

# STATIC-99 and RRASOR Predict Recidivism among Developmentally Delayed Sexual Offenders: A Cumulative Meta-Analysis

R. Karl Hanson, Chelsea L. Sheahan, Heather VanZuylen  
Public Safety Canada

[Sexual Offender Treatment, Volume 8 (2013), Issue 1]

## Abstract

*This cumulative meta-analysis examined the predictive validity of actuarial risk measures (RRASOR, Static-99, Static-99R) with developmentally delayed sexual offenders. In Study 1, a meta-analytic average was calculated from four studies using the RRASOR or Static-99. Based on a fixed-effect model, both measures were significantly related to the risk of sexual recidivism. Study 2 examined five actuarial risk measures (RRASOR, Static-99, Static-99R, Static-2002, Static-2002R) with 52 developmentally delayed sex offenders, finding good predictive accuracy for all measures ( $0.80 < d < 1.15$ ). When the effect sizes from all previous findings were combined in Study 3, the average effect size for the RRASOR was moderate ( $d = 0.56$ , 95% CI of 0.08 to 1.04,  $k = 4$ ,  $N = 280$ ) and large for the Static-99R ( $d = 1.04$ , 95% CI of 0.39 to 1.69,  $k = 2$ ,  $N = 66$ ) and Static-99 ( $d = 0.77$ , 95% CI of 0.45 to 1.09,  $k = 4$ ,  $N = 160$ ). Given the consistency of the current results with the findings with non-DD sexual offenders, we recommend the use of the Static-99/R and Static-2002/R with developmentally delayed sexual offenders.*

*Key words: sexual offender, developmentally delayed, risk assessment, static risk assessment, recidivism*

Interventions for sexual offenders are most likely to be effective when they follow the principles that guide effective interventions with general offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009). One of these principles, the Risk Principle, states that the intensity of interventions should be matched to the risk level of the offender, with high risk offenders receiving larger amounts of service than those who pose little or no risk to the community (Andrews & Bonta, 2010; Lovins, Lowenkamp, & Latessa, 2009). Application of the risk principle requires valid risk assessments tools.

Although there have been considerable advances in sexual offender risk assessment (Hanson, 2005; Hanson & Morton-Bourgon, 2009), there has been little research that has specifically examined the risk of developmentally delayed sexual offenders. The most commonly used sexual offender risk assessment tools, such as the Rapid Risk Assessment for Sexual Offence Recidivism (RRASOR; Hanson, 1997), Static-99/R (Hanson & Thornton, 2000; Helmus, Thornton, Hanson, & Babchishin, 2012), and Static-2002/R (Hanson & Thornton, 2003; Helmus, Thornton et al., 2012), have been validated on samples containing few (or no) individuals with significant developmental delays.

A developmental delay (DD) of the cognitive type is evident when individuals, since childhood, have had pervasive deficits in cognitive development that substantially impaired their adaptive function in multiple areas (e.g., education, communication, self-care, work, leisure). Formal diagnoses of DD

are associated with both functional impairment and IQ less than some threshold (e.g., < 70, American Psychiatric Association, 2000; World Health Organization, 2010).

Although many of the same psychological risk factors should apply to sexual offenders with or without developmental delays, the validity of any specific risk tool to DD offenders cannot be assumed. For example, the Static-99R item "ever lived 2 years with an intimate partner" may not be relevant because the rate of co-habitation is much lower among DD populations compared to other offender groups (Wilcox, Beech, Markall, & Blacker, 2009). It is also possible that the unique ways in which DD offenders are processed by the criminal justice system could change the diagnostic (and prognostic) significance of their officially recorded offence histories. In addition, there may be unique and important risk factors applicable to the DD populations that are not measured by risk scales developed on non-DD populations (Craig & Hutchinson, 2005; Lindsay, 2002). For example, the ARMIDILO-S, a specialized risk assessment tool for DD sexual offenders (Boer et al., 2012), considers the qualities of the caregivers and community supports, factors that may have particular significance for individuals with developmental delays.

## Actuarial Risk Assessment with Developmentally Delayed Offenders

There are a small number of recidivism studies that examined the RRASOR and Static-99 with DD sexual offenders. The first study to examine the predictive validity of the RRASOR and Static-99 was an unpublished doctoral dissertation by Tough (2001), involving 76 DD adult male sex offenders. Tough found that the RRASOR was more accurate than the Static-99, a finding that influenced subsequent research and practice in the area (see Boer, Tough, & Haaven, 2004). However, in a next available study, the RRASOR did not significantly distinguish between sexual recidivists and non-recidivists ( $AUC = .58$ ,  $n = 103$ ; McGrath, Livingston, & Falk, 2007). A subsequent study of 27 DD sexual offenders found that the Static-99 did better than the RRASOR, although neither was statistically significant (Wilcox et al., 2009). A recent study by Lofthouse and colleagues (in press) found good predictive accuracy for the Static-99 ( $AUC = 0.75$ ,  $p < .001$ ;  $n = 64$ ).

Overall, the research on the validity of the RRASOR and Static-99 with developmentally delayed offenders has been mixed. Some studies have supported the RRASOR, others supported Static-99, and others found negligible differences between them. Such fluctuations are to be expected given the small sample sizes. However, such fluctuations make it difficult for evaluators to decide which, if any, of these measures to use with DD sexual offenders. Consequently, there is a need for a systemic review of the research on the use of these actuarial risk tools with this population.

The following article is therefore divided into three parts. Study 1 is a meta-analysis of existing research on the ability of RRASOR and Static-99 to discriminate between DD offenders who become sexual recidivists and those who do not. In Study 2, we present new data on the predictive accuracy of five risk assessment measures (i.e., RRASOR, Static-99, Static-99R, Static-2002, and Static-2002R) with a sample of developmentally delayed offenders. Study 3 presents an overall summary of the predictive accuracy of these measures for DD sexual offenders in a cumulative meta-analysis.

### Study One

Study 1 is a meta-analysis of sexual recidivism prediction studies using RRASOR, Static-99, and Static-99R with DD sexual offenders. These measures were selected because they are the most

commonly used with routine samples of non-DD sexual offenders. We also looked for Static-2002/R studies, but were unable to identify any previous research using Static-2002/R with DD samples. We summarized the predictive validity using Cohen's  $d$ , an effect size measure of the extent to which the scores of recidivists are different from the scores of non-recidivists (a measure of discrimination). Overall, we were able to identify 6 relevant studies (RRASOR  $k = 3$ ; Static-99  $k = 2$ ; Static-99R  $k = 1$ ).

## Method

**Inclusion criteria.** Studies were eligible for inclusion if they examined adult male sex offenders who were identified as developmentally delayed (DD). The study also must have included scores for a sample of sexual recidivists and non-recidivists regarding one of the following actuarial risk scales: RRASOR, Static-99, Static-99R, Static 2002, and/or the Static-2002R. However, no Static-2002/R studies were identified. The risk assessment must have been completed on information available at time of release, and the coders must have been blind to recidivism status.

**Search strategy.** A search of relevant electronic databases *Criminal Justice Abstracts*, *Google Scholar*, *ProQuest Digital Dissertations*, *PsycINFO*, *Web of Science*, and *Social Sciences Full Text*, for relevant empirical articles was conducted using the following search terms: sex offend\*, child abuse\*, developmentally delayed, intellectually delayed, intellectually disabled, mental retard\*, recidivist, and recidivism. Reference lists of relevant papers, books, and theses were also examined for additional studies. First authors were contacted for data that was not reported in published articles when necessary.

**Studies included.** We identified six independent samples meeting the inclusion criteria (see Table 1). All studies examined adult male sex offenders who were diagnosed as developmentally delayed. Four of the studies were published in peer-reviewed journals (or in press). For two studies, effect sizes were coded from unpublished raw data provided by the author. Specifically, we used an updated dataset by Harris and Tough (2004a, 2004b), which included a longer follow-up time, replacing the findings by Tough (2001). As well, Static-99R data on the subgroup of DD offenders from McGrath, Lasher, and Cumming (2012a, 2012b) was provided by the authors.

**Table 1: Description of Studies Included in Meta-Analysis**

Study	Source	N	Age	Mean IQ	Country	RRASOR Mean (SD)	Static-99 Mean (SD)	Static-99R Mean (SD)
Harris & Tough (2004a)	Unpublished raw data	81	43.8	-	Canada	2.3 (0.88)	4.3 (1.4)	
McGrath et al. (2007)	Published	103	34.6	61.8	U.S.	Not reported		
Wilcox et al. (2009)	Published	27	-	-	U.K.		4.2 (1.9)	
Blacker et al. (2010)	Published	44	35	69.0a	U.K.	2.0 (1.2)		
McGrath et al.	Unpublished	14	33.7	-	U.S.			3.4 (2.3)

(2012a)	raw data						
Lofthouse et al. (in press)	Published	64	32	67.0	U.K.		Not reported

<sup>a</sup> Based on a subsample (n = 14).

Samples were included if the DD diagnoses used in the study was generally consistent with the criteria for mental retardation found in the major diagnostic systems (American Psychiatric Association, 2000; World Health Organization, 2010). Blacker, Beech, Wilcox, and Boer (2010) and Wilcox et al. (2009) identified DD based on intellectual difficulties, significant literacy deficits, and IQ less than 80. In most cases, IQ estimates were provided by the Shipley Institute of Living Scale (SILS, Zachary, 1986), and, occasionally, by WAIS-R scores.

The participants in Harris and Tough (2004) and McGrath et al. (2007) had been diagnosed with mental retardation based on the Diagnostic and Statistical Manual-IV-Text Revision criteria (DSM-IV-TR; American Psychiatric Association, 2000). The criteria used by McGrath et al. (2012b) were not specified, but presumably would have been the same as in McGrath et al. (2007), i.e., DSM-IV-TR. Lofthouse et al. (in press) diagnosed DD based on significant deficits in adaptive skills, onset before age 18, and IQ less than 70 as measured by the WAIS-R or WAIS-III.

**Overview of analyses.** Cohen's *d* was used as the effect size statistic, and was computed directly from means and standard deviations in 5 studies. When means and standard deviations were unavailable ( $k = 1$ ), Cohen's *d* was calculated using the Area Under the Curve (AUC) from receiver operating characteristic (ROC) analyses (Rice & Harris, 2005). A positive *d* value indicated a higher score on the risk measures for recidivists compared to non-recidivists. According to Cohen (1988), *d* values of .20 are considered a small effect, .50 are a medium effect, and .80 are a large effect.

Fixed-effect and random-effects meta-analysis was used to aggregate the results and examine variability across samples (Borenstein, Hedges, Higgins, & Rothstein, 2009). Fixed-effect analyses assume that each effect is a sampled from the same population, and fixed-effect analyses provides an estimate of whether the amount of between-study variability is more than would be expected by chance (Cochran's *Q*). In random-effects meta-analysis, in contrast, between-study variability is assumed and explicitly included in the error term, yielding wider and often more realistic confidence intervals (Schmidt, Oh, & Hayes, 2009). When assumptions are violated, the fixed-effect model is too liberal and the random-effects model is too conservative (Overton, 1998). The results of the random-effects and fixed-effect models converge as the amount of between study variability decreases. Reader should note that the random-effects estimates can be unstable when the number of studies is small (< 30; Schulze, 2007). Consequently, even though the random-effects model may be the best conceptual fit for certain research questions, the available data may be insufficient to estimate its parameters with precision.

The *I*<sup>2</sup> statistic was used to quantify the amount of variability, defined as  $I^2 = [(Q - df)/Q]*100$ . According to Higgins, Thompson, Deeks, and Altman (2003), *I*<sup>2</sup> values of 25%, 50%, and 75% are low, moderate, and high, respectively.

## Measures

**The Rapid Risk Assessment for Sexual Offence Recidivism (RRASOR; Hanson, 1997).** The RRASOR is an empirically derived actuarial risk assessment tool developed to assess the risk for

sexual offence recidivism in adult male sex offenders. The RRASOR consists of four items: (1) prior sexual offences, (2) any unrelated victims, (3) any male victims, and (4) if the offender is less than 25 years of age. Scores range from 0 to 6, with a higher score indicating greater risk of sexual recidivism.

**Static-99 (Hanson & Thornton, 2000).** The Static-99 is an empirically derived actuarial risk assessment tool developed to predict sexual recidivism in adult male sex offenders. It includes ten items related to criminal history, victim characteristics (e.g., gender, relationship to offender), and demographics (age, relationship history). The total scores (ranging from 0 to 12) can be used to place offenders into one of four risk categories: low risk (0, 1), moderate-low risk (2, 3), moderate-high risk (4, 5), and high risk (6+).

**Static-99R (Helmus, Thornton et al., 2012).** The Static-99R contains identical items to Static-99 with the exception of updated age weights, resulting in a range of -3 to 12. In Static-99R, the low risk category ranges from -3 to 1, instead of (0, 1) in Static-99.

## Results

Table 2 presents the recidivism rates and effect sizes for each study. The average effect size for the RRASOR was moderate ( $d = 0.49$  in both fixed-effect and random-effects analyses,  $k = 3$ ;  $N = 228$ ). However, there was significant variability across the three studies ( $Q = 6.44$ ,  $df = 2$ ,  $p = .04$ ,  $I^2 = 68.9$ ). For the RRASOR, the 95% confidence interval indicated a significant effect (i.e., the confidence interval did not contain zero) in the fixed-effect analysis (0.14, 0.84), but not the random-effects analysis (-0.14, 1.12).

**Table 2: Recidivism and Effect Sizes for Each Study Reviewed**

Study	Follow-up in Years (SD)	Recidivism Rate (%)	n/N	Difference between sexual recidivists and non-recidivists (Cohen's $d$ )		
				RRASOR	Static-99	Static-99R
Harris & Tough (2004a)	7.7 (5.0)	17.3	14/81	1.112	0.573	
McGrath et al. (2007)	5.8 (3.7)	10.7	11/103	0.286		
Wilcox et al. (2009)		29.6	8/27		0.327	
Blacker et al. (2010)	8.4 (1.9)					
- Official recidivism		25.0	11/44	0.142		
- Unofficial recidivism		45.4	20/44	0.071		
McGrath et al. (2012a)	5.0(0.0)	14.3	2/14			0.530
Lofthouse et al. (in press)	6.0	32.8	21/64		0.954	

The average effect for the Static-99 was also moderate ( $d = 0.69$  in both fixed-effect and random-effects meta-analysis; 95% *C.I.* of 0.33 to 1.05 for both;  $k = 3$ ;  $N = 172$ ). Across the three Static-99 studies, the amount of variability was no more than would be expected by chance ( $Q = 1.78$ ,  $df = 2$ ,  $p = .41$ ,  $I^2 = 0.0$ ).

The one study that examined Static-99R also found a moderate effect size ( $d = 0.53$ ). The small sample size ( $N = 14$ ) resulted in large confidence intervals for this estimate (-0.87 to 1.93).

## Discussion

The meta-analysis of existing follow-up studies found that both the RRASOR and Static-99 discriminated relatively well between recidivists and non-recidivists, with a slight advantage for Static-99. These findings contradict the early findings of Tough (2001), who observed a clear superiority of the RRASOR over the Static-99 with DD offenders. There was, however, significant variability in the RRASOR findings, with effect sizes ranging from trivial ( $d = 0.14$ ) to large ( $d = 1.11$ ).

Even though we aggregated all previous studies, the combined sample size was still not large. In general, samples containing at least 100 recidivists and 100 non-recidivists are needed for stable estimates (Vergouwe, Steyerberg, Eijkemans, & Habbema, 2005). Instead, the meta-analyses contained 43 recidivists for the Static-99, 36 for RRASOR, and only 2 for Static-99R. Clearly, more research is needed.

## Study Two

The present study examined the predictive validity of five actuarial risk assessment tools (i.e., RRASOR, Static-99, Static-99R, Static-2002, and Static-2002R) in a new sample of adult male developmentally delayed sex offenders. The offenders in this sample were part of a larger sample ( $N = 1,138$ ) collected for the Dynamic Supervision Project (DSP), a prospective study of sexual offenders on community supervision (Hanson, Harris, Scott, & Helmus, 2007). The original assessments were collected as part of the DSP, with updated recidivism information collected in 2011.

## Method

**Subjects.** Of the original sample, 53 offenders were identified as developmentally delayed as reported by the supervising officer. One offender was excluded from the present study as there was insufficient follow-up time (less than two months). Therefore, a total of 52 offenders were included and used as the sample in the present study.

The supervising officers were not trained in diagnosing mental retardation and were simply asked to record whether a DD diagnosis had already been recorded in the offender's file. Given how records are maintained in community supervision, old diagnoses would not have been recorded with consistency; instead, a record of DD would indicate sexual offenders who were currently considered to have special needs within the correctional service.

All offenders were adults starting a period of community supervision (probation or parole) for a recent sexual offence. A sexual offence was defined as an offence with a sexual motivation involving a non-consenting person or persons unable to provide consent (Category "A" offences in

the Static-99 coding rules; Harris, Phenix, Hanson, & Thornton, 2003). Offenders were excluded if they had been in the community for a period of six months prior to initial assessment, had successfully appealed their conviction, were serving sentences for crimes committed prior to the age of 18, or had only been convicted of sexual offences involving consenting adults (e.g., prostitution).

The offenders came from 10 jurisdictions: nine Canadian provinces/territories (Newfoundland, Prince Edward Island, New Brunswick, Ontario, Saskatchewan, Alberta, British Columbia, and Nunavut) and one American state (Iowa). Static-2002 and Static-2002R information was only available for offenders in Canada ( $n = 34$ ). Of the 52 offenders, the average age was 33 ( $SD = 9.2$ ; range of 19 to 60). Approximately 23% ( $n = 12$ ) self-identified as being of Aboriginal heritage, and 26% ( $n = 13$ ) had at least one overnight admission to a psychiatric facility. Thirty-five percent ( $n = 18$ ) had lived with a lover for two years, and 42% ( $n = 22$ ) had prior charges or convictions for sexual offences.

**Measures.** The measures used were the RRASOR, Static-99, Static-99R, Static-2002 and Static-2002R. Descriptions of the RRASOR, Static-99 and Static-99R are provided in Study 1 above.

**Static-2002 (Hanson & Thornton, 2003).** The Static-2002 was created with the same goals as Static-99, namely to predict sexual and violent recidivism using commonly available criminal history information. In comparison to Static-99, it has more coherent scoring rules, and its 14 items are grouped into meaningful subscales: age at release, persistence of sex offending, sexual deviance, relationship to victims, and general criminality. The total score (ranging from 0 to 14) can be used to place offenders in one of five risk categories: low (0 to 2), low-moderate (3, 4), moderate (5, 6), moderate-high (7, 8), and high (9+).

**Static-2002R (Helmus, Thornton et al., 2012).** The items in the Static-2002R are identical to the Static-2002 with the exception of updated age weights, resulting in the low risk category ranging from -2 to 2, and total scores ranging from -2 to 13.

**Data collection.** Data were collected as part of the routine supervision practices of the participating officers for offenders starting supervision between June, 2001 and September, 2005. Static-99 scores were scored prospectively by the supervising officers. In the original DSP study, the rater reliability for Static-99 was high (intraclass correlation of .91, based on 88 cases of mainly non-DD offenders; see Hanson et al., 2007). For the current study, Static-99R scores were computed from Static-99 scores and age. Static-2002 and Static-2002R were scored retrospectively using Static-99 scores, age, and criminal history records.

The assessment data used in this study were considered administrative records controlled by the specific jurisdictions and did not require the consent of offenders to collect. Formal agreements were developed with the participating jurisdictions to share the data with the researchers at Public Safety Canada (then Solicitor General Canada) for the common purpose of program evaluation.

**Recidivism criteria.** Information concerning new sexual offences was gathered through reviews of state, provincial, and Canadian national criminal history records, as well as from supervising officers and local police jurisdictions. For Canadian offenders, centralized criminal history records maintained by the Royal Canadian Mounted Police (RCMP, Canada's federal police service) were received in August 2005, June 2006, and March 2011. Recidivism information from Iowa was collected in July, 2004 and June, 2005.

Once a recidivism event was identified, we obtained the date on which the offence occurred and a brief description of the offence behavior. Offence information was provided by the supervising officers, provincial or state correctional systems, and through direct contact with the police jurisdictions responsible for the original charges.

The follow-up period was calculated from the date that the first assessment information was collected to the date of the last recidivism information received. For the few cases that did not appear on any official record, the follow-up end date was set one month after the last assessment information was received. The median follow-up time was 7.78 years, and ranged from 15 months to 9.8 years. None of these offenders were known to have died during the follow-up period.

Criminal recidivism was considered to have occurred if the agency reporting the information believed that the offence occurred. Given that criminal history records were the major source of recidivism information, the majority of recidivism events were linked to an officially recorded charge or conviction.

**Overview of analyses.** Cohen's *d* and AUCs were used as the effect size indicators. Cohen's *d* is computed using means and standard deviations, where a positive *d* value indicates higher scores on the measures for recidivists compared to non-recidivists. AUC are computed using ordinal ranks, and indicate the probability that a randomly selected recidivist would have a more deviant score than a randomly selected non-recidivist.

## Results

The overall sexual recidivism rate for the 52 developmentally delayed offenders in the sample was 19.2% (10/52). Of the 10 recidivists, 8 had only contact sexual reoffences, one had only non-contact reoffences, and one offender had new contact and non-contact sexual offences.

The effect sizes for the measures are presented in Table 3. All measures showed moderate to large discrimination between recidivists and non-recidivists, with AUC values ranging from a low of .70 for the RRASOR to a high of .79 (shared by Static-99R, Static-2002, and Static-2002R). The Cohen's *d*'s were all large, ranging from 0.80 for the RRASOR to 1.16 for Static-99R. All effects were statistically significant based on the confidence intervals for the AUCs and Cohen's *d*.

**Table 3: Means and Effect Sizes for Risk Tools with DD Sexual Offenders from the Dynamic Supervision Project**

Measure	Recidivists/Total	Mean	AUC	95% C.I.	Effect size ( <i>d</i> )	95% C.I.
RRASOR	10/52	2.1 (1.2)	.70	.51	.89	0.80 0.09 1.51
Static-99	10/52	4.2 (2.1)	.75	.66	.93	1.10 0.38 1.82
Static-99R	10/52	4.2 (2.3)	.79	.59	.91	1.16 0.44 1.88
Static-2002	8/34	5.8 (2.1)	.79	.63	.95	1.14 0.30 1.98
Static-2002R	8/34	5.4 (2.2)	.79	.63	.95	1.15 0.31 1.99

## Study 3

### Method and Results

Cumulative meta-analyses allow researchers to update effect sizes as new studies become available, and to consider the extent to which new findings are consistent with the previous research (Hanson & Broom, 2005). None of the effect sizes in the DSP study (Study 2) were significantly different from the previous meta-analytic averages reported in Study 1. The fixed-effect  $Q_{\text{change}}$  for the RRASOR was 0.57 ( $p = .45$ ), 0.98 ( $p = .32$ ) for the Static-99, and 0.54 ( $p = 0.46$ ) for Static-99R ( $df = 1$  for these comparisons). Consequently, the DSP were added to the previous findings to create new summaries of the effect sizes of these risk assessment tools for DD sexual offenders.

As seen in Table 4, the new meta-analytic averages indicated that all measures showed moderate to large ability to discriminate between recidivists and non-recidivists. All effect sizes were statistically significant in both fixed-effect and random-effects analyses. The average effect size was medium ( $d = .56$ ) for the RRASOR, and the variability was now no more than would be expected by chance ( $Q = 7.01$ ,  $df = 3$ ,  $p = .072$ ). The effect sizes were large for both Static-99 ( $d = 0.77$ ) and Static-99R ( $d = 1.04$ ), and the variability between studies was low ( $Q < df$ ).

**Table 4: Meta-Analysis of Recidivism Studies using the RRASOR, Static-99, or Static-99R with DD Sexual Offenders**

Measure	Fixed-effect			Random-effects			$N(k)$	$Q$	$I^2$
	$d$	LL	UL	$d$	LL	UL			
RRASOR	0.55	0.24	0.87	0.56	0.08	1.04	280 (4)	7.01	57.2
Static-99	0.77	0.45	1.09	0.77	0.45	1.09	160 (4)	2.76	0.00
Static-99R	1.04	0.39	1.69	1.04	0.39	1.69	66 (2)	0.54	0.00

LL – lower limit of 95% confidence interval; UP – upper limit of 95% confidence interval

Note: All  $Q$  were non-significant at  $p > .13$

## Discussion of Study 2 and Study 3

All the risk tools examined in Study 2 (DSP sample) showed moderate to large effect sizes. When the findings from the DSP sample were combined with previous findings, the overall cumulative meta-analysis indicated a moderate effect for the RRASOR and large effects for Static-99 and Static-99R. We could not compute strong tests of the differences between these measures because the data were collected from different studies. However, with the substantial overlap in the confidence intervals for all the measures, it would be reasonable to conclude that all these measures meaningfully discriminate between recidivists and non-recidivists.

Although the effects were not significantly higher in the DSP samples than in other studies, the sample size was insufficient for strong tests. It is quite possible, for example, that the effect sizes were particularly high in the DSP sample because all the DD offenders were selected due to a recent sexual offence conviction. It should not be surprising, then that the type of criminal history

information collected by these static risk tools has predictive power for offenders who have already entered the criminal justice system. The validity of these tools for predicting unofficial offences has yet to be determined. It is interesting to note that neither Blacker et al. (2010) nor Tough (2001) found that considering unofficial offences increased the predictive accuracy of either the RRASOR or Static-99.

## General Discussion

The purpose of this study was to examine the predictive validity of commonly used static risk assessment measures with developmentally delayed sex offenders. Overall, we found moderate to large effects for all measures (RRASOR, Static-99, Static-99R, Static-2002, and Static-2002R). In contrast to the early, influential study by Tough (2001), our cumulative meta-analysis pointed towards a superiority of Static-99/R over the RRASOR. Although we were unable to test whether this difference was statistically significant, this pattern would be consistent with a previous meta-analysis that found that the RRASOR had less predictive accuracy than either the Static-99R or Static-2002R with regular (non-DD) sexual offenders (Babchishin, Hanson, & Helmus, 2012).

It is worth noting that the change in relative strength of the Static-99 and RRASOR from that observed by Tough (2001) was not only attributable to new studies. Extending the follow-up of her original sample by 3 years resulted in a change in  $d$  value from 0.22 in Tough (2001) to 0.57 in Harris and Tough (2004a). Such fluctuations underscore the importance of replication and meta-analysis. With the small samples sizes typical of research with DD sexual offenders, it is difficult to intuit (without formal analysis) the effects of random sampling error.

One limitation of the current review was that different researchers used different criteria for identifying DD groups. Some researchers started with sexual offenders identified by the criminal justice system then separated out a subgroup of individuals with developmental delays; other researchers started with individuals already in treatment due to their cognitive deficits, and then separated out those with sex offence histories. The extent to which these risk scales work for the full range of individuals with developmental disability remains a question for future research.

Another outstanding question is the extent to which the recidivism rates associated with risk scores apply equally to sexual offenders with and without developmental delays. The current study only examined the extent to which recidivists had higher risk scores than non-recidivists (discrimination). We did not have sufficient data to examine the stability of the absolute recidivism rates associated with each score or risk category (calibration, see Moons, Royston, Vergouwe, Grobbee, & Altman, 2009).

A related question for future research is the extent to which the DD sexual offenders would be expected to have major criminogenic needs that are not addressed by the indicators in these risk tools. To the extent that DD offenders have, on average, more risk-relevant life problems compared to typical sexual offenders, then the recidivism rates associated with Static-99R or Static-2002R scores would underestimate the rates for DD offenders (i.e., poor calibration). Previous research has yet to determine whether DD status is a reliable risk indicator in itself (Lindsay & Beail, 2004). Hanson and Bussière (1998) found a small relationship between low intelligence and increased sexual recidivism ( $r = .09$ , 9 studies, 5,651 offenders, which is equivalent to  $d = .20$ ). Hanson and Bussière (1998), however, did not differentiate between normal variation in intelligence and specific developmental delays. There is a need for further research examining whether DD status is associated with increased risk, and whether it adds incrementally to risk prediction after considering other well-established risk factors.

**Implications for practice.** The RRASOR, Static-99/R and Static-2002/R all showed moderate to large ability to rank order sexual offenders in terms of their likelihood of sexual recidivism. Given the consistency of the current findings with the larger research literature for non-DD sexual offenders (Babchishin et al., 2012; Hanson & Morgon-Bourgon, 2009), we believe that it is reasonable to use the research on non-DD sexual offenders as a guide for selecting static measures for DD sexual offenders. All the measures in the current study showed acceptable discrimination. For evaluators wishing to select only one of these measures, we recommend either the Static-99R or the Static-2002R (Helmus, Thornton et al., 2012; [www.static.org](http://www.static.org)). Evaluators should use caution, however, in using any of these measures to estimate absolute recidivism rates because the risk for DD sexual offenders may be systematically higher or lower than the rates for non-DD samples.

None of the measures examined in the current study should be considered a comprehensive risk assessment. The RRASOR, Static-99/R, and Static-2002R only consider commonly available criminal history and demographic information. Much more risk-relevant information can be obtained by also considering the offenders' thoughts, behavioural patterns, social networks, and living conditions. Evaluators interested in comprehensive risk evaluations should consider other structured measures of offenders' risk and criminogenic needs, such as the ARMIDILO-S (Boer et al., 2012; Lofthouse et al., in press).

## Conclusion

The present study found support for the use of the RRASOR and Static-99/R with developmentally delayed sex offenders. Although past research suggested that the RRASOR was superior to the Static-99 (Tough, 2001), our findings conclude that the Static-99/R is preferable to the RRASOR. There is, however, enough research to justify the use of either tool. Although not included in the cumulative meta-analysis, we also found that the Static-2002/R had good predictive validity with developmentally delayed offenders. As all findings were based on limited sample sizes, future research is needed to validate the use of these tools with DD offenders. We are optimistic, however, that these tools will continue to perform well with DD sexual offenders given the consistency of the current results with the much larger research on non-DD sexual offenders.

## Author Note

The views expressed are those of the authors and not necessarily those of Public Safety Canada. We would like to thank Andrew J.R. Harris, Bill Lindsay, Robert McGrath, and Mike Lasher for access to unpublished data. We would also like to thank Kelly Babchishin and Leslie Helmus for assistance throughout this project.

## References

Studies with an asterisk were included in the meta-analysis.

1. American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
2. Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence: LexisNexus Matthew Bender.
3. Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, 19, 442-461.

4. \*Blacker, J., Beech, A., Wilcox, D. T., & Boer, D. P. (2010). The assessment of dynamic risk and recidivism in a sample of special needs sexual offenders. *Psychology, Crime, & Law*, 17, 75-92.
5. Boer, D. P., Haaven, J. L., Lambrick, F., Lindsay, W. R., McVilly, K., Sakdalan, J., & Frize, M. (2012). ARMIDILO-S Manual: Web Version 1.0. Available at [www.armidilo.net](http://www.armidilo.net).
6. Boer, D.P., Tough, S., & Haaven, J. (2004). Assessment of risk manageability of intellectually disabled sex offenders. *Journal of Applied Research in Intellectual Disabilities*, 17, 275-283.
7. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, United Kingdom: Wiley.
8. Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
9. Craig, L., & Hutchinson, R. B. (2005). Sexual offenders with learning disabilities: risk, recidivism and treatment. *Journal of Sexual Aggression*, 11, 289-304.
10. Hanson, R. K. (1997). The development of a brief actuarial risk scale for sexual offence recidivism (User Report 97-04). Ottawa: Department of the Solicitor General of Canada.
11. Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20(2), 212-217.
12. Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, 36, 865-891.
13. Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: A Journal of Research and Treatment*, 17, 357-373.
14. Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362.
15. \*Hanson, R. K., Harris, A. J. R., Scott, T-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The dynamic supervision project*. Ottawa, ON: Public Safety Canada.
16. 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.
17. Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behaviour*, 24, 119-136.
18. Hanson, R. K., & Thornton, D. (2003). Notes on the development of Static-2002. User Report 2003-01. Ottawa: Department of the Solicitor General of Canada.
19. Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa, ON: Solicitor General Canada.
20. \*Harris, A. J. R., & Tough, S. E. (2004a). [Recidivism follow-up of 81 sexual offenders with developmental delays from York Central Hospital]. Unpublished raw data.
21. Harris, A. J. R., & Tough, S. E. (2004b). Should actuarial risk assessments be used with sex offenders who are intellectually disabled? *Journal of Applied Research in Intellectual Disabilities*, 17, 235-241.
22. Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment*, 24(1), 64-101.
23. Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557-560.
24. Lindsay, W. R., (2002). Research and literature on sex offenders with intellectual and developmental disabilities. *Journal of Intellectual Disability Research*, 46, 74-85.

25. Lindsay, W. R., & Beail, N. (2004). Risk assessment: Actuarial prediction and clinical judgment of offending incidents and behaviour for intellectual disability services. *Journal of Applied Research in Intellectual Disabilities*, 17, 229-234.
26. \*Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., Boer, D. P., & Haaven, J. L. (in press). Prospective dynamic assessment of risk of sexual reoffending in individuals with an intellectual disability and a history of sexual offending behaviour. *Journal of Applied Research in Intellectual Disabilities*.
27. Lovins, B., Lowenkamp, C. T., Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse? *The Prison Journal*, 89, 344-357.
28. \*McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012a). [Static-99R scores of 14 developmentally delayed sexual offenders from Vermont who were either recidivists or non-recidivists]. Unpublished raw data.
29. McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012b). The sex offender treatment intervention and progress scale (SOTIPS): Psychometric properties and incremental predictive validity with Static-99R. *Sexual Abuse*, 24, 431-458.
30. \*McGrath, R. J., Livingston, J. A., & Falk, G. (2007). Community management of sex offenders with intellectual disabilities: Characteristics, services, and outcome of a statewide program. *Intellectual and Developmental Disabilities*, 45, 391-398.
31. Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Prognosis and prognostic research: What, why and how? *British Medical Journal*, 338, 1317-1320.
32. Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354-379.
33. 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.
34. Schmidt, F.L., Oh, I., & Hayes, T.L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62, 97-128.
35. Schulze, R. (2007). Current methods for meta-analysis: Approaches, Issues, and Developments. *Zeitschrift für Psychologie*, 215, 90-103.
36. Tough, S. E. (2001). Validation of two standard risk assessments (RRASOR, 1997; Static-99; 1999) on a sample of adult males who are developmentally disabled with significant cognitive deficits. Unpublished Master's Thesis. University of Toronto, Toronto, Ontario.
37. Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J. C., & Habbema, D. F. (2005). Substantial effective samples sizes are required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology*, 58, 475-483.
38. \*Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. *Journal of Sexual Aggression*, 15, 97-106.
39. World Health Organization. (2010). International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10 Online). Downloaded March 10, 2013, from <http://www.who.int/classifications/icd/en/>
40. Zachary, R.A. (1986). Shipley Institute of Living Scale: Revised Manual. Los Angeles, CA: Western Psychological Services.

## Author address

**R. Karl Hanson, Ph.D.**  
*Corrections Research*

*Public Safety Canada  
340 Laurier Ave., West  
Ottawa, Ontario K1A 0P8  
Canada  
[Karl.Hanson@ps-sp.gc.ca](mailto:Karl.Hanson@ps-sp.gc.ca)*