Article

Sexual Abuse
1–29

DOI: 10.1177/1079063219871572
journals.sagepub.com/home/sax
SAGE

STABLE-2007 Demonstrates Predictive and Incremental Validity in Assessing Risk-Relevant Propensities for Sexual Offending: A Meta-Analysis

Andrew E. Brankley[1,2], Kelly M. Babchishin[1], and R. Karl Hanson[1,2]

Abstract

STABLE-2007 is a measure of risk-relevant propensities for adult males convicted of a sexual offense. This meta-analysis evaluated the ability of STABLE-2007 and its items to discriminate between recidivists and nonrecidivists, and the extent to which STABLE-2007 improves prediction over and above Static-99R. Based on 21 studies (12 unique samples, $N = 6{,}955$), we found that STABLE-2007 was significantly and incrementally related to sexual recidivism, violent (nonsexual) recidivism, violent (including sexual) recidivism, and any crime. Scores on STABLE-2007 items and the three STABLE-2000 attitude items also discriminated between individuals who sexually reoffended and those who did not sexually reoffend. These findings support the use of STABLE-2007 in applied risk assessment practice and the interpretation of STABLE-2007 items as indicators of treatment and supervision targets.

Keywords

risk assessment, STABLE-2007, Static-99R, meta-analysis, sexual offender, dynamic risk factors

[1]Public Safety Canada, Ottawa, Ontario, Canada
[2]Ryerson University, Toronto, Ontario, Canada

Corresponding Author:
Andrew E. Brankley, Department of Psychology, Ryerson University, 350 Victoria St., Toronto, Ontario, Canada M5B 2K3.
Email: Andrew.Brankley@psych.ryerson.ca

Effective interventions for men convicted of sexual offenses require addressing the factors responsible for their offending (Hanson, Bourgon, Helmus, & Hodgson, 2009). To reduce the likelihood of reoffending, professionals need to focus their supervision and treatment efforts on client's risk-relevant propensities (i.e., problems that are related to the persistence of sexual offending). Identifying these propensities, however, require empirical justification.

Adult males with a history of sexual offenses may have a wide range of life problems, not all of which would be related to offending. For example, some professionals may assess low-victim empathy and lack of motivation at treatment intake because of their relevance to the process of treatment, but neither factor predicts sexual recidivism (Hanson & Morton-Bourgon, 2005; Mann, Hanson, & Thornton, 2010). In contrast, reliable predictors of sexual recidivism include atypical sexual interests (Hanson & Morton-Bourgon, 2005), antisocial tendencies (Hawes, Boccaccini, & Murrie, 2013), emotional congruence with children (McPhail, Hermann, & Nunes, 2013), and attitudes tolerant of sexual offending (Helmus, Hanson, Babchishin, & Mann, 2013). No single risk factor provides the sufficient or necessary conditions for the persistence of sexual offending, and consequently, evaluators typically consider a range of risk factors organized into structured risk assessment tools (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Kelley, Ambroziak, Thornton, & Barahal, 2018).

Bonta (1996) emphasized the distinction between risk assessment tools addressing static factors and those addressing dynamic risk factors. Static risk factors provide efficient indicators of risk based on criminal history and demographic information (e.g., age; Hanson & Morton-Bourgon, 2009). In contrast, dynamic risk factors capture putatively changeable risk-relevant problems (e.g., antisocial tendencies; Mann et al., 2010). It is important to remember, however, that both static and dynamic risk factors predict recidivism because they are markers for risk-relevant propensities (Mann et al., 2010). Indeed, static and dynamic risk items load onto similar risk-relevant propensities when submitted to factor analysis (Brouillette-Alarie & Hanson, 2015; Stockdale, Olver, & Wong, 2014). The difference between static and dynamic items is the type of information used to measure a specific risk-relevant propensity.

Tools assessing dynamic risk factors usually require more information and time to score than tools assessing static risk factors. Evaluators using dynamic risk tools typically integrate information from interviews, file review, and other collateral sources to assess the client's overall functioning. Compared with static risk tools, scoring dynamic risk tools typically requires greater professional expertise, background knowledge, and training. Dynamic risk factors are, however, more easily used to infer appropriate supervision and intervention strategies (i.e., case formulation). For example, a high score on an item addressing poor problem-solving skills can be used to guide discussion with a client about the need for vocational and skills training to help reduce reoffending risk.

Evaluators can now choose from a growing number of tools measuring static and dynamic risk factors. Although many features may influence evaluators' preference for a particular tool (or group of tools, Blais & Forth, 2014), relative predictive validity is an important consideration (Helmus & Babchishin, 2016; Steyerberg et al.,

2010). Relative predictive validity refers to the extent to which risk scores differ between those who sexually reoffend and those who do not. If dynamic factors have improved discrimination over and beyond static risk factors alone, this could also justify the additional time and resources required for scoring dynamic risk factors (Harris & Rice, 2003). Evaluating the predictive validity of individual dynamic items is also important because dynamic items are often used to direct intervention efforts (Wheeler, George, & Stephens, 2005).

A meta-analysis by van den Berg and colleagues (2018) found that dynamic tools and dynamic coding protocols predicted recidivism among men convicted of sexual offenses. These measures included, for example, the Risk for Sexual Violence Protocol (Hart et al., 2003), the Structured Risk Assessment–Forensic Version (Thornton & Knight, 2015), the Sex Offender Treatment Intervention and Progress Scale (McGrath, Lasher, & Cumming, 2012), the Violence Risk Scale: Sexual Offender version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003), and STABLE-2007 (Hanson, Harris, Scott, & Helmus, 2007), the latter being the measure that was the focus of the current review. Some of the dynamic risk measures reviewed by van den Berg et al. (2018) were also demonstrated to be incremental to static factors—k(samples) = 52, N = 13,446. Their meta-analysis considered sexual recidivism, violent recidivism, and nonviolent recidivism. There was, however, significant variability between studies for violent recidivism and the incremental effect for any crime. Their meta-analysis did not consider the predictive accuracy of individual risk tools.

The current review examined the predictive validity of one measure of dynamic risk factors: STABLE-2007. STABLE-2007 (Fernandez, Harris, Hanson, & Sparks, 2014; Hanson et al., 2007; Hanson, Helmus, & Harris, 2015) is the most widely used measure of dynamic risk for sexual recidivism in Canada and the United States (Bourgon, Mugford, Hanson, & Coligado, 2018; Kelley et al., 2018; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). STABLE-2007 items were empirically selected by identifying supervision problems that distinguished between adult males convicted of a sexual offense who had sexually reoffended while on community supervision from those who had a similar static risk score but did not reoffend (Hanson & Harris, 1998, 2000). Item definition and scoring of STABLE-2007 were refined over two previous versions and tested in a large prospective study (Hanson et al., 2007; Hanson et al., 2015).

Reports from evaluators who have implemented STABLE-2007 support its utility for guiding case management and treatment. Probation officers in Ireland, England, and Wales indicated that using STABLE-2007 helped them examine a wider range of relevant issues for their case reports and improved their confidence and consistency in decision making (McNaughton-Nicholls et al., 2010; Walker & O'Rourke, 2013). German treatment providers found STABLE-2007 items (i.e., sexual preoccupation and deviant sexual interests) useful to characterize the severity of paraphilic disorders (Briken & Muller, 2014). In New Zealand, ratings on earlier versions of STABLE-2007 were strongly related to professional judgments about whether adult males convicted of sexual offenses against children should receive long-term supervision orders (Ryan, Wilson, Kilgour, & Reynolds, 2014; Watson & Vess, 2007).

Although STABLE-2007 appears to provide helpful information for case management, previous research on the predictive and incremental validity of STABLE-2007 has been inconsistent. Replicating results from the construction sample, STABLE-2007 predicted all types of recidivism and was incremental to Static-99 for violent and nonviolent recidivism in an Austrian prison sample ($n = 263$; Eher, Matthes, Schilling, Haubner-Maclean, & Rettenberger, 2012). STABLE-2007, however, did not predict any type of recidivism in a high-intensity treatment program in Canada ($n = 180$; Sowden & Olver, 2017). Meta-analysis is a useful analytical tool to aggregate information across studies and determine if seemingly inconsistent findings are due to sampling error, or due to true variance that can be explained by the methodological features of the studies (e.g., length of follow-up, sample type).

The purpose of the current study was to conduct a meta-analysis of the predictive and incremental validity of STABLE-2007. We expected that STABLE-2007 total scores would consistently predict sexual recidivism, nonsexual violence, any violence (including sexual), and any crime, as well as that STABLE-2007 total scores would add incrementally to Static-99R. We also expected that STABLE items would be predictive of sexual recidivism.

Method

Selection of Studies

Online searches for studies examining the predictive and incremental validity of STABLE-2007 were conducted through Google Scholar, PubMed, PsycARTICLES, ProQuest Dissertations and Theses, and PsycINFO using the following keywords: "sex* offend*," "rap*," "molest*." These were integrated in two other searches using the Boolean "and" with "stable." These terms were further integrated using the Boolean "and" with the following key words: "reoffend*," "recidi*." We also used Google Scholar to search all documents citing the original government report describing the development of STABLE-2007 (i.e., Hanson et al., 2007). In addition, one of the developers of the tool (R. K. Hanson) reviewed the final study list for any known omissions.

As of April 25, 2018, our search yielded 21 studies consisting of 12 nonoverlapping samples (see the appendix). The vast majority of studies were produced in English ($k = 18$), and two studies were in German (Eher et al., 2013; Rettenberger, Matthes, Schilling, & Eher, 2011). Six studies reported on nonoverlapping samples, one study (Looman & Goldstein, 2015) contained two different samples that were divided into a routine and a high-risk/need sample, and a further 13 studies represented four overlapping samples, resulting in 12 unique samples. Table 1 presents the descriptive information of the included samples. The sample size ranged from 42 to 4,291 ($Mdn = 179.5$, $N_{total} = 6{,}955$). Six samples were described in peer-reviewed journal articles; the remainder were from published dissertations ($k = 2$), conference presentations ($k = 3$), and an unpublished dataset ($k = 1$). In addition to the information in the reports, the meta-analysis included additional data supplied by the authors of the original studies (see below).

Table 1. Descriptive Information of Samples.

						STABLE-2007						Static-99R		
									Low	Moderate	High			
S. No	*N*	Country	Location	Type	Age	*M* (*SD*)	Minimum	Maximum	*N* (%)	*N* (%)	*N* (%)	*M* (*SD*)	Minimum	Maximum
1	613	Canada	National	Routine	41.1 (13.7)	7.4 (5.0)	0	26	153 (24.9)	335 (54.6)	126 (20.5)	2.4 (2.4)	−3	10
2	4,291	Canada	British Columbia	Routine	40.8 (13.7)	7.5 (4.8)	0	25	931 (21.7)	2,529 (58.9)	832 (19.4)	2.4 (2.5)	−3	11
3	320	Canada	CSC	Routine	45.0 (13.3)	9.6 (4.7)	1	22	28 (8.8)	182 (56.9)	110 (34.4)	2.3 (3.0)	−3	11
4	179	The United States	Iowa State	Routine	33.9 (12.3)	7.2 (4.3)	0	21	37 (20.7)	112 (62.6)	30 (16.8)	3.0 (2.2)	−3	9
5	175	The United States	North Dakota	Treatment	45.5 (11.6)	13.7 (4.8)	1	26	7 (4.0)	41 (23.4)	127 (72.6)	1.2 (2.3)	−3	8
6	185	The United Kingdom	London Probation	Treatment	42.8 (12.9)	9.4 (4.2)	1	19	13 (7.0)	117 (63.2)	55 (29.7)	2.8 (2.4)	−3	9
7[a,b]	638	Austria	FECVSO Prison	Other[c]	—	12.2 (3.6)	4	23	0 (0.0)	—	—	3.3 (2.1)	0	10
8	76	The United Kingdom	Challenge Project	High Risk/ Need	42.3 (12.6)	10.8 (4.9)	2	21	6 (7.8)	36 (46.8)	35 (45.5)	3.8 (2.4)	−2	9
9[a,b]	92	Austria	FECVSO Forensic	High Risk/ Need	—	11.1 (3.6)	4	21	0 (0.0)	56 (60.9)	36 (39.1)	4.5 (2.2)	0	10
10	164	Canada	RTC	High Risk/ Need	34.6 (8.8)	14.6 (4.2)	3	24	2 (1.2)	32 (19.6)	129 (79.1)	4.5 (2.2)	−1	9
11	42	Canada	RTC	High Risk/ Need	39.3 (11.1)	14.6 (3.5)	8	22	0 (0.0)	9 (21.4)	33 (78.6)	6.1 (2.4)	0	10
12[b,d]	180	Canada	RPC	High Risk/ Need	39.0 (10.4)	14.8 (3.4)	4	24	0 (0.0)	31 (17.2)	149 (82.8)	4.9 (2.2)	−1	11
	6,955				40.0 [33.3, 46.6]	10.4 [10.1, 10.7]	0	26	1,177 (18.6)	3,480 (55.1)	1,662 (26.3)	3.1 [2.9, 3.3]	−3	11

Note. Fixed-effect meta-analysis, with 95% confidence intervals, presented for age, STABLE-2007 mean score, and Static-99R mean score. CSC = Correctional Service of Canada; FECVSO = Federal Evaluation Center for Violent and Sexual Offenders of the Austrian Prison Directorate; RTC = Regional Treatment Centre, Central District of CSC; RPC = Regional Psychiatric Center, Prairie Region of CSC.

[a]Sample used Static-99 instead of Static-99R.

[b]Original datasets were available for all samples except these three. Authors for these studies provided additional effect sizes and information for coding.

[c]This sample was coded as "other" because it had greater preselection than the treatment sample, but less preselection than the high-risk/need.

[d]Pretreatment effect sizes were used.

Recidivism was defined in most samples by new convictions ($k = 7$); three samples used new arrests or charges, and two samples did not specify the recidivism criteria. Sources of recidivism information included national criminal records ($k = 9$), local criminal records ($k = 5$), self-report ($k = 1$), institutional reports ($k = 2$), online news articles ($k = 1$), and probation officers ($k = 1$); note that these numbers do not equal to 12 as three samples used multiple sources of recidivism ($M = 1.5$, range = 1-4). Follow-up periods ranged from 23.5 to 169.8 months ($M = 78.4$; $k = 12$) for sexual recidivism, 23.8 to 173.4 months ($M = 75.4$; $k = 11$) for violent (nonsexual) recidivism, 23.4 to 130.1 months ($M = 69.9$; $k = 11$) for violent (including sexual) recidivism, and 22.9 to 111.6 months ($M = 66.8$; $k = 11$) for any crime. The observed recidivism rates were 8.2% (560/6,825) for sexual recidivism, 15.6% (1,043/6,679) for violent nonsexual recidivism, 19.9% (1,331/6,696) for violent (including sexual) recidivism, and 28.2% (1,937/6,876) for any crime.

Measures

STABLE-2007. STABLE-2007 (Fernandez et al., 2014; Hanson et al., 2015) is a measure of risk-relevant propensities for adult males convicted of a sexual offense against a child or nonconsenting adult. The 13 items of the tool describe problems that could be addressed as part of community supervision or correctional rehabilitation (e.g., impulsivity, hostility toward women). For individuals who did not offend against a child less than 14 years of age, there are only 12 items. Items are scored on a 3-point rating system of severity and summed to produce a total score (ranging from 0 to 26). The total score reflects the density of current risk-relevant propensities, and can be divided into three levels based on ± 1 standard deviation from the median score (*Mdn* = 7) in routine/complete samples: low (0-3), moderate (4-11), and high (12-26) density. STABLE-2007 can be mechanically combined with Static-99R (Hanson & Thornton, 1999, 2000), Static-2002R (Hanson, Helmus, & Thornton, 2010; Hanson & Thornton, 2003), or Risk Matrix 2000 (Thornton, 2010; Thornton et al., 2003) to produce an overall risk level and recommendations for intervention targets and dosage (Brankley, Helmus, & Hanson, 2017).

Static-99R. Static-99R is an empirically derived actuarial risk assessment tool designed to predict sexual recidivism in adult males charged or convicted of sexual offenses (Hanson & Thornton, 1999, 2000; Helmus, Thornton, Hanson, & Babchishin, 2011, 2012). Static-99R items are identical to Static-99 with the exception of updated age weights. The tool's 10 items assess criminal history, victim characteristics, and relationship history. For each item, one point is awarded for higher risk (with the exception of two items that are worth more than one point: prior sex offenses and age). The total score (ranging from −3 to 12) is calculated by summing all item points and can be used to place evaluees in one of the five standardized risk levels: Level I: "Very Low Risk" (–3, –2), Level II: "Below Average" (–1, 0), Level III:"Average Risk" (1, 2, 3), Level IVa: "Above Average" (4, 5), and Level IVb: "Well Above Average"

(6-12). Static-99R has been found to have moderate predictive accuracy for sexual recidivism (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012).

Coding Procedure

To be included in the current meta-analysis, the study had to include adult males convicted of a sexual offense, ratings on STABLE-2007 (or STABLE-2000 and sufficient information to compute STABLE-2007 scores), a follow-up period in which recidivism was recorded, and sufficient information to calculate an effect size. Authors from nine of the 12 samples provided their original datasets for the current meta-analysis; additional effect sizes and information were collected from the authors of the remaining samples (Eher et al., 2012; Eher et al., 2013; Rettenberger et al., 2011; Sowden & Olver, 2017). A coding manual containing a standard list of variables and explicit coding rules is included in the Supplemental Material (and available upon request). Each study was coded by the first and second author to conduct interrater analyses and to generate final consensus ratings. Ethics approval was not required as the current study presents a reanalysis of existing data.

Interrater reliability. Interrater reliability was based on nine samples (representing 17 studies) and excluded three practice cases. Continuous variables were assessed using absolute intraclass correlations (ICC) based on a two-way mixed design. Cicchetti (1994) suggested interpretive guidelines for ICC ratings of .40 as fair agreement, .60 as good agreement, and .75 as excellent agreement. Categorical variables were assessed using Cohen's κ statistic and percent agreement. Landis and Koch (1977) suggested interpretive guidelines for Cohen's κ of .21 for fair agreement, .41 for moderate agreement, .61 for substantial agreement, and .81 for almost perfect agreement.

Both raters coded effect sizes with high levels of agreement. The absolute ICC based on two-way mixed model and single measure was .999 (n effect size = 28). Interrater reliability for continuous descriptive variables ranged from ICC = .647 to 1 (Mdn = 1, n variable = 88). For categorical variables, interrater reliability ranged from 77.8% to 100% agreement (Mdn = 100%, n variable = 22) and kappa (κ) ranged from .609 to 1.0 (Mdn = 1.0, n variable = 22). None of the variables were excluded due to unacceptable interrater reliability. For all samples, a consensus rating between the two raters was completed after interrater reliability was conducted.

Overview of Analyses

Effect size. The effect size indicator used to summarize the predictive accuracy of STABLE-2007 was the hazard ratio from Cox regression survival analysis, which allows for unequal follow-up periods (Allison, 1984). In the current context, the hazard ratio is the ratio of the hazards (i.e., the likelihood that individuals will reoffend if they have not already done so) for individuals with different STABLE-2007 scores. It is an index of relative risk in that it provides information on a particular individual's likelihood of reoffending compared with others. For example, a hazard ratio of 1.10

indicated that each 1 score increase on STABLE-2007 increased the hazard by a factor of 1.10, or 10%. To provide summary descriptive information on the samples, the following variables were also averaged by fixed-effect meta-analysis: age of sample, STABLE-2007 mean score, and Static-99R mean scores (see Table 1).

To facilitate comparison with other studies, the direct effect of STABLE-2007 on sexual recidivism was also reported as the area under the receiver operating characteristic curve (AUC, Ruscio, 2008). Although AUC values are not influenced by base rate of the outcome, they are, unlike hazard ratios, influenced by the variance in the scores of the predictor variable (Hanson, 2008; Humphreys & Swets, 1991). There are no universal standards for describing effect sizes; however, Cohen's (1988) heuristics are widely used in psychology. For AUC values, Cohen's labels describe .56 as a small effect, .64 as a moderate effect, and .71 as a large effect (Rice & Harris, 2005).

Aggregation of findings. Findings across samples were aggregated using fixed-effect and random-effect meta-analysis weighted by the inverse of the variance (Borenstein, Hedges, Higgins, & Rothstein, 2009). What differentiates the fixed-effect and random-effect meta-analytic models is how they deal with variability between samples. A fixed-effect method assumes that all variability between samples are attributable to sampling error and that the participants come from the same population. A random-effect model incorporates variability across samples into the error term, allowing for the samples to represent different populations. When there is little variability across samples ($Q <$ degrees of freedom), random-effect and fixed-effect meta-analyses produce identical results (Borenstein et al., 2009). Random-effect models are often preferred because they are designed to reduce the possibility of Type 1 error (Borenstein et al., 2009; Schulze, 2007). However, as variability increases, the random-effect method gives more weight to smaller samples. Given that the between-study variability estimate needed for random-effect analyses loses precision when the analysis includes a small number of samples ($k < 30$; Schulze, 2007), we reported both types of analyses, but based our interpretations on the fixed-effect results.

To examine variability across samples, we used Cochran's Q statistic and the I^2 statistic (Borenstein et al., 2009). The Q statistic is a significance test for variability, whereas the I^2 is a measure of effect size for variability and can, thus, be compared across analyses, with varying number of samples. The I^2 statistic describes the proportion of the overall variability (i.e., Q) that is beyond what would be expected from random sampling error (Borenstein et al., 2009). The formula for I^2 is normally given as

$$I^2 = \frac{(Q - df)}{Q} \text{ x } 100,$$

where Q is the sum of squared deviations (of each effect size from the mean effect size) on a standardized scale and the degrees of freedom (df) is 1 less than the number of samples (k). I^2 was reported as a percentage and 25%, 50%, and 75% represents low,

moderate, and high variability, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

The process for identifying outliers followed rules established by Hanson and Bussière (1998). First, a sample was considered an outlier if (a) the variability (Q) was more than expected by chance ($p < .05$), (b) the effect size of the sample was the most extreme value, and (c) removing the sample reduced the variability across study (Q) by more than 50%. Meta-analyses can also be strongly influenced by unusually large sample sizes. The sample size of Sample 2 (Hanson, Helmus, Babchishin, & Zabarauckas, 2017) was 4,291, which contributed over half of the total sample size of this meta-analysis. To account for this usually large sample, the weight of Sample 2 (the largest weight) was reduced so that it would have 50% more weight than the next largest study.

Moderator analyses. Moderator analyses were conducted using $Q_{between}$ analyses. In these analyses, the overall variability among samples (measured by the Q statistic) was partitioned into the variability within each level of the moderator (e.g., published vs. not published). The $Q_{between}$ is calculated as $Q_{overall} - Q_{level1} - Q_{level2} - Q_{level\,i}$. The $Q_{between}$ can be tested for significance using a χ^2 distribution with the degrees of freedom equal to 1 less the number of categories in the moderator. Statistical power to detect significant moderators is influenced by the number of samples, the distribution of the moderator, and the proportion of variability between samples (Borenstein et al., 2009). As the current meta-analysis includes 12 samples, moderator analyses were limited to examining publication status, sample selection type (routine vs. high-risk/need), and country of origin (Canada vs. other).

Results

Sample Description

Table 1 provides basic descriptive information for the included samples. The samples came from Canada ($k = 6$), the United States ($k = 2$), the United Kingdom ($k = 2$), and Austria ($k = 2$) and were equally drawn from institutional ($k = 6$) and community settings ($k = 6$). Studies were produced between 2006 and 2017 ($Mdn = 2014$). Four of the samples were routine (representative of the general population of adult males convicted of a sexual offense); two samples included individuals who had been preselected on the basis of treatment needs; one sample was categorized as "other" because it had greater preselection than the treatment sample, but less preselection than the high-risk/need; and five samples included individuals who were preselected as high-risk/need (see Table 1).

The fixed-effect weighted average age for the samples was 40.4 years (95% confidence interval [CI] = [39.9, 41.0], $k = 10$, $n = 6{,}225$). Caucasian individuals were predominant in most samples ($k = 10$), Indigenous individuals were predominant in one sample (12), and the ethnic majority was unknown for one sample (11). Individuals

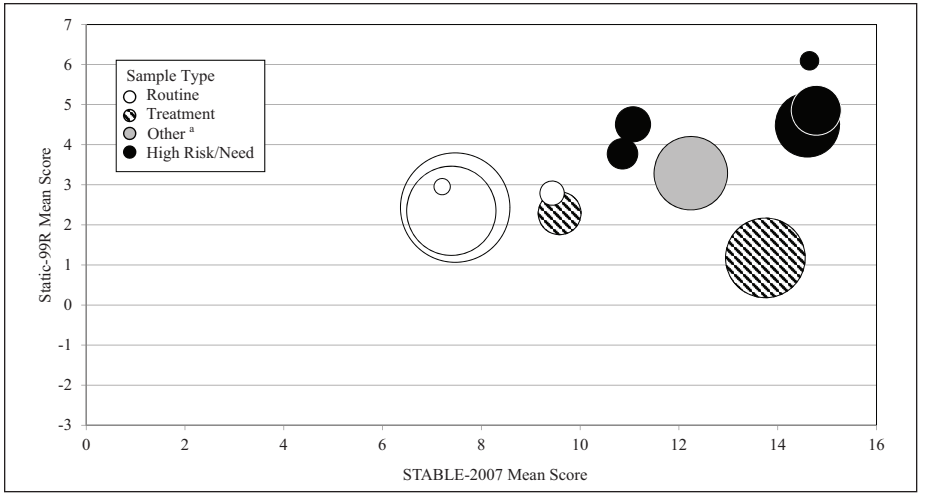

Figure 1. Relationship between sample type and mean scores on STABLE-2007 and Static-99R.

Note. Bubble sizes are directly proportional to the weight of the study in Table 3.

[a]Sample had greater preselection than the treatment sample, but less preselection than the high-risk/need.

with developmental or intellectual disabilities were included in five samples (i.e., <50% of the sample); their prevalence in the remaining samples ($k = 7$) was unknown. All samples included individuals who had sexually offended against adults or children, except for one sample (Webb, Craissati, & Keen, 2007) that had a mix of individuals convicted exclusively for possessing child sexual exploitation materials (CSEM) and individuals convicted of contact sexual offenses against children. Most samples ($k = 8$) included individuals who had received at least some treatment ($k = 4$ without treatment information).

STABLE-2007 and Static-99R scores were scored in a variety of methods in the primary studies: purely archival ($k = 6$), mixed archival and prospective ($k = 5$), and purely prospective ($k = 1$). Interrater reliability was reported using ICC in four studies. The median ICC for STABLE-2007 was .90 (range = .38-.92) and for Static-99R was .94 (range = .91-.98). The fixed-effect weighted average STABLE-2007 score was 10.4 (95% CI = [10.1, 10.7], $k = 12$, $n = 6{,}955$) and the fixed-effect weighted average Static-99R score was 3.1 (95% CI = [2.9, 3.3], $k = 12$, $n = 6{,}881$). More individuals were placed in the high STABLE-2007 level ($n = 1{,}662$, 26.3%) than the low STABLE-2007 level ($n = 1{,}177$, 18.6%); however, most individuals were in the moderate STABLE-2007 level ($n = 3{,}480$, 55.1%). Figure 1 presents the relationship between Static-99R, STABLE-2007, and sample type. Routine samples ($k = 3$) consistently had mean scores around the expected normative values for both tools (i.e., STABLE-2007 score of 7 and Static-99R score of 2); treatment samples had higher STABLE-2007 scores but not higher Static-99R scores compared with routine sam-

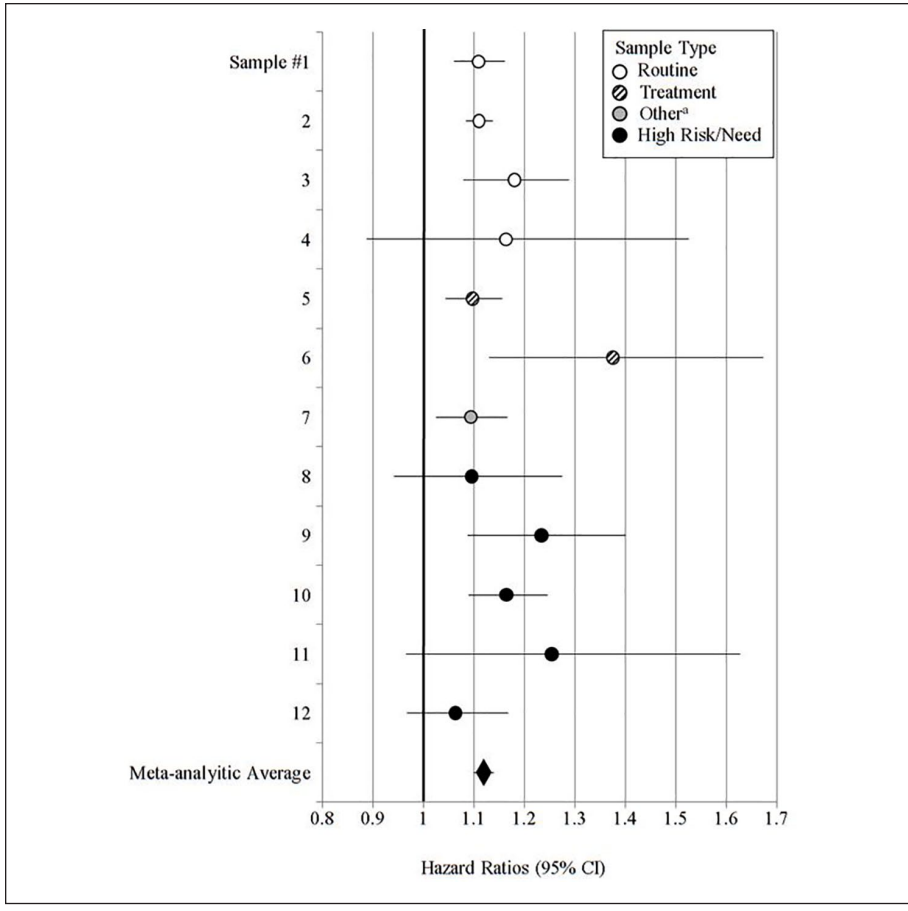

Figure 2. Distribution of raw hazard ratios across samples and the weighted average from a fixed-effect meta-analysis for STABLE-2007 discriminating between individuals who have and have not sexually reoffended.

Note. CIs that cross 1 are not significant at $p = .05$. CI = confidence interval.

[a]Sample had greater preselection than the treatment sample, but less preselection than the high-risk/need.

ples. The "other" type sample had higher Static-99R and STABLE-2007 scores compared with the routine sample, but not as high as the high-risk/need samples.

Direct Effect of STABLE-2007 Predicting Recidivism

The fixed-effect weighted hazard ratio for sexual recidivism was 1.12 (95% CI = [1.10, 1.14], $k = 12$), meaning that there was a 12% increase in the likelihood of sexual recidivism for every unit increase on STABLE-2007 (see Figure 2 for the distribution of hazard ratios across samples). The difference in the relationship between

STABLE-2007 and sexual recidivism observed across studies was small ($I^2 = 13.6\%$) and was not significant beyond what could be expected by sampling error ($Q = 12.73$, $p = .311$; see Table 2). That is, the predictive accuracy of the STABLE-2007 was consistent across studies. STABLE-2007 significantly predicted all types of recidivism (fixed-effect hazard ratios ranging from 1.09 to 1.10), with little between study variability across studies except for the outcome variable of any crime, which had moderate variability ($Q = 28.88$, $p = .002$, $I^2 = 61.9\%$). No statistically significant outlier was identified.

The AUC for the direct effect of STABLE-2007 predicting sexual recidivism was .674 (95% CI = [0.646, 0.702], $k = 12$, $n = 6{,}845$). This means that there was a 67% chance that a randomly selected sexual recidivist ($n_{\text{recidivists}} = 532$) would have a higher score than a randomly selected nonrecidivist ($n_{\text{nonrecidivists}} = 6{,}313$). The relationship between STABLE-2007 and sexual recidivism was consistent across studies ($I^2 = 0\%$) and was not more than what could be expected by sampling error ($Q = 10.20$, $df = 12$, $p = .512$).

Incremental Effect of STABLE-2007 to Static-99R

STABLE-2007 added incrementally to Static-99R in the prediction of all four recidivism outcomes (fixed-effect hazard ratios ranging from 1.05 to 1.07, see Table 3). The fixed-effect weighted hazard ratios for STABLE-2007 and sexual recidivism was 1.07 (95% CI = [1.04, 1.09], $k = 12$), meaning that for every unit increase on STABLE-2007, there was a 7% increase in the likelihood of sexual recidivism, after controlling for Static-99R scores. The fixed-effect weighted hazard ratios for Static-99R and sexual recidivism was 1.24 (95% CI = [1.19, 1.30], $k = 12$). These two effect sizes cannot be directly compared as both are relative to the range of scores on each tool. For all analyses, the variability between studies was no more than what would be expected by chance.

Predictive Validity of the STABLE Items

Table 4 presents the weighted hazard ratios for STABLE-2007 items and the three STABLE-2000 attitude items (i.e., sexual entitlement, rape attitudes, and child molester attitudes) for sexual recidivism. All 13 items significantly predicted sexual recidivism using either a fixed- or random-effect model—exp(β) ranged from 1.34 to 1.83, k ranged from 3 to 9, N ranging from 2,024 to 6,005. Variability across studies was nonsignificant for all but three items: general social rejection/loneliness ($Q = 28.88$, $p < .001$, $I^2 = 65.0\%$, $k = 9$), poor cognitive problem solving ($Q = 19.68$, $p = .012$, $I^2 = 59.4\%$, $k = 9$), and child molester attitudes ($Q = 12.06$, $p = .007$, $I^2 = 75.1\%$, $k = 4$). A statistical outlier was identified for child molester attitudes item (Sample 1).

Moderator Analyses

Three moderator analyses were conducted to examine factors that could influence the effect size for STABLE-2007 predicting sexual recidivism. The first moderator was

Table 2. Direct Effects of STABLE-2007 for Predicting Recidivism.

	Fixed-Effect		Random-Effects						
Recidivism Type	Exp(β)	[95% CI]	Exp(β)	[95% CI]	Q	I^2	N_{rec}/N_{total}	k	Studies
Sexual	1.12	[1.10, 1.14]	1.12	[1.10, 1.15]	12.73	13.6%	567/6901	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Violent (Non-Sexual)	1.09	[1.07, 1.11]	1.09	[1.06, 1.12]	15.19	34.2%	1056/6749	11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12
Violent (Including Sexual)	1.10	[1.08, 1.12]	1.11	[1.08, 1.13]	16.73	40.2%	1347/6766	11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12
Any Crime	1.10	[1.08, 1.11]	1.10	[1.08, 1.12]	28.88†	61.9%	1966/6952	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Note. Meta-analyses were conducted with Study 2's weight reduced to be 50% greater than the next largest study weight. No statistical outlier(s) identified.
Exp(β) = hazard ratio; CI = confidence interval.
†$p = .002$.

Table 3. Incremental Effects of the STABLE-2007 and Static-99R.

	Fixed-Effect		Random-Effects						
Recidivism Type	Exp(β)	[95% CI]	Exp(β)	[95% CI]	Q	$I2$	N_{rec}/N_{total}	k	Studies
Sexual									
STABLE-2007	1.07	[1.04, 1.09]	1.07	[1.04, 1.10]	11.86	7.2%	560/6825	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Static-99R	1.24	[1.19, 1.30]	1.24	[1.19, 1.30]	10.44	0.0%			
Violent (Non-Sexual)									
STABLE-2007	1.05	[1.02, 1.07]	1.05	[1.02, 1.07]	11.97	16.5%	1043/6679	11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12
Static-99R	1.21	[1.16, 1.25]	1.20	[1.16, 1.25]	11.17	10.5%			
Violent (Including Sexual)									
STABLE-2007	1.06	[1.04, 1.07]	1.06	[1.03, 1.08]	14.06	28.8%	1331/6696	11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12
Static-99R	1.21	[1.17, 1.25]	1.21	[1.16, 1.25]	11.98	16.5%			
Any Crime									
STABLE-2007	1.05	[1.03, 1.06]	1.05	[1.03, 1.07]	17.39	36.8%	1937/6876	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Static-99R	1.24	[1.20, 1.27]	1.23	[1.19, 1.27]	14.92	26.3%			

Note. Meta-analyses were conducted with Study 2's weight reduced to be 50% greater than the next largest study weight. Exp(β) = hazard ratio; CI = confidence interval.

Table 4. Predicting Sexual Recidivism With STABLE-2000/2007 Items.

	Fixed-effect		Random-effect						
STABLE-2000/2007 Items	Exp (β)	[95% CI]	Exp (β)	[95% CI]	Q	I^2	N_{rec}/N_{total}	k	Studies
1) Significant Social Influences	1.35	[1.20, 1.53]	1.36	[1.19, 1.55]	8.43	5.1%	444/5974	9	1, 2, 3, 4, 5, 6, 8, 10, 11
2) Capacity for Relationship Stability	1.68	[1.39, 2.02]	1.68	[1.39, 2.02]	6.09	1.4%	277/3941	7	1, 2, 3, 4, 8, 10, 11
3) Emotional Identification with Children	1.53	[1.20, 1.94]	1.53	[1.20, 1.94]	3.26	0.0%	168/2744	6	1, 2, 3, 8, 10, 11
4) Hostility Towards Women	1.42	[1.22, 1.64]	1.43	[1.20, 1.69]	9.57	16.4%	444/6005	9	1, 2, 3, 4, 5, 6, 8, 10, 11
5) General Social Rejection/Loneliness	1.34	[1.17, 1.54]	1.39	[1.06, 1.82]	22.88**[a]	65.0%	444/5979	9	1, 2, 3, 4, 5, 6, 8, 10, 11
6) Lack of Concern for Others	1.70	[1.47, 1.96]	1.70	[1.47, 1.96]	3.98	0.0%	441/5978	9	1, 2, 3, 4, 5, 6, 8, 10, 11
7) Impulsive Acts	1.80	[1.55, 2.10]	1.80	[1.54, 2.11]	8.22	2.6%	443/5971	9	1, 2, 3, 4, 5, 6, 8, 10, 11
8) Poor Cognitive Problem Solving	1.68	[1.44, 1.98]	1.61	[1.22, 2.12]	19.68*a	59.4%	443/5971	9	1, 2, 3, 4, 5, 6, 8, 10, 11
9) Negative Emotionality/Hostility	1.32	[1.12, 1.56]	1.34	[1.08, 1.66]	12.57	36.4%	442/5972	9	1, 2, 3, 4, 5, 6, 8, 10, 11
10) Sex Drive/Preoccupation	1.67	[1.44,1.94]	1.69	[1.43, 2.00]	9.47	15.5%	444/5977	9	1, 2, 3, 4, 5, 6, 8, 10, 11
11) Sex as Coping	1.58	[1.37, 1.82]	1.58	[1.37, 1.82]	7.58	0.0%	444/5975	9	1, 2, 3, 4, 5, 6, 8, 10, 11
12) Deviant Sexual Interests	1.41	[1.22, 1.64]	1.41	[1.22, 1.64]]	4.93	0.0%	277/3945	7	1, 2, 3, 4, 8, 10, 11
13) Cooperation with Supervision	1.59	[1.40, 1.81]	1.60	[1.31, 1.95]	14.35	44.2%	443/5977	9	1, 2, 3, 4, 5, 6, 8, 10, 11
Sexual Entitlement	1.65	[1.35, 2.01]	1.68	[1.20, 2.37]	8.43	64.4%	232/2608	4	1, 2, 5, 6
Rapist Attitudes	1.75	[1.38, 2.22]	1.75	[1.38, 2.22]	1.65	0.0%	231/2605	4	1, 2, 5, 6
Child Molester Attitudes	1.53	[1.22, 1.92]	1.64	[1.04, 2.57]	12.06*	75.1%	231/2600	4	1, 2, 5, 6
Outlier removed	1.83	[1.41, 2.38]	1.93	[1.26, 2.96]	5.03	60.2%	167/2024	3	2, 5, 6

Note. Exp(β) = hazard ratio. Meta-analyses were conducted with Study 2's weight reduced to 50% of the next largest study weight. Studies were included in analyses for Items 2, 3, and 12 if they were scored using STABLE-2007 rules or if there was sufficient information to transform items into reasonable STABLE-2007 approximations.
*$p < .05$. **$p < .01$. [a]No statistical outlier(s) identified.

the publication status (i.e., published vs. unpublished). The $Q_{between} = 0.24$, $p = .624$, was not statistically significant; published—fixed-effect $\exp(\beta)_{direct} = 1.11$, 95% CI = [1.08, 1.14], $Q = 8.66$, $k = 8$—and unpublished studies—fixed-effect $\exp(\beta)_{direct} = 1.14$, 95% CI = [1.10, 1.18], $Q = 2.40$, $k = 4$—provided similar hazard ratios. The predictive accuracy of STABLE-2007 was also found to be equivalent in routine and high-risk/need samples—$Q_{between} = 0.79$, $df = 1$, $p = .374$; routine fixed-effect $\exp(\beta)_{direct} = 1.12$, 95% CI = [1.09, 1.15], $Q = 1.78$, $I^2 = 0\%$, $k = 4$; and high risk/need fixed-effect $\exp(\beta)_{direct} = 1.14$, 95% CI = [1.09, 1.20], $Q = 4.79$, $I^2 = 16.6\%$, $k = 5$. Finally, country of origin (i.e., Canada vs. other) was also not a statistically significant moderator ($Q_{between} = 0.01$, $df = 1$, $p = .920$), with the hazard ratios being similar for Canadian—fixed-effect $\exp(\beta)_{direct} = 1.12$, 95% CI = [1.09, 1.15], $Q = 4.97$, $I^2 = 0\%$, $k = 6$—and non-Canadian—fixed-effect $\exp(\beta)_{direct} = 1.12$, 95% CI = [1.08, 1.16], $Q = 7.75$, $I^2 = 35.5\%$, $k = 6$—samples.

Discussion

The purpose of the current study was to evaluate the predictive and incremental validity of STABLE-2007, a commonly used measure of risk-relevant propensities for adult males convicted of a sexual offense. As expected, STABLE-2007 predicted all types of recidivism and was incremental to Static-99R. In addition, each STABLE-2007 item predicted sexual recidivism, supporting their utility in informing treatment and supervision targets.

Meta-analysis is a valuable tool to determine whether apparent variability in results across studies is due to true variability across samples or if the variability is within the range expected by sampling error. In the 12 samples examined in this meta-analysis, the relationship between STABLE-2007 and recidivism was generally within the range expected by chance. The only exception was the significant variability in the univariate (direct) effect of STABLE-2007 for any recidivism, which was no longer significant in the multivariate analyses that included STABLE-2007 along with Static-99R total scores. As well, for 12 of the 15 STABLE-2007/STABLE-2000 items, the between-studies variability was no greater than would be expected by chance. Although nonsignificant variability does not preclude the existence of moderators, no moderator effects were found.

STABLE-2007 was incremental to Static-99R across all outcomes. Incremental findings are necessary, but not sufficient, to justify combining measures into an overall estimate of recidivism risk. Previous research has shown that combining even highly correlated risk tools can improve prediction (Babchishin, Hanson, & Helmus, 2012; Lehmann et al., 2013); however, the practical gains may be small. To justify the inclusion of an additional risk assessment tool, evaluators need to consider the costs (e.g., effort, professional resources) against the benefits.

STABLE-2007 was designed to support inferences concerning priority areas for treatment and supervision, and practitioners have perceived it as useful for this function (Briken & Muller, 2014; Walker & O'Rourke, 2013). Although it is possible to

infer certain risk-relevant propensities from static risk items (Brouillette-Alarie & Hanson, 2015), STABLE-2007 items more directly speak to the needs areas to be addressed in recidivism reduction interventions. For example, "ever lived with a lover" from Static-99R and "capacity for relationship stability" from STABLE-2007 are both related to the capacity to form and maintain intimate relationships. Assessing this propensity in Static-99R is limited to the duration of a consenting age-appropriate relationship, whereas the STABLE-2007 item also directs evaluators to consider the client's motivation and the quality of the relationship (Fernandez et al., 2014). STABLE-2007 items allow for reassessment to identify improvements or deterioration. In contrast, Static-99R scores only describe risk at one time point, that is, at the time of release of index sexual offense. Contrary to expectation, the STABLE-2000 attitude items that were removed from STABLE-2007 predicted sexual recidivism as strongly as the other STABLE-2007 items. This finding should caution evaluators and test developers from placing too much emphasis on the findings from any single study. Although the STABLE-2000 attitude items did not predict sexual recidivism in the development sample (Hanson et al., 2007), subsequent studies found that diverse measures of sex offense–supportive attitudes (including the STABLE-2000 items) predicted sexual recidivism (Helmus et al., 2013). These findings highlight an intrinsic challenge of using empirically derived actuarial risk tools. Given that risk tools cannot claim to measure all relevant risk factors, evaluators often struggle with how to address risk factors known to be important, but which are not included in the actuarial scheme (Hanson, 1998). Considering the importance of attitudes to cognitive-behavioral interventions, their exclusion from STABLE-2007 now seems regrettable.

The difference in average STABLE-2007 scores across sample types reinforces the importance of using representative samples when developing norms for actuarial risk tools. For routine samples, the average scores for both Static-99R and STABLE-2007 were consistent. In contrast, STABLE-2007 scores were higher than expected in samples preselected for treatment or preselected as high-risk/need samples. Interestingly, the increase in STABLE-2007 scores for treatment samples was not associated with a corresponding increase in Static-99R scores. That is, the same Static-99R total score was associated with greater densities of risk-relevant propensities in treatment samples than in routine samples. Although it is possible to make inferences about risk-relevant propensities based solely on static risk factors (Brouillette-Alarie, Babchishin, Hanson, & Helmus, 2016; Brouillette-Alarie, Proulx, & Hanson, 2018), such inferences are strengthened when the sample type is known. Conversely, knowing the density of an individual's criminogenic needs (i.e., STABLE-2007 total score) can support inferences about the appropriate normative reference group for Static-99R and Static-2002R users (Hanson, Thornton, Helmus, & Babchishin, 2016; Phenix, Helmus, & Hanson, 2016).

Implications for Research

The relationship between STABLE-2007 items and recidivism other than sexual recidivism has yet to be examined, and may show greater variability than that observed for

sexual recidivism. Factor analyses of STABLE-2007 reveal that some items load onto a general propensity for rule-violation domain whereas the remaining items load onto a sex crime–specific problems domain (Brouillette-Alarie & Hanson, 2015). Given that the inclusion of static items assessing sex crime–specific problems depreciates the prediction of general and violent recidivism (Babchishin, Hanson, & Blais, 2016), the sex crime–specific items from STABLE-2007 (e.g., emotional congruence with children, deviant sexual interests) may be unrelated, or even negatively related, to nonsexual recidivism. Consequently, it may be possible to improve STABLE-2007's ability to predict general or violent recidivism by only considering items related to general criminality.

It is likely that STABLE-2007 items represent more than two underlying constructs (i.e., general and sexual criminality). For example, diverse paraphilias are coded under a single item for Deviant Sexual Interests, and the clinical correlates of sadism are quite different from those of pedophilia. Sexual Preoccupation and Sexualized Coping may be part of an another, distinct construct (Kafka, 2010). The Emotional Identification With Children item explicitly addresses (a) emotional attraction to children and childhood, (b) sexual interest in imagining themselves as a child (autopedophilia; Hsu & Bailey, 2017), and (c) attributing adult qualities to children (Fernandez et al., 2014). The links between these constructs are largely unknown, and may vary based on the preexisting relationship between the offender and the victim (McPhail et al., 2013). Furthermore, the constructs measured by the STABLE-2007 Emotional Identification With Children item may predict sexual recidivism largely due to their association with pedophilic sexual interests (Brankley, 2019; Hermann, McPhail, Helmus, & Hanson, 2017). Consequently, further research on the construct validity of the STABLE-2007 items could lead to new or revised items that advance the measurement of psychologically meaningful risk factors for sexual recidivism.

Given that static and dynamic risk factors are generally incremental to one another (van den Berg et al., 2018), further research is also needed on how best to integrate and interpret tools measuring static and dynamic risk factors into an overall judgment on risk. One approach is to use assessment tools to inform an individual's placement within the standardized risk levels (Hanson, Bourgon, et al., 2017). For example, evaluators could start with risk-level placement based on Static-99R scores (Hanson, Babchishin, et al., 2017), and then adjust risk-level placement based on unusually high or unusually low STABLE-2007 scores (see Brankley et al., 2017). Placement within the standardized risk levels could also be achieved by other empirical actuarial tools with static and dynamic risk factors, such as VRS-SO (Olver et al., 2018), or even using a structured professional judgment approach. The ease, accuracy, and utility of these various approaches remain an important research topic given that evaluators typically use multiple risk tools (Kelley et al., 2018; Neal & Grisso, 2014).

Although STABLE-2007 was intended to measure change (Hanson et al., 2007), there is surprisingly little evidence that the risk factors it measures are actually dynamic (Kraemer et al., 1997). There is ongoing scientific debate about the mechanisms or extent of the change, but individuals' capacity for change is generally accepted (e.g., Laub & Sampson, 2001). To justify reassessments, however, changes in STABLE-2007

scores should be related to changes in recidivism risk. Such an evaluation requires three or more time-points to determine if the observed change is a true change rather than measurement error (Babchishin, 2013). The current meta-analysis used only the first STABLE-2007 assessment, effectively treating STABLE-2007 scores as static (unchanging) risk indicators. Further research, using multiple time-points, is needed to evaluate the rate of change over time and the extent to which subsequent assessments improve prediction.

A final implication is that ad hoc comparison groups for treatment outcome studies are likely to be biased if matching is based solely on static risk. Random group assignment is difficult in correctional settings; consequently, researchers often use ad hoc comparison groups and attempt to correct for this bias by controlling for static risk scores (Hanson et al., 2002). Individuals, however, are not selected into treatment or high-risk/need groups based only on static risk factors; they are also likely to have an unusually high density of treatment needs. Consequently, controlling only for static risk scores cannot be expected to result in comparable groups. To the extent that the groups differ on these additional risk factors, solely controlling for static risk will underestimate the treatment effect in most cases. A more appropriate comparison could be made after controlling for both static (e.g., Static-99R) and dynamic risk tools or other indicators of treatment needs (e.g., STABLE-2007 total scores). The conclusion of a recent, large scale, matching study of a prison-based sexual offender treatment program in England and Wales (Mews, Di Bella, & Purver, 2017) is severely limited given they did not additionally control for treatment needs. As such, the groups could not be determined to be equal in risk at the start of treatment, and thus, conclusions regarding the treatment effect cannot assume group equality.

Implication for Practice

For Static-99R and Static-2002R users, the current findings support the use of STABLE-2007 to inform the selection of recidivism tables. Static-99R and Static-2002R have separate recidivism rate tables for routine and for high-risk/need samples (Hanson et al., 2016). The user guidance for these tools (Phenix et al., 2016) recommends that the decision concerning the reference group is based on the density of treatment needs as measured by a risk tool external to Static-99R, such as STABLE-2007. In the current collection of studies, the average STABLE-2007 score for the high-risk/need groups ranged from 10.8 to 14.8 ($Mdn = 14.6$) compared with average STABLE-2007 scores ranging from 7.2 to 9.6 ($Mdn = 7.4$) for the routine samples. Although there is no definitive threshold that distinguishes the routine from high-risk/need groups, it is plausible to expect that individuals with high STABLE-2007 scores of 12 or higher are to be more commonly found in the high-risk/need samples than in the routine samples. Evaluators are cautioned, however, that there is still no research on whether choosing between Static-99R/Static-2002R reference groups improves calibration.

Another method to integrate Static-99R (or Static-2002R) and STABLE-2007 is to directly combine the scores based on the standardized risk levels (Hanson, Babchishin, et al., 2017) and reference the expected recidivism rates provided for the combined

tools (Brankley et al., 2017). This option is justified given the incremental validity of STABLE-2007 and Static-99R; however, the Static/STABLE combination rules only provide recidivism rates up to 5 years, whereas 10-year recidivism rates are provided for the STATIC norms (Phenix et al., 2016). In addition, the recidivism information provided by the Static/STABLE combination is based on a much smaller sample size compared with the recidivism information provided from the STATIC norms.

The current results are also relevant for evaluators working with individuals whose criminal histories include CSEM offenses but do not include contact sexual offenses (i.e., CSEM-exclusive; Babchishin, Hanson, & VanZuylen, 2015). For individuals who exclusively have CSEM offenses, the use of existing risk tools developed and validated on individuals with contact sexual offenses is not recommended (e.g., Babchishin, Merdian, Bartels, & Perkins, 2018). Currently, the Child Pornography Offender Risk Tool (CPORT; Seto & Eke, 2015) is the only risk tool developed and validated on individuals with CSEM offenses (including those with contact sexual offenses). Although the CPORT predicted recidivism for the subgroup of individuals who exclusively have CSEM offenses, the predictive accuracy was lower than individuals with both CSEM and contact sexual offenses and was not predictive of recidivism for individuals who exclusively had CSEM offenses without any other criminal history (Seto & Eke, 2015). Given most individuals who exclusively have CSEM offenses would score zero on the criminal history items of the CPORT, these findings are likely symptomatic of the fact that it is difficult to discriminate between individuals who have similar scores on a tool (see Howard, 2017, for distributional limitations of AUCs). STABLE-2007 does not include criminal history items and, as such, offers a promising method of assessing the risk among adult males who exclusively have CSEM offenses. Within a sample (6) that included a significant number of participants with CSEM offenses (n = 72, 68 were CSEM-exclusive), STABLE-2007 predicted sexual recidivism (AUC = 0.94, 95% CI = [0.85, 1.00], $N_{\text{recidivists}} = 2$, $N_{\text{total}} = 72$). Both new convictions were for new CSEM offenses. Individuals with CSEM offenses were also present in four additional samples (1, 4, 5, and 7) in the current meta-analysis. This provides initial, albeit limited, evidence to support the use of STABLE-2007 with adult males convicted of CSEM-exclusive offenses. Future studies replicating the applicability of dynamic tools on samples of adult males convicted of CSEM offenses would be important to support this recommendation.

Conclusion

Effective intervention for individuals with sexual offenses requires addressing the factors responsible for their offending. Our findings indicated that the items included in STABLE-2000 and STABLE-2007 predicted sexual recidivism, and that STABLE-2007 total scores incrementally contributed over Static-99R scores to the prediction of sexual, violent, and any recidivism. Consequently, these items are plausible targets for supervision and rehabilitation efforts. Future research is needed to determine the extent to which interventions designed to address these risk-relevant propensities are actually effective in reducing the recidivism risk of individuals with a history of sexual offending.

Appendix. Studies Used in the Meta-Analysis.

Study number	Sample	Study
1.1	Dynamic Supervision Project	Hanson, Harris, Scott, and Helmus (2007)
1.2	Dynamic Supervision Project	Helmus and Hanson (2012)
1.3	Dynamic Supervision Project	Helmus, Babchishin, and Blais (2012)
1.4	Dynamic Supervision Project	Helmus, Hanson, Babchishin, and Thornton (2015)
1.5	Dynamic Supervision Project	Hanson, Helmus, and Harris (2015)
2.1	British Columbia Corrections	Lussier, Deslauriers-Varin, and Râtel (2010)
2.2	British Columbia Corrections	Hanson, Helmus, Babchishin, and Zabarauckas (2017)
3	Millhaven 2015	Looman and Goldstein (2015)[a]
4	Iowa	Smeth (2013)
5	North Dakota	Saum (2004)
6	London Probation	Webb, Craissati, and Keen (2007)
7.1	Austrian Prison	Eher, Rettenberger, Matthes, and Schilling (2010)
7.2	Austrian Prison	Rettenberger, Matthes, Schilling, and Eher (2011)
7.3	Austrian Prison	Eher, Matthes, Schilling, Haubner-Maclean, and Rettenberger (2012)
7.4	Austrian Prison	Eher, Olver, Heurix, Schilling, and Rettenberger (2015)
7.5	Austrian Prison	Etzler, Eher, and Rettenberger (2018)
8	Challenge	Craissati and Blundell (2013)
9	Austrian Forensic	Eher et al. (2013)
10	Regional Treatment Centre 2012	Looman and Abracen (2012)
11	Regional Treatment Centre 2015	Looman and Goldstein (2015)[a]
12.1	Clearwater	Sowden (2013)
12.2	Clearwater	Sowden and Olver (2017)

[a]Looman and Goldstein (2015) contributed two unique, nonoverlapping samples (3 and 11).

Authors' Note

The views expressed are those of the authors and not necessarily those of Public Safety Canada. The authors take responsibility for the integrity of the data and the accuracy of the data analyses, and have made every effort to avoid inflating statistically significant results.

Acknowledgments

The authors thank the following colleagues who provided data analyses not included in other reports: Jeff Abracen, Jackie Craissati, Reinhard Eher, Sonja Etzler, Jan Looman, Mark Olver, Martin Rettenberger, and Angela Smeth.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: R. K. Hanson is a certified Static-99R/STABLE-2007 trainer. The copyright for these instruments is held by the Government of Canada.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this project was provided in part by the Vanier Canada Graduate Scholarship Program (A. E. Brankley).

ORCID iD

Andrew E. Brankley https://orcid.org/0000-0001-7606-7779

Supplemental Material

Supplemental material for this article is available online.

References

References marked with asterisk indicate studies included in one or more of the meta-analyses. Study numbers are provided in square brackets for these references.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data* (*Vol. 46*). Thousand Oaks, CA: SAGE

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*, 84-94. doi:10.1207/s15327752jpa8701_07

Babchishin, K. M. (2013). *Sex offenders do change on risk-relevant propensities: Evidence from a longitudinal study of the Acute-2007* (Unpublished doctoral dissertation). Carleton University, Ottawa, Ontario, Canada. Retrieved from https://curve.carleton.ca/system/files/theses/27626.pdf

Babchishin, K. M., Hanson, R. K., & Blais, J. (2016). Less is more: Using Static-2002R subscales to predict violent and general recidivism among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *28*, 187-217. doi:10.1177/1079063215569544

Babchishin, K. M., Hanson, R. K., & Helmus, L.-M. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, *19*, 442-461. doi:10.1177/1073191112458312

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*, 45-66. doi:10.1007/s10508-014-0270-x

Babchishin, K. M., Merdian, H. L., Bartels, R. M., & Perkins, D. (2018). Child sexual exploitation materials offenders: A review. *European Psychologist*, *23*, 130-143. doi:10.1027/1016-9040/a000326

Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior*, *38*, 531-543. doi:10.1037/lhb0000082

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley

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*, 167-205. doi:10.3138/cjccj.2016-0024

Brankley, A. E. (2019). *A taxometric analysis of pedophilia in adult males convicted of sexual offences: Evidence for a taxon* (Unpublished doctoral dissertation). Ryerson University, Toronto, Ontario, Canada

Brankley, A. E., Helmus, L.-M., & Hanson, R. K. (2017). *STABLE-2007 evaluator workbook: Updated recidivism rates (includes combinations with Static-99R, Static-2002R, and Risk Matrix 2000)*. Unpublished report, Public Safety Canada, Ottawa, Ontario.

Briken, P., & Muller, J. L. (2014). Beurteilung der schuldfahigkeit bei paraphiler storung; kann der schweregrad der storung mithilfe von kriterien aus prognoseinstrumenten erfasst werden? [Assessment of criminal responsibility in paraphilic disorder: Can the severity of the disorder be assessed with items of standardized prognostic instruments?]. *Der Nervenarzt*, *85*, 304-311. doi:10.1007/s00115-013-3901-x

Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L.-M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution. *Assessment*, *23*, 96-111. doi:10.1177/1073191114568114

Brouillette-Alarie, S., & Hanson, R. K. (2015). Comparaison de deux mesures d'évaluation du risque de récidive des délinquants sexuels. [Comparison of two assessment measures of recidivism for sex offenders]. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *47*, 292-304. doi:10.1037/cbs0000019

Brouillette-Alarie, S., Proulx, J., & Hanson, R. K. (2018). Three central dimensions of sexual recidivism risk: Understanding the latent constructs of Static-99R and Static-2002R. *Sexual Abuse: A Journal of Research and Treatment*, *30*, 676-704. doi:10.1177/1079063217691965

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284-290. doi:10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum

*Craissati, J., & Blundell, R. (2013). A community service for high-risk mentally disordered sex offenders: A follow-up study. *Journal of Interpersonal Violence*, *28*, 1178-1200. doi:10.1177/0886260512468235 [Sample 8]

*Eher, R., Matthes, A., Schilling, F., Haubner-Maclean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 5-28. doi:10.1177/1079063211403164 [Sample 7.3]

*Eher, R., Olver, M. E., Heurix, I., Schilling, F., & Rettenberger, M. (2015). Predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment. *Law and Human Behavior*, *39*, 571-580. doi:10.1037/lhb0000144 [Sample 7.4]

*Eher, R., Rettenberger, M., Gaunersdorfer, K., Haubner-MacLean, T., Matthes, A., Schilling, F., & Mokros, A. (2013). Über die treffsicherheit der standardisierten risikoeinschätzungsverfahren Static-99 und Stable-2007 bei aus einer sicherungsmaßnahme entlassenen sexualstraftätern [On the accuracy of the standardardized risk assessment procedures Static-99 and Stable-2007 for sexual offenders released from detention]. *Forensische Psychiatrie, Psychologie, Kriminologie*, *7*, 264-272. doi:10.1007/s11757-013-0212-9 [Sample 9]

*Eher, R., Rettenberger, M., Matthes, A., & Schilling, F. (2010). Stable dynamic risk factors in child sexual abusers: The incremental predictive power of narcissistic personality traits beyond the Static-99/STABLE-2007 priority categories on sexual reoffense. *Sexual Offender Treatment*, *10*, 1-12. Retrieved from http://www.sexual-offender-treatment.org/2-2010_02.html [Sample 7.1]

*Etzler, S., Eher, R., & Rettenberger, M. (2018). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the STABLE-2007. *Assessment*. Advance online publication. doi:10.1177/1073191118754705 [Sample 7.5]

Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual—Revised 2014*. Unpublished report, Public Safety Canada, Ottawa, Ontario

Hanson, R. K. (1998). What do we know about sexual offender risk assessment? *Psychology, Public Policy, and Law*, *4*, 50-72. doi:10.1037/1076-8971.4.1-2.50

Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene*, *14*, 15-17

Hanson, R. K., Babchishin, K. M., Helmus, L.-M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment*, *29*, 582-597. doi:10.1037/pas0000371

Hanson, R. K., Bourgon, G., Helmus, L.-M., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, *36*, 865-891. doi:10.1177/0093854809338545

Hanson, R. K., Bourgon, G., McGrath, R., 'Kroner, D. D., Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. New York, NY: The Council of State Governments Justice Center. Retrieved from https://csgjusticecenter.org/wp-content/uploads/2017/01/A-Five-Level-Risk-and-Needs-System_Report.pdf

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*, 348-362. doi:10.1037//0022-006X.66.2.348

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment of sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 169-194. doi:10.1177/107906320201400207

Hanson, R. K., & Harris, A. J. R. (1998). *Dynamic predictors of sexual recidivism* (User Report 1998-01). Ottawa, Ontario: Department of the Solicitor General of Canada. Retrieved from www.ps-sp.gc.ca/res/cor/rep

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offence recidivism. *Criminal Justice and Behavior*, *27*, 6-35. doi:10.1177/0093854800027001002

*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, Corrections Research). Ottawa, Ontario: Public Safety Canada. Retrieved from https://vawnet.org/material/assessing-risk-sexual-offenders-community-supervision-dynamic-supervision-project [Sample 1.1]

*Hanson, R. K., Helmus, L.-M., Babchishin, K. M., & Zabarauckas, C. L. (2017). *Risk assessment and recidivism information of sex offenders in the Province of British Columbia* [Data file]. [Sample 2.2]

Hanson, R. K., Helmus, L.-M., & Thornton, D. (2010). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior*, *34*, 198-211. doi:10.1007/s10979-009-9180-1

*Hanson, R. K., Helmus, L.-M., & Harris, A. J. R. (2015). Assessing the risk and needs of supervised sexual offenders: A prospective study using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior*, *42*, 1205-1224. doi:10.1177/0093854815602094 [Sample 1.5]

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*, 1154-1163. doi:10.1037/0022-006X.73.6.1154

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*, 1-21. doi:10.1037/a0014421

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 1999-02). Ottawa: Department of the Solicitor General of Canada. Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/sttc-mprvng-actrl/index-en.aspx

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for individuals convicted of sexual offences: A comparison of three actuarial scales. *Law and Human Behavior*, *24*, 119-136. doi:10.1023/A:1005482921333

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of the Static-2002* (User report 2003-01). Ottawa: Department of the solicitor General of Canada. Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/nts-dvlpmnt-sttc/index-en.aspx

Hanson, R. K., Thornton, D., Helmus, L.-M., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sexual Abuse: A Journal of Research and Treatment*, *28*, 218-252. doi:10.1177/1079063215574710

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. *Annals of New York Academy of Sciences*, *989*, 198-210. doi:10.1111/j.1749-6632.2003.tb07306.x

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk for sexual violence*. Vancouver, British Columbia, Canada: The Institute Against Family Violence

Hawes, S. W., Boccaccini, M. T., & Murrie, D. C. (2013). Psychopathy and the combination of psychopathy and sexual deviance as predictors of sexual recidivism: Meta-analytic findings using the Psychopathy Checklist–Revised. *Psychological Assessment*, *25*, 233-243. doi:10.1037/a0030391

Helmus, L.-M., & Babchishin, K. M. (2016). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, *44*, 8-25. doi:10.1177/0093854816678898

*Helmus, L.-M., Babchishin, K. M., & Blais, J. (2012). Predictive accuracy of dynamic risk factors for aboriginal and non-aboriginal sex offenders: An exploratory comparison using STABLE-2007. *International Journal of Offender Therapy and Comparative Criminology*, *56*, 856-876. doi:10.1177/0306624X11414693 [Sample 1.3]

*Helmus, L.-M., & Hanson, R. K. (2012, October). *Dynamic risk assessment using STABLE-2007: Updated follow-up and new findings from the dynamic supervision project*. Paper presented at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, *Denver, CO*. [Sample 1.2]

Helmus, L. M., Hanson, R. K., Babchishin, K. M., & Mann, R. E. (2013). Attitudes supportive of sexual offending predict recidivism: A meta-analysis. *Trauma, Violence, & Abuse*, *14*, 34-53. doi:10.1177/1524838012462244

*Helmus, L.-M., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2015). Sex offender risk assessment with the Risk Matrix 2000: Validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression*, *21*, 136-122. doi:10.1080/13552600.2013.870241 [Sample 1.4]

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 & Behavior*, *39*, 1148-1171. doi:10.1177/0093854812443648

Helmus, L.-M., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2011). *Assessing the risk of older individuals convicted of sexual offences: Developing the Static-99R and Static-2002R* (Corrections User Report 2011-01). Ottawa, Ontario: Public Safety Canada

Helmus, L.-M., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older individuals convicted of sexual offences: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 64-101. doi:10.1177/1079063211409951

Hermann, C. A., McPhail, I. V., Helmus, L.-M., & Hanson, R. K. (2017). Emotional congruence with children is associated with sexual deviancy in sexual offenders against children. *International Journal of Offender Therapy and Comparative Criminology*, *61*, 1311-1334. doi:10.1177/0306624X15620830

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557-560. doi:10.1136/bmj.327.7414.557

Howard, P. D. (2017). The effect of sample heterogeneity and risk categorization on area under the curve predictive validity metrics. *Criminal Justice and Behavior*, *44*, 103-120. doi:0093854816678899

Hsu, K. J., & Bailey, J. M. (2017). Autopedophilia: Erotic-target identity inversions in men sexually attracted to children. *Psychological Science*, *28*, 115-123. doi:10.1177/0956797616677082

Humphreys, L. G., & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *The Journal of Applied Psychology*, *76*, 316-321. doi:10.1037/0021-9010.76.2.316

Kafka, M. P. (2010). Hypersexual disorder: A proposed diagnosis for DSM-V. *Archives of Sexual Behavior*, *39*, 377-400. doi:10.1007/s10508-009-9574-7

Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2018). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse: A Journal of Research and Treatment*. Advance online publication. doi:10.1177/1079063218800474

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupler, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, *54*, 337-343. doi:10.1001/archpsyc.1997.01830160065009

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174. Retrieved from http://www.jstor.org/stable/2529310

Laub, J. H., & Sampson, R. J. (2001). Understanding desistance from crime. *Crime and Justice*, *28*, 1-69. doi:0192-3234/2001/0028-0001

Lehmann, R. J., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K. P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment*, *25*, 1019-1024. doi:10.1037/a0033098

*Looman, J., & Abracen, J. (2012, October). *Long-term follow-up of two groups of sex offenders*. Presentation at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, *Denver, CO* [Sample 10]

*Looman, J., & Goldstein, J. (2015, October). *Incremental validity of the STABLE-2007 with the Static-99R*. Presentation at the 34th Annual Research and Treatment Conference for the Treatment of Sexual Abusers, *Montreal, QC* [Sample 3 and 11]

*Lussier, P., Deslauriers-Varin, N., & Râtel, T. (2010). A descriptive profile of high-risk sex offenders under in intensive supervision in the province of British Columbia, Canada. *International Journal of Offender Therapy and Comparative Criminology*, *54*, 71-91. doi:10.1177/0306624X0832323 [Sample 2.1]

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, *22*, 191-217. doi:10.1177/1079063210366039

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

McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012). The sex offender treatment intervention and progress scale (SOTIPS): Psychometric properties and incremental predictive validity with static-99R. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 431-458. doi:10.1177/1079063211432475

McNaughton-Nicholls, C., Callanan, M., Legard, R., Tomaszewski, W., Purdon, S., & Webster, S. (2010). *Examining implementation of the STABLE and ACUTE dynamic risk assessment tool pilot in England and Wales*. Unpublished report. Retrieved from https://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/sexual-offenders-risk.pdf

McPhail, I. V., Hermann, C. A., & Nunes, K. L. (2013). Emotional congruence with children and sexual offending against children: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *81*, 737-749. doi:10.1037/a0033248

Mews, A., Di Bella, L., & Purver, M. (2017). *Impact evaluation of the prison-based Core Sex Offender Treatment Programme*. London, England: Ministry of Justice. Retrieved from www.gov.uk/government/publications

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*, 1406-1421. doi:10.1177/0093854814548449

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*, 941-955. doi:10.1037/pas0000538

Phenix, A., Helmus, L.-M., & Hanson, R. K. (2016). *Static-99R & Static-2002R evaluators' workbook*. Unpublished report, Public Safety Canada, Ottawa, Ontario

*Rettenberger, M., Matthes, A., Schilling, F., & Eher, R. (2011). Die validität dynamisch-veränderbarer risikofaktoren bei der vorhersage einschlägiger rückfälle pädoseueller straftäter [The validity of dynamic and changeable risk factors for the prediction of sexual recidivism in a sample of pedosexual offenders]. *Forensische Psychiatrie, Psychologie, Kriminologie*, *5*, 45-53. doi:10.1007/s11757-010-0086-z [Sample 7.2]

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29*, 615-620. doi:10.1007/s10979-005-6832-7

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*, 19-30. doi:10.1037/1082-989X.13.1.19

Ryan, J., Wilson, N., Kilgour, G., & Reynolds, N. (2014). Balancing efficiency and accuracy: Guidelines for the psychological assessment of extended supervision orders for child sex offenders in New Zealand. *Psychiatry, Psychology and Law*, *21*, 139-157. doi:10.1080/13218719.2013.793154

*Saum, S. (2004). *A comparison of an actuarial risk prediction measure (Static-99) and a stable dynamic prediction measure (Stable-2000) in making risk predictions for a group of sexual offenders* (Unpublished doctoral dissertation). The Fielding Institute, Santa Barbara, CA [Sample 5]

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie / Journal of Psychology*, *215*, 90-103. doi:10.1027/0044-3409.215.2.90

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*, 416-429. doi:10.1037/lhb0000128

*Smeth, A. (2013). *Evaluating risk assessments among sex offenders: A comparative analysis of static and dynamic factors* (Unpublished master's thesis). Ottawa, Ontario, Canada: Carleton University. Retrieved from https://curve.carleton.ca/system/files/theses/27517_1.pdf [Sample 4]

*Sowden, J. N. (2013). *Examining the relationship of risk, treatment readiness, and therapeutic change to recidivism in a sample of treated sex offenders* (Unpublished doctoral dissertation). University of Saskatchewan, Saskatoon, Canada. Retrieved from http://ecommons.usask.ca/bitstream/handle/10388/ETD-2013-11-1305/SOWDEN-DISSERTATION.pdf?sequence=4 [Sample 12.1]

*Sowden, J. N., & Olver, M. E. (2017). Use of the Violence Risk Scale-Sexual Offender version and the STABLE-2007 to assess dynamic sexual violence risk in a sample of treated sexual offenders. *Psychological Assessment*, *29*, 293-303. doi:10.1037/pas0000345 [Sample 12.2]

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology*, *21*, 128-138. doi:101097/EDE0b013e3181c30fb2

Stockdale, K. C., Olver, M. E., & Wong, S. C. P. (2014). The validity and reliability of the Violence Risk Scale–Youth version in a diverse sample of violent young offenders. *Criminal Justice and Behavior*, *41*, 114-138. doi:10.1177/0093854813496999

Thornton, D. (2010). *Scoring guide for Risk Matrix 2000.10/SVC*. Unpublished document.

Thornton, D., & Knight, R. A. (2015). Construction and validation of SRA-FV need assessment. *Sexual Abuse: A Journal of Research and Treatment*, *27*, 360-375. doi:10.1177/1079063213511120

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management* (pp. 225-235). New York: New York Academy of Sciences

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, *30*, 179-191. doi:10.1037/pas0000454

Walker, M., & O'Rourke, M. (2013). Probation officers' experience of using Risk Matrix 2000 and Stable & Acute 2007 when supervising sex offenders living in the community. *Irish Probation Journal*, *10*, 162-176. Available from http://www.probation.ie

Watson, T., & Vess, J. (2007). Risk assessment of child victim sex offenders on extended supervision in New Zealand. *Journal of Forensic Psychiatry and Psychology*, *18*, 235-247. doi:10.1080/14789940601116267

*Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 449-465. doi:10.1007/s11194-007-9063-2 [Sample 6]

Wheeler, J. G., George, W. H., & Stephens, K. (2005). Assessment of sexual offenders: A model for integrating dynamic risk assessment and relapse prevention approaches. In D. M. Donavan & G. A. Marlatt (Eds.), *Assessment of addictive behaviors* (2nd ed., pp. 392-424). New York, NY: Guilford Press

Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The violence risk scale: Sexual offender version* (VRS-SO). Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and University of Saskatchewan