levels of accuracy than tools developed for general or violent offending. Because of the lack of fourth-generation tools specific to sexual offending and the resultant lack of information about their efficacy with this specific population, it is difficult to draw any strong conclusions or make strong comparisons between these and other existing measures in the prediction of sexual offending. The remainder of this chapter will therefore limit the discussion of risk prediction largely to second- and third-generation tools and their various applications.

The construct validity of dynamic risk assessments

As mentioned above, most third and fourth generation risk assessment tools include dynamic factors to measure the level of risk posed by a given individual. For this reason, the construct validity of dynamic risk factors becomes important when considering how to interpret risk assessments and what they are saying about individual offenders. Construct validity is commonly defined as whether a particular measure represents or measures what it

is supposed to (Colliver, Conlee, & Verhulst, 2012), and is a key to the notion that psychological assessments are useful in capturing the latent characteristics that a given individual may possess. –Because construct validity relates to the meaningfulness of what is being measured, itis not necessarily of primary importance in assessing the utility of risk assessments to predict offending (Helmus & Babchishin, in press); in this instance, the ability of the items to accurately predict the outcome is of paramount importance. Understanding what it is that these items are actually capturing is of lesser, if any, importance. However, construct validity becomes more important when one considers the additional ways in which risk assessments are commonly utilized, particularly in the construction of clinical explanations and treatment plans (Ward & Fortune, 2016). In particular, third generation risk assessments are often viewed as being an improvement over second generation tools because they are able to both predict risk *and* identify possible treatment targets for interventions. The transition here is that between the context of prediction and that of explanation. However, in order for these tools to be useful for informing treatment formulation and targets, the dynamic risk factors being measured should arguably represent meaningful psychological constructs that are causally linked with offending (Ward & Beech, 2015; Ward & Fortune, 2016). In other words, construct validity becomes an important facet of these risk assessment tools as they are applied to areas other than pure prediction.

In attempting to demonstrate the construct validity of dynamic risk assessments, most research has focused on demonstrating concurrent validity – the extent to which a particular measure correlates with existing measures of the same constructs or outcomes – and predictive validity – the extent to which a measure accurately predicts a specific relevant outcome (typically recidivism). Because of the importance of these two concepts in assessing the utility of risk assessments, we will briefly address each of these separately.

Concurrent validity

Overall, research has supported the concurrent validity of common dynamic risk assessment tools, in that assessments of risk tend to remain consistent across different risk assessments (Beggs & Grace, 2010; Nunes & Babchishin, 2012). Although general concordance is typically found between measures, situations in which different risk assessments result in different risk ratings is still surprisingly common. For example, Barbaree, Langton and Peacock (2006) assessed risk ratings across five different risk assessment tools and found that only 5% of cases were consistently rated as high risk or low risk.

One of the contributing factors to inconsistent ratings across measures is the lack of a universal method for determining risk or a common language with which to communicate risk (Hanson & Bourgon, 2016). Although the names of risk categories is often consistent across measures (e.g. low, medium, and high risk), the expected outcomes for each of these categories (e.g. recidivism rates and relative risk ratios) differ between measures. This poses a difficulty in allowing individuals to compare the information provided by different measures, particularly if they are not familiar with the details of each given assessment. One potential solution to this issue is being provided by work supported by the Justice Center of the Council of State Governments to develop a common risk language (see Hanson & Bourgon, 2016, for further details). Under their system, there are five standardised levels of risk (Levels I-V) that individuals can be assigned to on the basis of their anticipated rate of reoffending. These categories could therefore be applied to any risk assessment tool no matter their assessment approach, provided that normed recidivism estimates are developed for each potential risk score or group. Research utilising these new common risk categories is in its infancy, however this may provide an effective way of promoting more comprehensive approaches to assessing offender risk whilst avoiding issues with inconsistencies between measures.

Predictive validity

A strong empirical base has developed over recent years which supports the view that measures of dynamic risk can significantly improve the accuracy of risk prediction over and above the ability of static risk alone (Beggs & Grace, 2010; Hanson & Harris, 2000). In general, studies indicate that empirically-derived actuarial measures are significantly more predictive of reoffending compared to both structured clinical judgement and unstructured clinical judgement (Hanson & Morton-Bourgon, 2009). Somewhat unsurprisingly, studies have also found that the predictive accuracy of a particular measure changes depending on the match between the target behaviour being predicted and the type of behaviour the measure was developed to predict; for instance, a measure developed to measure risk of future sexual reoffending is typically more predictive of sexual recidivism than violent or general recidivism (Hanson & Morton-Bourgon, 2009).

Despite the evidence for moderate predictive validity of commonly-used risk assessment tools, it is important to note that these measures are not perfect predictors of recidivism. It could instead be argued that many of these tools tend to over-estimate the risk posed by a given individual. For example, individuals categorised into the highest risk band for sexual recidivism using the VRS-SO are estimated to offend at a rate of approximately 36%, meaning that just under two-thirds of these 'high risk' offenders are effectively predicted to not re-offend sexually, according to official recidivism data (Olver, Beggs Christofferson, Grace, & Wong, 2013). Although official recidivism data is likely to be an under-estimate of true recidivism rates and the actual reoffending rates of these high-risk individuals is likely to be somewhat higher than the 36% reported, it is clear that these individuals are potentially not as "high risk" as one might expect.

This clearly poses a problem for the assessment of risk for highly restrictive approaches to offender management such as preventative detention or Sexually Violent Predator evaluations; is this level of accuracy sufficient to impose such severe restrictions of freedom on individuals? Our ability to communicate what the scales say about the risk posed by an individual offender will likely improve after implementing the common risk language outlined above, however this will not change the ability for the risk assessments themselves to accurately identify whether a given individual will reoffend or not. In part this is because risk assessments have been developed as prognostic tools (i.e. tools that provide probabilities of a future event) rather than diagnostic tools (i.e. tools that provide a "yes" or "no" assessment of a given outcome). The implications of this distinction are further discussed below, however suffice to say that the information provided by risk assessments must be carefully considered when applied to real life situations, despite the strong support for their predictive ability overall. *A new conceptualisation of construct validity*

The high level of concurrent and predictive validity demonstrated by research into risk assessments should suggest strong support for the construct validity of common risk assessment approaches, however, scholars have recently questioned the common conceptualization of construct validity as being demonstrated by concurrent and predictive validity (Borsboom, Mellenbergh, & van Heerden, 2004; Colliver et al., 2012; Haig, 2012). Instead, they argue that a given measure should be considered to have good construct validity only if it is able to demonstrate a causal or explanatory link between the attributes it measures and the outcome of interest.

The implication is that dynamic risk assessment could be considered to have good construct validity only if researchers are able to demonstrate a causative or explanatory link between dynamic risk factors and recidivism. Although dynamic risk factors are in theory causally related to offending (and would therefore meet this explanatory requirement of

construct validity), there has been some recent doubts cast on whether the current reliance on correlational analyses and significance testing in order to identify the dynamic risk factors used in risk assessments is a valid method of identifying truly causal factors relating to recidivism risk (Haig, 2012; Heffernan & Ward, 2015; see also, Ward, 2016; Ward & Fortune, 2016). Additionally, because of the current ethical and pragmatic constraints around research designs required for a strong determination of causality, it could be argued that the true causes of sexual offending are unlikely to be empirically determined in the near future. It is therefore unclear whether the current conceptualization of what constitutes as a "risk factor" is demonstrably valid in a more meaningful sense of the term.

A further problem with validity in the area of risk assessment relates to the multi-dimensional and indistinct nature of many dynamic risk factors (Heffernan & Ward, 2015). "Cognitive distortions" is a dynamic risk factor that is commonly incorporated into risk assessment tools, and provides a good example of what is meant here. Typically, cognitive distortions are conceptualized as non-normative belief structures that include justifications and rationalizations for sexual offending (Gannon, Ward, & Collie, 2007). For example, an offender may attempt to justify his actions by claiming that the child victim was a willing participant. However, as noted by Ó Ciardha and Gannon (2011), 'cognitive distortions' has been applied to a multitude of different constructs including 'maladaptive beliefs', 'defensiveness', 'rationalizations', 'incorrect or deviant cognitive practices', and 'etiological cognitions'.

This variability in the definition of dynamic risk factors, as demonstrated by 'cognitive distortions', poses a significant problem for developing valid measures of dynamic risk. Without clear definitions of what is (and is not) being represented by each individual risk factor, it is uncertain which features of each factor are linked to recidivism, how strongly it is linked, and how to create clear scoring guidelines that reflect this. Such uncertainty will

not only potentially degrade the accuracy and discriminative validity of a measure, but also inter-rater and test-retest reliability. It also has implications for construct validity – how can we be sure that we are measuring what we want to measure, and that the link between this construct and recidivism can be explained, when we are unable to define clearly what the construct actually is?

Changing our approach to developing dynamic risk assessments

The failure of most dynamic risk factors used in risk assessment in demonstrating the explanatory power necessary for strong construct validity is partially due to the method in which these factors are currently identified. Overreliance on the hypothetico-deductive methodology in psychological research, where empirical data are used to identify, describe and/or discover correlates of constructs to inform theory, has long been criticized (e.g., Cohen, 1994; Rozeboom, 1960). In response to these criticisms, there has recently been increasing discussion of the potential for the abductive method to provide a more meaningful and valid approach to research (Borsboom et al., 2004; Haig, 2014). Haig (2014) describes the abductive approach as "reasoning from factual premises to *explanatory* conclusions" (p.60), noting that "phenomena, not data, as evidence for the abducted theories" (p.61). Phenomena[JC1] in this instance are distinct from data in that they are enduring characteristics or patterns of a particular construct that have been identified across multiple studies. This differs from a purely inductive approach, whereby conclusions or theories are the 'same in kind' as the data used to generate them, meaning that they are more descriptive than explanatory in nature. Therefore, the appeal of the abductive approach for identifying dynamic risk factors is that in using this method, we are more likely to discover risk factors that have strong causal links to offending, thereby increasing the accuracy and validity of risk assessments that incorporate these factors. Use of the abductive method would also increase our ability to explain and conceptualize the links between the factors and offending, including

an understanding of aetiology and causal networks, which has clear benefits for the use of risk assessments in guiding interventions and assessing changes in risk (for a further discussion of the use of abductive method and dynamic risk factors see Ward & Beech, 2015).

The abductive critique of the hypothetico-deductive method is clearly applicable to prior research on dynamic risk assessment, which has largely identified risk factors that should be included through the use of regression-based statistical methods, rather than being primarily driven by theory (Heffernan & Ward, 2015; Ward, 2016; Ward & Beech, 2015). One could argue that variables which were studied as potential dynamic risk factors – such as lack of empathy for victims – were selected based on prior theoretical grounds (e.g., Marshall, Hamilton, & Fernandez, 2001), however acceptance of these dynamic risk factors was reliant largely on evidence of their ability to predict recidivism, ideally beyond the contribution made by static factors. Thus, they have ultimately been accepted as important criminogenic factors on the basis of their predictive accuracy, rather than on whether they represent meaningful causal factors for sexual offending.

The dual uses of risk assessments

Reasons for the use of the hypothetico-deductive method are understandable when one considers the primary goal of risk assessments: to predict the likelihood of future offending for a given individual. Thus it is logical to identify dynamic risk factors by their ability to predict reoffending beyond the static, actuarial factors that had already been shown to have predictive validity. However, this becomes an issue due to the use of risk assessments to assess individual needs (and therefore inform treatment decisions) as well as predict risk. Currently, both of these tasks are typically performed using the same tools, despite the differences in what is required of the tools for these two tasks: the assessment of needs is in

essence a diagnostic task, in that the measure is being used to determine the presence or absence of a certain condition or characteristic (and in treatment contexts result in warranted causal inferences; Ward & Fortune, 2016), whereas the assessment of risk is a prognostic task, in that the measure is being used to assess the probability of a future outcome (Helmus & Babchishin, in press).

The difference between these two tasks has several practical implications for scale development, including the selection of items to include in the measure. Diagnostic scales are inherently norm-referenced (i.e., they are trying to capture the degree to which an individual displays a particular characteristic), whereas prognostic scales are inherently criterion-referenced (i.e., they are designed to specifically predict a particular outcome (Helmus & Babchishin, in press). Whereas norm-referenced scales should ideally include multiple items that assess the same construct in different ways to ensure the reliability of the 'diagnosis', criterion-referenced scales are solely concerned with predictive accuracy, and are therefore largely atheoretical. Because practical reasons often require measures to be as short as possible whilst still serving their purpose, criterion-referenced scales should ideally include a small number of items that each represent a distinct factor that has been linked with offending; overlap between what is predicted by different items is to be avoided. The primary concern here is efficiency and predictive accuracy of the scale, rather than the theoretical implications of the construct being measured. The implication is that in order to meet these competing requirements whilst ensuring that measures have construct validity where this is important, we may need to develop different measures for assessing risk as opposed to identifying treatment targets (or incorporate two different scales into the one measure, similar to the format of fourth-generation tools). This implication suggests that it is unwise to translate dynamic risk factors from risk prediction measures into causal constructs to be used

in the explanation of offending and to direct treatment, without considerable theoretical reworking (Ward & Fortune, 2016).

This is particularly important where the risk assessments are currently being used in treatment. Merely demonstrating correlations between dynamic risk factors and recidivism falls short of providing evidence of a causal linkage, or of providing a strong explanatory theory behind the correlation. This is likely to result in an incomplete picture of the risk posed by an individual offender that is lacking in an explanation of the historical causes or current maintenance of behavior. Consequently, implications for treatment formulation in terms of the most important needs to target to reduce risk are compromised. A greater understanding of the etiology of serious offending would allow us to develop more effective strategies for early intervention, ideally to reduce first-time sexual and violent offending rather than reoffending. In moving away from the purely data-driven approach to risk/needs tool development, it is hoped that we can develop a deeper and more meaningful understanding of how these various factors may contribute to the generation and maintenance of offending, as well as how they combine to determine the overall level of risk of an offender. It also avoids issues with developing risk assessments that are generalizable and that can be used for offenders for whom it is difficult to obtain data; we turn to this discussion in the next section.

Risk assessment and data availability

Although the predictive validity of dynamic risk measures has been well-established through large volumes of research, questions remain about the generalizability of these measures to different groups, such as female or youth offenders, or offenders from other cultural groups. Despite the high levels of heterogeneity found amongst the population of individuals who offend sexually (see Robertiello & Terry, 2007, for review), research

focussing on the identification of important dynamic risk factors to incorporate into risk assessment tools is overwhelmingly conducted using samples of white adult male offenders (Reisig, Holtfreter, & Morash, 2006; Schwalbe, 2008) under the general assumption that dynamic risk factors are broadly consistent across gender and age.

Focussing specifically on gender, a number of scholars have rejected the "gender-neutral assumption" (Yesberg, Scanlan, Hanby, Serin, & Polaschek, 2015), instead asserting that there are a number of important differences between men and women who sexually offend which challenge the idea that dynamic risk assessments developed solely on male populations are equally valid for females. These include differences in motivations for offending, history of offending, acceptance of offending and offence characteristics (e.g., use of force; Robertiello & Terry, 2007). Additionally, even though some risk factors may initially appear to apply to both male and female offenders, it has been found that there are sometimes differences in how these factors manifest as a function of gender, which has important implications for how they should be measured and incorporated into risk assessment (Ford, 2010).

For example, offense-supportive cognitions are an important dynamic risk factor that has been identified for both female and male sexual offenders, however the content of these cognitions are gender-specific. Offence-supportive cognitions for females have been shown to often incorporate perceptions of males as being threatening and entitled to behave in sexually harmful ways, and these beliefs have an impact on the offending behavior of females (Gannon & Rose, 2015; Gannon, Rose, & Cortoni, 2010). Furthermore, other studies have concluded that the content of females' offence-supportive cognitions can change depending on whether the offending was committed alone or with others, with the latter often incorporating distortions about their co-offender(s) that have an influence on their own offending behavior (Beech, Parrett, Ward, & Fisher, 2009). Although we are beginning to

build an understanding of the unique factors specific to each gender, due to the difficulty in obtaining samples of female sex offenders – and therefore in gathering data that can inform theories about factors that are important for the etiology or maintenance of offending for females – there are still large gaps in our understanding of the similarities and differences between female and male sexual offenders. This has led some to argue that applying risk assessment tools that have been developed largely with males to female offenders is erroneous and could possibly lead to inappropriate risk-related decisions (Cortoni, 2010). At the time of writing, we are unaware of any studies that have specifically assessed the predictive accuracy of risk assessments specific to sexual offending for men versus women. However even if existing risk assessment tools were to be validated on female populations, the question would remain whether the inclusion of gender-specific items would further improve the predictive ability of risk assessments for females (Yesberg et al., 2015). For this reason, there have been suggestions that a more beneficial and productive approach to this issue would be to develop female-specific assessments regardless of the predictive validity of existing measures (Gannon et al., 2010). These arguments of course also apply to other offender groups as well, including youth offenders and offenders of cultural groups other than White North American populations.

The lack of specific risk assessment tools for female offenders despite estimates that females are responsible for approximately 5% of sexual offenses (Cortoni, Hanson, & Coache, 2010) highlights the issue of relying only on available data (rather than theory) to identify dynamic risk factors that can be incorporated into risk assessment approaches. This is not an issue that is limited only to female offenders; there is a lack of research in general, and specific risk assessments in particular, for other large groups within the sexual offender population, including those who offend against intimate partners, and individuals with predominantly adult victims; most existing risk assessment tools are designed to be used for

both individuals with predominantly adult victims *and* those with predominantly child victims, despite studies suggesting that they are distinctly different populations with unique needs (e.g., Beauregard, Leclerc, & Lussier, 2012; Sigre-Leirós, Carvalho, & Nobre, 2015; Vess & Skelton, 2010).

Leading on from our discussion in the previous section, we suggest that one of the major reasons for this lack of risk assessment tools for important sub-groups of sexual offenders is because of the data-driven nature of research into dynamic risk factors and assessment. This leads to an *over-reliance* on empirical evidence that is available to researchers – including being limited in which dynamic risk factors *can* be investigated by what measures are currently available to assess them - and *under-reliance* on theory or aetiology to guide our understanding of dynamic risk factors as scientific phenomena, and how these factors might be combined into an overall meaningful measure of risk. As explained by Haig (2013, p. 137), "[d]ata themselves are of scientific interest and importance only because they serve as evidence for the phenomena under investigation." Another area in which limited empirical information has impacted on the rate of progression of risk assessment approaches is in the incorporation of change into risk assessments, which we discuss in the following section.

Incorporating change into assessments of risk

As discussed at the beginning of this chapter, one of the major benefits of third-generation assessments is that they are theoretically able to assess changes in risk over time. However, due to the potential issues with the accuracy of clinical judgement outlined previously, it is important that any approach to assessing changes in risk is structured and minimizes the possibility of bias. Given previous findings, it is fair to hypothesize that unstructured clinical judgement will *over-estimate* the influence of change on risk level;

although dynamic risk factors are indeed theoretically able to change (thereby altering overall risk), it is important to remember that most factors measured are relatively enduring, with pre-treatment assessments of risk commonly found to remain significantly predictive of recidivism with follow-ups of several years *post*-treatment (Beech, Friendship, Erikson, & Hanson, 2002). This suggests that dynamic factors may be more stable than one might initially consider them to be, leading evaluators to incorrectly quantify the extent to which these factors change over the course of an often relatively short period of treatment. Indeed, previous research has underscored issues with clinical judgement of change, showing that factors unrelated to treatment progress can affect ratings; for example, individuals with a more positive view of treatment and sex offenders are more likely to report identification of positive treatment outcomes (Kivlighan & Tarrant, 2001).

Any adjustments to risk assessment on the basis of change observed over the course of a treatment programme will therefore need to acknowledge the gradual and complex nature of desistance from sexual offending. One issue here is the relatively small amount of research that has assessed whether a) dynamic factors really do change over time, and b) whether these changes can reliably be linked with changes in risk. Although dynamic risk factors are in theory changeable and they have been empirically linked with recidivism, studies have often only assessed this link at one point in time (e.g., Beech et al., 2002; Dempster & Hart, 2002). This means that there is very little empirical guidance as to how to best provide structure around the adjustment of risk on the basis of observable changes in risk-related factors. Where more recent studies have begun to explore change, there have been some mixed findings. For example, one study of treated sexual offenders found no significant change on measures of cognitive distortions and deviant attraction post-treatment (Jung & Gulayets, 2011). This raises questions about the nature of dynamic risk factors and whether they truly are changeable (and therefore whether assessments of risk *should* be

amended), however this becomes less clear when one considers other possible reasons for the lack of measureable change on these factors, including ineffective treatment, or inappropriate or insensitive measurement techniques. In contrast, other studies have reported significant improvements across treatment, with medium to large effect sizes found for change on dynamic factors including sexual interests, anger, pro-offending attitudes, and interpersonal skills (Beggs & Grace, 2011; Hudson et al., 2002).

Although the ability for risk factors to change over time has been supported by some studies, recidivism outcomes are often not included as part of the investigation into change. However, in order to justify the amendment of risk assessments on the basis of observable change, it is necessary that these changes be demonstrated to be *meaningful* (i.e., linked to changes in actual reoffending risk; Cording, Beggs Christofferson, & Grace, 2016). Despite the widespread assumption that changes on dynamic risk factors will be associated with changes in observed recidivism rates, research has only more recently begun to test this assumption empirically and again, results are mixed. Whereas some studies have been able to demonstrate a significant link between assessed pro-social change (i.e., improvements in a given dynamic risk factor across treatment, as described in the paragraph above) across treatment with reduced sexual recidivism rates at follow up (Beggs & Grace, 2011; Olver, Nicholaichuk, Kingston, & Wong, 2014), others have failed to find any significant link after controlling for static risk (Olver, Kingston, Nicholaichuk, & Wong, 2014). The failure to identify significant links between changes on dynamic risk factors and reductions in observed recidivism may be partially attributable to methodological issues with appropriately controlling for pre-treatment levels of risk; individuals who are higher risk pre-treatment have more ‘room to move’ or opportunity to demonstrate prosocial change than those who were lower risk to begin with, and therefore large amounts of change do not necessarily correspond directly with a lower overall risk level (Beggs & Grace, 2011). Studies that have controlled

for this by partialling pre-treatment scores out of the prediction equation (e.g., Beggs & Grace, 2011; Olver, Nicholaichuk, & Wong, 2014) have had greater success in being able to demonstrate a link between prosocial change and reductions in recidivism.

Other studies assessing the impact of change on risk level have employed a different approach known as clinically significant change methodology (e.g., Barnett, Wakeling, Mandeville-Norden, & Rakestrow, 2013; Olver, Beggs Christofferson, & Wong, 2015). In this method, post-treatment scores are evaluated against non-deviant norms to determine whether a given individual has qualitatively "improved", "recovered", is "already ok" (i.e., never scored outside the normative range), or remained "unchanged". Although this method of assessing changes that may impact on risk is user-friendly and easy to interpret, Olver and colleagues (2015) have overviewed the limitations of the method and noted mixed findings, in particular that the usefulness of the output is dependent on the quality of the measures used. In general, Olver et al. (2015) suggested that the use of a single, purpose-designed risk tool containing multiple dynamic factors, such as the VRS-SO or the STABLE 2007, may offer advantages over the psychometric battery approach for the consistent and meaningful applied measurement of risk and related change across relevant factors.

Although a number of existing third generation tools do not provide guidance as to the amendment of risk levels for a given individual over time, there are some measures that explicitly provide a formalized approach to the incorporation of individual change, including the VRS-SO. The VRS-SO includes both static and dynamic risk components, as well as protocols for the structured measurement of change across time that have a theoretical basis in the Transtheoretical Change Model (Prochaska, Diclemente, & Norcross, 1992). In this model, scores on dynamic items are assessed in conjunction with an assessment of an individual's motivational stage on each factor (pre-contemplation, contemplation, preparation, action, or maintenance). Modifications to risk ratings are obtained by adjusting

initial risk ratings according to progression across motivational stages for individual dynamic factors, although this adjustment is designed to be relatively modest. The interrater reliability of VRS-SO stages of change scoring was supported in a study by Olver, Wong, Nicholaichuk, and Gordon (2007), who found a "good" level of agreement in scoring (ICC = 0.68). This particular measure therefore allows for the identification of individual treatment targets, and provides a measure of static and dynamic risk, some assessment of responsivity issues, and information on treatment gains all incorporated into a single measure (for these reasons, the VRS-SO could arguably be considered a fourth-generation, rather than a third-generation, tool).

Clearly, more research is needed on the assessment of changes in risk, with numerous challenges having been identified for applied settings. As discussed above, dynamic risk assessments in applied settings typically impact greatly on individuals' progress through the criminal justice system, and assessments of change (across treatment or with continued repeat assessments) are certainly no different. For clinicians this carries a great responsibility, and the need to ensure that the methods we select to assess the changes made by our clients are both capable of detecting change that has occurred, and meaningful in terms of being predictive of actual reductions in the likelihood of recidivism. The issue of how to incorporate information related to individual change into assessments of risk becomes especially relevant when one considers the implications that a reduction in estimated risk may have for an individual and their progress through the system (and, by extension, implications for the community they will be returned to), including their chances for parole and what their parole conditions and level of oversight will be, decisions around whether to extend their custodial sentence, and the availability of particular rehabilitation programmes.

Conclusion

Given the high level of importance placed on the ability to accuracy predict whether a given individual is likely to offend in the future, the advancements that have been made in the development of risk assessment tools that are easy to understand and use have been important steps forward for the area of forensic psychology. The generational improvement upon risk assessment approaches has resulted in tools that are increasingly accurate in predicting risk, and have the potential to assist with case management beyond pure prediction. However, the utility and accuracy of these tools is diminished if incorrectly or inappropriately applied, and it is therefore vital that the limitations and challenges facing current favoured risk assessment approaches are highlighted and discussed. Some of these limitations have been outlined in the current chapter, including issues with the theoretical meaningfulness of factors being used in risk assessment tools, and the indistinct nature of factors being measured in risk assessments. Additionally, although the utility and potential applications for risk assessments have also improved over time, there are different conceptual and practical requirements for risk assessment tools depending on the tasks for which they are used, and these requirements are not always congruent. Further, there are issues with the amount of information available for the development of risk assessment approaches specific to important sub-groups within the sex offender population, including females and intimate-partner perpetrators, and for the clear understanding of how to best incorporate change (including treatment-induced change) into assessments of risk. Given the pervasive impact that assessments of risk have on the wellbeing and management of offenders, as well as the safety and wellbeing of those around them, we hope that the current discussion will prompt further thinking in any or all of these areas so that these challenges can be readily addressed and solutions incorporated into the assessment of risk for sexual offenders.

References

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric Assessment of Dynamic Risk Factors for Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, *19*(4), 347–367. http://doi.org/10.1007/s11194-007-9052-5

Andrews, D. A., & Bonta, J. (2016). *The psychology of criminal conduct* (6th ed.). Boston: Anderson Publishing Ltd.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory*. Toronto, Canada: Multi-Health Systems.

Andrews, D.A., Bonta, J., & Wormith, J. S. (2006). The Recent Past and Near Future of Risk and/or Need Assessment. *Crime & Delinquency*, *52*(1), 7–27. http://doi.org/10.1177/0011128705281756

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.

Barnett, G. D., Wakeling, H., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law*, *19*(1), 1–26. http://doi.org/10.1080/1068316X.2011.607820

Beauregard, E., Leclerc, B., & Lussier, P. (2012). Decision Making in the Crime Commission Process: Comparing Rapists, Child Molesters, and Victim-Crossover Sex Offenders. *Criminal Justice and Behavior*, *39*(10), 1275–1295. http://doi.org/10.1177/0093854812453120

Beech, A. R., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between

static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse : A Journal of Research and Treatment*, *14*(2), 155-167-197. http://doi.org/10.1177/107906320201400206

Beech, A. R., Parrett, N., Ward, T., & Fisher, D. (2009). Assessing female sexual offenders' motivations and cognitions: an exploratory study. *Psychology, Crime & Law*, *15*(2–3), 201–216. http://doi.org/10.1080/10683160802190921

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. http://doi.org/10.1177/1079063210369014

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), 182–92. http://doi.org/10.1037/a0022900

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. Vancouver, BC, Canada: Mental Health, Law and Policy Institute.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. http://doi.org/10.1037/0033-295X.111.4.1061

Cohen, J. (1994). The Earth Is Round ($p < .05$). *The American Psychologist*, *49*(12), 997–1003. http://doi.org/10.1037/0003-066X.49.12.997

Colliver, J. A., Conlee, M. J., & Verhulst, S. J. (2012). From test validity to construct validity... and back? *Medical Education*, *46*(4), 366–371. http://doi.org/10.1111/j.1365-

2923.2011.04194.x

Cording, J. R., Beggs Christofferson, S. M., & Grace, R. C. (2016). Challenges for the theory and application of dynamic risk factors. *Psychology, Crime and Law*, *22*(1–2), 84–103. http://doi.org/10.1080/1068316X.2015.1111367

Cortoni, F. (2010). The assessment of female sexual offenders. In T. A. Gannon & F. Cortoni (Eds.), *Female Sexual Offenders: Theory, assessment and treatment* (pp. 87–100). West Sussex, UK: John Wiley.

Cortoni, F., Hanson, R. K., & Coache, M.-E. (2010). The recidivism rates of female sexual offenders are low: A meta-analysis. *Sexual Abuse: Journal of Research and Treatment*, *22*(4), 387–401. http://doi.org/10.1177/1079063210372142

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. R. (2004). Limitations in actuarial risk assessment of sexual offenders: A methodological note. *British Journal of Forensic Practice*, *6*, 16–32.

Craissati, J., & Beech, A. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of Sexual Aggression*, *9*(1), 41–55. http://doi.org/10.1080/355260031000137968

Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment*, *14*(2), 121–138. http://doi.org/10.1177/107906320201400204

Ford, H. (2010). The treatment needs of female sexual offenders. In T.A. Gannon & F. Cortoni (Eds), *Female Sexual Offenders: Theory, assessment and treatment* (pp. 101-117). West Sussex, UK: John Wiley.

Gannon, T. A., & Rose, M. R. (2015). Bias in Female Child Molesters, 194–207.

Gannon, T. A., Rose, M. R., & Cortoni, F. (2010). Developments in female sexual offending and considerations for future research and treatment. In T. A. Gannon & F. Cortoni (Eds.), *Female Sexual Offenders: Theory, assessment and treatment* (pp. 181–198). West Sussex, UK: John Wiley.

Gannon, T. A., Ward, T., & Collie, R. (2007). Cognitive distortions in child molesters: Theoretical and research developments over the past two decades. *Aggression and Violent Behavior*, *12*(4), 402–416. http://doi.org/10.1016/j.avb.2006.09.005

Haig, B. D. (2012). From construct validity to theory validation. *Measurement: Interdisciplinary Research and Perspectives*, *10*(1–2), 59–62.

Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *The American Journal of Psychology, 126*(2), 135-153. doi:10.5406/amerjpsyc.126.2.0135

Haig, B. D. (2014). *Investigating the Psychological World: Scientific Method in the Behavioral Sciences*. Cambridge, MA.: MIT Press.

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism. (User report 97-04).* Ottawa.

Hanson, R. K., & Bourgon, G. (2016). Advancing sexual offender risk assessment: Standardized risk levels based on psychologically meaningful offender characteristics. In F. S. Taxman (Ed.), *Handbook on Risk and Need Assessment: Theory and Practice*. London, UK: Routledge.

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.

Hanson, R. K., & Harris, A. J. R. (2000). Where Should We Intervene?: Dynamic Predictors of Sexual Offense Recidivism. *Criminal Justice and Behavior*, *27*(1), 6–35. http://doi.org/10.1177/0093854800027001002

Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *13*(2), 105–122.

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 (No. 2007-05)*. Ottawa: Public Safety Canada.

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), 1–21. http://doi.org/10.1037/a0014421

Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders. (User report 1999-02).* Ottawa.

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP).* Burnaby, Canada: Simon Fraser University, Mental Health, Law and Policy Institute.

Heffernan, R., & Ward, T. (2015). The conceptualization of dynamic risk factors in child sex offenders: An agency model. *Aggression and Violent Behavior*, *24*, 250–260. http://doi.org/10.1016/j.avb.2015.07.001

Helmus, L. M., & Babchishin, K. M. (in press). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*.

Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: the Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment*, *14*(2), 103–

119. http://doi.org/10.1177/107906320201400203

Judge, J., Quayle, E., O'Rourke, S., Russell, K., & Darjee, R. (2014). Referrers' views of structured professional judgement risk assessment of sexual offenders: A qualitative study. *Journal of Sexual Aggression, 20*(1), 94-109. http://doi.org/10.1080/13552600.2013.767948

Jung, S., & Gulayets, M. (2011). Using clinical variables to evaluate treatment effectiveness in programmes for sexual offenders. *Journal of Sexual Aggression*, *17*(2), 166–180. http://doi.org/10.1080/13552601003802238

Kivlighan, D. M., & Tarrant, J. M. (2001). Does group climate mediate the group leadership-group member outcome relationship?: a test of Yalom's hypotheses about leadership priorities. *Group Dynamics*, *5*(3), 220–234. http://doi.org/10.1037//1089-2699.5.3.220

Marshall, W. L., Hamilton, K., & Fernandez, Y. (2001). Empathy deficits and cognitive distortions in child molesters. *Sexual Abuse : A Journal of Research and Treatment*, *13*(2), 123–130. http://doi.org/10.1177/107906320101300205

Nunes, K. L., & Babchishin, K. M. (2012). Construct Validity of Stable-2000 and Stable-2007 Scores. *Sexual Abuse: A Journal of Research and Treatment*, *24*(1), 29–45. http://doi.org/10.1177/1079063211404921

Ó Ciardha, C., & Gannon, T. A. (2011). Are the cognitive distortions of child molesters in need of treatment? *Journal of Sexual Aggression*, *17*(2), 130–141. http://doi.org/10.1080/13552600.2011.580573

Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2013). Incorporating Change Information Into Sexual Offender Risk Assessments Using the Violence Risk Scale-Sexual Offender Version. *Sexual Abuse: A Journal of Research and*

Treatment, *20*(10), 1–28. http://doi.org/10.1177/1079063213502679

Olver, M. E., Beggs Christofferson, S. M., & Wong, S. C. P. (2015). Evaluation and applications of the clinically significant change method with the Violence Risk Scale-Sexual Offender Version: Implications for Risk-Change Communication. *Behavioral Sciences and the Law*, *33*(1), 92–110.

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. (2014). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior*, *38*(6), 544–559. http://doi.org/http://dx.doi.org/10.1037/lhb0000086

Olver, M. E., Nicholaichuk, T. P., Kingston, D. a, & Wong, S. C. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology*, *82*(2), 312–24. http://doi.org/10.1037/a0035340

Olver, M. E., Nicholaichuk, T. P., & Wong, S. C. P. (2014). The predictive and convergent validity of a psychometric battery used to assess sexual offenders in a treatment programme: an 18-year follow-up. *Journal of Sexual Aggression*, *20*(2), 216–239. http://doi.org/10.1080/13552600.2013.816791

Olver, M. E., Wong, S. C. P., Nicholaichuk, T. P., & 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. http://doi.org/10.1037/1040-3590.19.3.318

Prentky, R. A., Lee, A. F. S., Knight, R. A., & Cerce, D. (1997). Recidivism Rates among Child Molesters and Rapists : A Methodological Analysis. *Law and Human Behavior*,

21(6), 635–659.

Prochaska, J. O., Diclemente, C. C., & Norcross, J. C. (1992). In Search of How People Change: Applications to Addictive Behaviors. *American Psychologist*, *47*, 1102–1114. http://doi.org/doi:10.3109/10884609309149692

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing Recidivism Risk Across Female Pathways to Crime. *Justice Quarterly*, *23*(3), 384–405. http://doi.org/10.1080/07418820600869152

Robertiello, G., & Terry, K. J. (2007). Can we profile sex offenders? A review of sex offender typologies. *Aggression and Violent Behavior*, *12*(5), 508–518. http://doi.org/10.1016/j.avb.2007.02.010

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*(5), 416–428. http://doi.org/10.1037/h0042040

Schwalbe, C. S. (2008). A Meta-Analysis of Juvenile Justice Risk Assessment Instruments. Predictive Validity by Gender. *Criminal Justice and Behavior*, *35*(11), 1367–1381. http://doi.org//10.1177/0093854808324377

Sigre-Leirós, V., Carvalho, J., & Nobre, P. (2015). Cognitive schemas and sexual offending: Differences between rapists, pedophilic and nonpedophilic child molesters, and nonsexual offenders. *Child Abuse and Neglect*, *40*, 81–92. http://doi.org/10.1016/j.chiabu.2014.10.003

Vess, J., & Skelton, A. (2010). Sexual and violent recidivism by offender type and actuarial risk: reoffending rates for rapists, child molesters and mixed-victim offenders.

Psychology, Crime & Law, *16*(7), 541–554. http://doi.org/10.1080/10683160802612908

Ward, T. (2013). Addressing the dual relationship problem in forensic and correctional practice. *Aggression and Violent Behavior, 18*(1), 92-100

Ward, T. (2016). Dynamic risk factors: scientific kinds or predictive constructs. *Psychology, Crime & Law*, *22*(1–2), 2–16. http://doi.org/10.1080/1068316X.2015.1109094

Ward, T. & Beech, A. (2015). Dynamic risk factors: A theoretical dead-end?. *Psychology, Crime & Law, 21,* 100-113. doi.org/10.1080/1068316X.2014.917854

Ward, T & Fortune, C. (2016). The role of dynamic risk factors in the explanation of offending. *Aggression and Violent Behavior, 29,* 79-88. doi: 10.1016/j.avb.2016.06.007

Wong, S. C. P., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale - Sexual Offender version (VRS-SO).* Saskatoon: Regional Psychiatric Centre and University of Saskatchewan.

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, *39*(12), 1511–1538. http://doi.org/10.1177/0093854812455741

Yesberg, J. A., Scanlan, J. M., Hanby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women's recidivism: Validating a dynamic community-based "gender-neutral" tool. *Probation Journal*, *62*(1), 33–48. http://doi.org/10.1177/0264550514562851