The Predictive Properties of Dynamic Sex Offender Risk Assessment Instruments: A Meta-Analysis

Article *in* Psychological Assessment · April 2017

DOI: 10.1037/pas0000454

7 authors, including:


Jan Willem Van den Berg
Van der Hoeven Kliniek
11 PUBLICATIONS **8** CITATIONS


Edwin Wever
De Forensische Zorgspecialisten
11 PUBLICATIONS **39** CITATIONS


Erick Janssen
Institute for Family and Sexuality Studies
130 PUBLICATIONS **3,622** CITATIONS


Luk Gijs
University of Leuven
47 PUBLICATIONS **455** CITATIONS

Psychological Assessment

The Predictive Properties of Dynamic Sex Offender Risk Assessment Instruments: A Meta-Analysis

Jan Willem van den Berg, Wineke Smid, Klaartje Schepers, Edwin Wever, Daan van Beek, Erick Janssen, and Luk Gijs

Online First Publication, April 3, 2017. http://dx.doi.org/10.1037/pas0000454

Psychological Assessment

1040-3590/17/$12.00 http://dx.doi.org/10.1037/pas0000454

The Predictive Properties of Dynamic Sex Offender Risk Assessment Instruments: A Meta-Analysis

Jan Willem van den Berg
Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry, Utrecht, the Netherlands, and University of Leuven, Belgium

Wineke Smid
Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry, Utrecht, the Netherlands

Klaartje Schepers
Trubendorffer, Utrecht, the Netherlands

Edwin Wever
Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry, Utrecht, the Netherlands

Daan van Beek
Private practice, Utrecht, the Netherlands

Erick Janssen and Luk Gijs
University of Leuven, Belgium

This meta-analysis is the first to our knowledge to evaluate the predictive properties of dynamic sex offender risk assessment instruments, which are designed to assess factors associated with recidivism that are amenable to change. Based on 52 studies (N = 13,446), we found that dynamic risk assessment instruments have small-to-moderate predictive properties, with Cohen's d ranging between 0.71 for sexual recidivism (41 studies, 22 unique samples, N = 5,699) and 0.43 for violent (including sexual) recidivism (27 studies, 14 unique samples, N = 10,368). Incremental predictive validity of dynamic over static risk assessment instruments was significant but modest; Cox hazard ratios varied between 1.08 for sexual recidivism (19 studies, 13 unique samples, N = 3,747) and 1.05 for any recidivism (11 studies, 8 unique samples, N = 2,511). Cox hazard ratios for the predictive validity of change scores on dynamic risk assessment instruments, controlling for static and initial dynamic scores, varied between 0.91 for sexual recidivism (6 studies, 6 unique samples, n = 1,980) and 0.95 for any recidivism (3 studies, 3 unique samples, n = 1,172). These findings indicate that dynamic risk assessment instruments can, in terms of Andrews and Bonta's (2010) risk and need principles, be a useful tool for improving sex offender treatment. They have the potential to contribute to the selection of appropriate, more individually tailored treatment approaches (focusing on individually relevant criminogenic need factors) and can assist in the evaluation of treatment effects. Considering this, further development of dynamic risk assessment instruments is warranted.

Public Significance Statement
Most research on the prediction of recidivism risk in sex offenders has relied on the use of risk assessment instruments containing predominantly static risk factors, factors that cannot be changed by treatment (e.g., the number of previous offenses). The findings of this meta-analytical study suggest that recidivism risk can also be predicted using "dynamic risk factors"—factors that are amenable to treatment—and support continued research on dynamic risk factors and their application to clinical practice.

Keywords: dynamic risk assessment, risk factors, recidivism, sex offenders, predictive validity, predictive properties

Supplemental materials: http://dx.doi.org/10.1037/pas0000454.supp

Jan Willem van den Berg, Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry, Utrecht, the Netherlands, and Institute for Family and Sexuality Studies, Department of Neurosciences, University of Leuven, Belgium; Wineke Smid, Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry; Klaartje Schepers, Trubendorffer, Utrecht, the Netherlands; Edwin Wever, Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry; Daan van Beek, Private practice, Utrecht, the Netherlands; Erick Janssen and Luk Gijs, Institute for Family and Sexuality Studies, Department of Neurosciences, University of Leuven, Belgium.

We thank the following colleagues who provided raw or unpublished data or who conducted additional analyses not included in other reports: Anthony Beech, Janine Blacker, Jackie Craissati, Karl Hanson, Stephen Hart, Maaike Helmus, Michael Lasher, Rachael Lofthouse, Jan Looman, Patrick Lussier, Robert McGrath, David Thornton, and Helen Wakeling. In addition, we thank Kelly Babchishin for her feedback on the statistics used in this meta-analysis.

Correspondence concerning this article should be addressed to Jan Willem van den Berg, Van der Hoeven Kliniek, Center for Clinical Forensic Psychiatry, Location De Voorde, Utrechtseweg 272a, 3818 EV Amersfoort, the Netherlands. E-mail: jwvandenberg@hoevenkliniek.nl

Sexual violence is a major public health concern. Many victims of sexual violence suffer long-term negative consequences, including depressive disorders, trauma and stressor-related disorders, sleep–wake disorders and sexual and relationship problems (Martin, Macy, & Young, 2011). In addition to the psychological consequences for victims, as well as for their family and significant others, there often are additional effects and needs, many of which pose a significant financial burden to society. For example, beyond the costs associated with psychological and medical treatment of victims, significant financial resources are required for the criminal investigation, detention, treatment, and supervision of sex offenders (Elliott & Beech, 2013).

One traditional way to reduce sexual recidivism is providing treatment to apprehended sex offenders. Although the effectiveness of this approach can be considered modest at best, in particular cognitive–behavior therapy (CBT) has been found to reduce recidivism in sex offenders (Andrews & Bonta, 2010; Hanson, Bourgon, Helmus, & Hodgson, 2009; Schmucker & Lösel, 2015). In addition, application of the Risk-Need-Responsivity model (RNR; Andrews & Bonta, 2010) is associated with greater reductions in recidivism in offender treatment. This has been found in a wide range of correctional interventions (Andrews & Bonta, 2010; French & Gendreau, 2006; Landenberger & Lipsey, 2005; Lowenkamp, Latessa, & Holsinger, 2006; Marlowe, Festinger, Lee, Dugosh, & Benasutti, 2006; Wilson, Bouffard, & Mackenzie, 2005), including those for sex offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009; Olver, Nicholaichuk, & Wong, 2013; Schmucker & Lösel, 2015; Smid, Kamphuis, Wever, & Van Beek, 2014).

The RNR model includes three general principles, also known as the "What Works" principles, for the optimization of offender treatment. The Risk Principle dictates that the involvedness of treatment services must be proportional to the offender's risk of reoffending; that is, high-risk offenders should receive the most intensive treatments. Second, the Need Principle emphasizes that treatments should focus on dynamic risk factors that are related to the risk of reoffending—the offender's "criminogenic needs." For instance, sexual preoccupation could be described as a criminogenic need and is empirically related to the risk of sexual reoffending. In contrast, sexual knowledge and social skills could be considered non-criminogenic needs. They can involve needs the offender may lack but that are not directly related to sexual recidivism (Hanson & Morton-Bourgon, 2005). Finally, the Responsivity Principle dictates the tailoring of treatment programs to the individual learning styles, capabilities, and limitations of offenders.

Risk Assessment

Given the Risk Principle, the effectiveness of policies and interventions aimed at the reduction of recidivism depends on our ability to predict who is most at risk of reoffending. For this purpose, a number of so-called "actuarial" risk assessment instruments have been developed (Dawes, Faust, & Meehl, 1989; Meehl, 1954). These instruments do not rely on clinical judgment but provide a statistical based method for estimating the risk of recidivism. Previous meta-analyses have shown that the use of so-called "actuarials" leads to more accurate prediction of recidivism in both sexual and nonsexual (violent) offenders than assessments that rely on clinical judgment (Andrews & Bonta, 2010; Andrews, Bonta, & Wormith, 2006; Hanson & Morton-Bourgon, 2007, 2009). Until recently, actuarials predominantly focused on static risk factors that are associated with recidivism, such as the number of prior sex offenses and the gender of the victim(s). The advantage of this type of instrument is that they take relatively little effort to score and that they are reliable (Hanson & Morton-Bourgon, 2009), although recent field studies (e.g., Boccaccini et al., 2012; Edens, Penson, Ruchensky, Cox, & Smith, 2016) indicate that inter-rater reliability may be lower in applied than in controlled research settings. A more important concern, however, is that static risk factors are not amenable to change and, therefore, provide little to no guidance for the selection of the most appropriate treatment focus. In response to this limitation of static actuarial risk assessment instruments, researchers have started to develop risk assessment instruments that focus on dynamic risk factors. Dynamic risk assessment instruments evaluate risk factors empirically related to recidivism and believed to be amendable to treatment, such as sexual preoccupation and conflicts in intimate relationships (e.g., Mann, Hanson, & Thornton, 2010). Thus, in addition to predicting recidivism risk, and consistent with the Need Principle, these measures can be used to help develop and evaluate treatment goals (e.g., addressing sexual preoccupation, deviant sexual interests, impulsive tendencies).

Although dynamic risk factors have been studied in sex offenders since the mid-1990s, dynamic risk assessment instruments developed for male sex offenders have, as yet, not been the focus of systematic reviews or meta-analyses. Hanson and Bussière (1998) were the first to publish a meta-analysis on "precursors" of sexual recidivism, which included a mix of static and dynamic variables. Following this, Hanson and Harris (2000a) conducted a retrospective study with the specific goal of identifying dynamic factors that are associated with recidivism. For this purpose, they interviewed community supervision officers and reviewed the files of 208 sexual recidivists and 201 non-recidivists. The identified dynamic risk factors were subsequently tested in prospective research (Hanson, Harris, Scott, & Helmus, 2007). This work forms the basis for most dynamic risk assessment instruments that are currently being used in the evaluation of adult male sex offenders.

Current Study

Recently, the number of studies on the predictive properties of dynamic risk assessment instruments, developed for use in adult male sex offenders, has increased substantially. The current meta-analysis is the first to our knowledge focusing on the predictive properties of such risk assessment instruments and, as a secondary goal, examines whether dynamic risk assessment instruments have predictive validity above and beyond that of static risk assessment instruments. As a step toward elucidating possible causal links between dynamic risk factors and reoffending behavior, our third goal is explore to what degree changes in dynamic risk factors over time are associated with recidivism risk.

Summarized, the current meta-analysis will address the following three questions:

(1) What are the predictive properties of dynamic risk assessment instruments developed for adult male sex offenders?

(2) What is the incremental predictive validity of these dynamic risk assessment instruments above and beyond that of static risk assessment instruments?

(3) What is the predictive validity of change scores on these dynamic risk assessment instruments?

Method

Search Strategy

A literature search was performed on empirical papers, published up to the end of March 2014, that present the findings of studies on the predictive properties of dynamic risk assessment instruments (or risk assessment instruments containing a section measuring dynamic risk factors) developed for adult male sex offenders. The search was conducted in PubMed, PsycINFO, ProQuest, SciVerse, and Web of Science databases using combinations of the following search terms: sex offending behavio(u)r (e.g., child molesting, exhibitionism), predictive validity (e.g., accuracy, area under the curve [AUC]-value), recidivism (e.g., relapse, reoffense), and dynamic risk assessment (e.g., dynamic risk factors, names of dynamic risk assessment instruments). In addition, papers were selected and retrieved using the reference sections of empirical studies and previous reviews, and authors of published dynamic risk assessment instruments were contacted and asked about possible other relevant studies and databases, whether published, submitted for publication, or, at that time, unpublished.

Eligibility

All relevant studies published in English, German, French, or Dutch were included. Study design was not used as a criterion for inclusion or exclusion, although we included this information to test the impact of study design and publication bias by post-hoc statistical tests (moderator analyses). To be included in the analyses, samples needed to consist of adult men with a history of at least one sexual offense, and the risk assessment instrument used should contain only dynamic risk factors or at least include a subscale specifically addressing dynamic risk factors. To be included, studies also needed to include the assessment of the predictive validity of the total score of dynamic risk factors, which could involve the prediction of new charges, convictions, or the revocation of conditional release because of new offending behaviors.

Following the recommendations of Hanson and Morton-Bourgon (2009), recidivism was operationalized in three ways: (1) sexual recidivism (vs. no recidivism or nonsexual recidivism) - this category includes possession of child pornography and other noncontact sexual offenses; (2) violent (including sexual) recidivism (vs. no recidivism or nonviolent recidivism); and (3) any recidivism (vs. no recidivism).

Final Selection of Studies

Out of an initial set of 148 potentially relevant studies, 58 met our inclusion criteria. A total of six studies were excluded because effect sizes were not presented in the original publication and could not be obtained from the authors. The data of the final set of 52 studies were used to address the three main questions of our meta-analysis (see Figure 1 for a flowchart of study selection). Some studies used partly or fully overlapping samples (e.g., because they repeated analyses after extending the follow-up period). For these cases, the data involving the longest follow-up period were included, but only if the sample size did not decrease by more than 30%. If the decrease in sample size was more than 30%, the data for the next and shorter follow-up period were used.

Included risk assessment instruments. The studies that were part of this meta-analysis included a total of 14 dynamic and nine static risk assessment instruments (or subscales). The static instruments were used to assess incremental predictive validity of the scores on the dynamic instruments and/or to evaluate the predictive validity of change scores controlled for both static and initially dynamic scores.

Included dynamic risk assessment instruments contain risk factors within at least one of the following domains (Thornton, 2013): sexual interests (e.g., sexualized coping, multiple paraphilias, sexual preoccupation, sexual interest in children), distorted attitudes (e.g., pro-offending attitudes, supportive attitudes towards child sexual abuse and/or rape), relational style (e.g., emotional congruence with children, lack of sustained marital type relations, callousness), or self-management (non-compliance with supervision, violation of conditional release, impulsivity/recklessness, poor coping). A complete listing of all static and dynamic instruments used in the included studies can be found in Appendix B, which is available in the online supplemental materials.

Overall predictive properties. To evaluate the overall predictive properties of dynamic risk assessment instruments (Research Question 1), 46 papers were eligible for inclusion. Twenty-nine of these papers consisted of studies with (partly) overlapping samples. These overlapping samples were selected as described previously, resulting in the final inclusion of 27 unique samples in this component of the meta-analysis (see Appendix B1, available in the online supplemental materials).

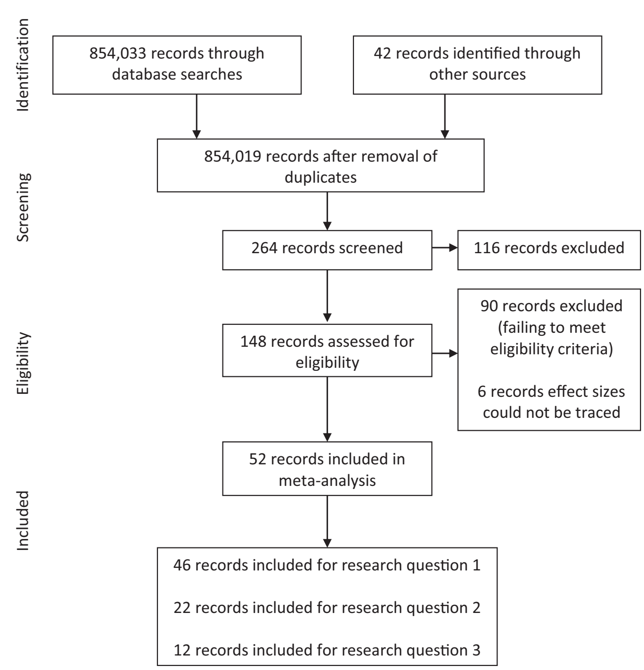

Figure 1. Selection steps for meta-analysis.

The studies on which these 27 samples were based were conducted in Australia (1), Austria (2), Canada (12), New Zealand (1), the United Kingdom (8), and the United States (3), and were published between 2000 and 2014 (mode 2013, median 2010/2011). The vast majority of the papers (44) was published in English. The two remaining papers were published in German. The effect sizes of these studies were taken directly from the publications for 18 samples. For the remaining nine samples, effect sizes were obtained from the authors, who either conducted additional analyses or provided us with the raw data, allowing us to compute effect sizes.

Most samples (21) included sex offenders who had offended against children, as well as sex offenders who had offended against adults. Four samples included only sex offenders who had offended against children, and for two samples their composition in terms of offender type was unknown. In 11 of the samples, all offenders received some form of treatment; 10 samples included only untreated offenders; five samples included both treated and untreated offenders, and for one sample, the treatment status was unknown.

Incremental validity. For the evaluation of incremental validity of dynamic risk assessment instruments (Research Question 2), 22 papers were eligible for inclusion. Ten of these papers consisted of studies with (partly) overlapping samples. These overlapping samples were selected as described previously, resulting in the final inclusion of 16 unique samples in this component of the meta-analysis (see Appendix B2, available in the online supplemental materials).

The studies on which these 16 samples were based were conducted in Australia (1), Austria (1), Canada (7), New Zealand (1), the United Kingdom (4), and the United States (2), and were published between 2004 and 2014 (mode 2012/2013, median 2011/2012). The majority of these papers (18) were published in English. Four were published in German. The effect sizes of these studies were taken directly from the publications for seven samples. For the remaining nine samples, effect sizes were obtained from the authors, who either conducted additional analyses or provided us with the raw data, allowing us to compute effect sizes.

Most samples (13) included sex offenders who had offended against children, as well as sex offenders who had offended against adults. Two samples included only sex offenders who had offended against children, and for one samples the composition in terms of offender type was unknown. In six of the samples, all offenders received some form of treatment; six samples included only untreated offenders; three samples included both treated and untreated offenders, and for one sample the treatment status was unknown.

Predictive validity of change scores. Beggs and Grace (2011) recommended that research into the predictive validity of change scores should ideally include statistically controlling for static and initial, pre-treatment, dynamic risk score to rule out the effect of preexisting risk levels. In other words, controlling for the fact that it is more likely to find (greater) positive changes in high-risk (HR) offenders and (greater) negative changes in low-risk (LR) offenders (regression toward the mean). For that reason, we included two analyses of change scores, one with regular change scores and one with change scores controlled for static and initial dynamic risk scores.

For the evaluation of the predictive validity of change scores (Research Question 3), 12 papers were eligible for inclusion. Six of these papers consisted of studies with overlapping samples. These overlapping samples were selected as described previously, resulting in the final inclusion of nine unique samples in this component of the meta-analysis (see Appendices B3 and B4, available in the online supplemental materials).

The studies on which these nine samples were based were conducted in Canada (5), New Zealand (1), the United Kingdom (3), and were published between 2007 and 2014 (mode 2013, median 2012). All of these papers were published in English. The effect sizes of these studies were taken directly from the publications for six samples. For the remaining three samples, effect sizes were obtained from the authors, who either conducted additional analyses or provided us with the raw data, allowing us to compute effect sizes.

Seven samples included sex offenders who had offended against children, as well as sex offenders who had offended against adults. Two samples included only sex offenders who had offended against children. In seven of the samples, all offenders received some form of treatment; one sample included only untreated offenders; one sample included both treated and untreated offenders.

Inter-rater Reliability

The first and third author independently examined all 148 studies for eligibility and agreed on 98.6% of them (146 of the 148 studies). After a discussion, the other two studies were included in consensus.

The first and second author independently examined 11 randomly chosen studies out of the 52 included studies to assess the reported effect sizes most suitable for inclusion in the three components of the meta-analysis. While the included studies presented a number of effect sizes and reported these in various ways (e.g., Cohen's *d*, AUC, etc.) the authors coded both the preferred unit of measurement for the meta-analysis and the effect size itself. Rater one coded 25, and rater two coded 29 effect sizes. There was a perfect agreement on the preferred unit of measurement (100%). There was agreement on the choice of 42 of the 54 actual effect sizes (77.8%). Most differences involved simple mistakes and omissions.

Statistical Analyses

Calculation of effect sizes. For studies reporting the predictive properties of several dynamic risk assessment instruments within the same sample, the weighted mean of the effect sizes of all included instruments was computed. For studies reporting different effect sizes for the incremental validity of one dynamic risk assessment instrument over several static instruments within the same sample, the weighted mean of the effect sizes on that particular dynamic risk assessment instrument was computed. For studies reporting the predictive validity of both pre- and post-treatment scores, the weighted mean of the effect size of both assessments was computed. Transformations of the effect sizes of individual studies and the meta-analytic integration was conducted using Comprehensive Meta-Analysis (CMA; Version 2; Borenstein, Hedges, Higgins, & Rothstein, 2005).

Effect size metric outcome. For Research Questions 1 and 3[a], regarding the predictive properties of dynamic risk assessment instruments and the predictive validity of regular change scores, we used the standardized mean difference, Cohen's *d*, as effect size metric outcome. It was computed as follows (Hasselblad & Hedges, 1995):

$$d = \frac{(M_1 - M_2)}{SD_{pooled}}$$

where M_1 is the mean score of the group recidivist, M_2 is the mean score of the group non-recidivists, and SD_{pooled} is the pooled-within *SD* of both groups, computed as follows:

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

If not all information necessary for these calculations was available, Cohen's *d* was calculated using the receiver operating characteristic area (AUC), correlation coefficient, or the LogOddsRatio (see Appendix A in the online supplemental materials for the formulas used to convert these effect sizes into Cohen's *d*).

For Research Question 2 and 3[b], regarding the incremental validity of scores on dynamic risk assessment instruments over scores on static risk assessment instruments and the predictive validity of change scores corrected for static and initial dynamic scores, we used the Cox estimate of the hazard ratio as effect size measure, and included static risk scores and for Research Question 3[b] also initial dynamic scores as a covariate. The hazard ratio reflects the chance of recidivism occurring in the group of offenders scoring low on dynamic risk scores divided by the chance of recidivism occurring in the group of offenders scoring high on dynamic risk scores, controlling for their static and initial dynamic (in case of Research Question 3[b]) scores. Thus, it represents the difference in recidivism of presumed LR versus HR sex offenders, above and beyond the effect of their static risk level for Research Question 2, and both static risk level and initial dynamic risk level for Research Question 3[b], independent of follow up time. Because the distribution of log hazard ratio is nearly normal, the log transformation is applied for the meta-analytic integration. The formula for hazard rate is:

$$HR_{CM} = \frac{H_{LR}}{H_{HR}} = \frac{O_{LR}/E_{LR}}{O_{HR}/E_{HR}}$$

where O_i is the observed number of recidivists in group *i*, E_i is the expected number of non-recidivists in group *i*, and H_i is the overall hazard rate for the i^{th} group. A confidence interval for hazard ratios is computed by first transforming to the log scale, which more accurately approximates a normal distribution, calculating the limits, and subsequently transforming back to the original scale. The calculation is made using the formula:

$$In(HR_{CM}) \pm z1 - \alpha/2(SE_{InHR_{CM}})$$

where

$$SE_{InHR_{CM}} = \sqrt{1/E_T} + 1/E_C$$

Meta-analytic integration. For the meta-analytic integration of the effect sizes of the included studies, both fixed and random effects model are used. These two meta-analytic models produce identical results when variability across the included studies is low. With increasing variability across studies, random-effect models give relatively more weight to smaller studies, resulting in confidence intervals that exceed those of fixed-effect models (Helmus, Hanson, Babchishin, & Mann, 2013). For this reason, we used fixed-effect models to construct a combined confidence interval when effect sizes were relatively homogeneous across included studies. However, if effect sizes across studies were heterogeneous, a random-effect model was used. Cochran's *Q* was used to test for heterogeneity. In addition, the I^2 statistic was used to describe the variability in effect sizes that could not be explained by chance.

Because the findings of meta-analyses can be strongly influenced by outlying effect sizes and large sample sizes, we removed outliers following Hanson and Bussière (1998) and Hanson and Morton-Bourgon (2009). A study was classified as an outlier if the overall *Q* was significant and the effect size of that particular study was the most extreme value and accounted for more than 50% of the variability (effect sizes will be presented both with and without outliers). Following Helmus, Babchishin, and Hanson (2013), the variance of studies with the largest weights were adjusted so their weights would not exceed the next largest weight by factor 2.

Moderator analyses. Moderator analyses were performed, after excluding studies that were considered outliers, to assess the influence of method variations and sample characteristics. To be qualified as a moderator, a sufficient number of studies with information on that moderator needed to be available (at least three studies in each category for categorical moderators, at least 10 studies for continuous moderators), and they should be meaningfully related to theoretical models or prior research findings. The following categorical moderators were analyzed by means of a fixed-effect test of the $Q_{between}$, measuring the variability accounted for the levels of the moderator; publication (published study/unpublished study), development (instrument developed with study sample/replication in new sample), design (prospective study/retrospective study), blind (scored blind to recidivism outcome/unclear if scored blind to recidivism outcome), location type (samples from institutions/community/mixed samples), setting (corrections in prison or community/treatment in prison or community/treatment in psychiatric setting/mixed group), and offender type (child molesters and child porn/rapists/mixed group). The following continuous moderators were analyzed by means of meta-regression: sample size, length of follow-up, and year of publication. If multiple publications were based on (partly) overlapping samples, the first publication date was used for the moderator analyses.

Appendix B, available in the online supplemental materials, provides a detailed description of the included samples.

Results

Overall Predictive Properties

The average follow-up time for studies predicting sexual recidivism was 67.1 month (range = 18–240), for violent (including sexual) recidivism 53.7 months (range = 36–240) and for any recidivism (sexual, violent or any) 75.5 months (range = 9–240). Observed recidivism rates were 16.9% for sexual recidivism (23 samples, n = 5,877), 11.8% for violent (including sexual) recidivism (14 samples, n = 10,368), and 33.3% for any recidivism (10 samples, n = 3,405). Table 1 summarizes the effect sizes of the predictive properties of dynamic risk assessment instruments de-

Table 1
Effect Sizes for Research Question 1: Predictive Properties of Dynamic Risk Assessment Instruments by Recidivism Type

	Fixed-effect		Random-effect						
Recidivism type	Cohen's *d*	95% CI	Cohen's *d*	95% CI	*Q*	I^2 (%)	*N*	K	L
Sexual	.71	[.63, .79]	.70	[.61, .79]	25.5	17.7	5,699	41	22
Includes outlier (Allan & Dawson, 2002)	.74	[.66, .82]	.76	[.63, .90]	59.9**	63.3	5,877	42	23
Violent	.37	[.30, .44]	.43	[.29, .57]	42.7**	70.0	10,368	27	14
Any	.64	[.56, .72]	.64	[.56, .73]	9.8	7.6	3,405	20	10

Note. K = number of studies; L = number of unique samples; CI = confidence interval.
** $p < .01$.

veloped for adult male sex offenders on the three types of recidivism. Effect sizes of the individual samples are reported in Appendix B1 (available in the online supplemental materials).

Predictive properties for sexual recidivism. After excluding one outlier (Allan & Dawson, 2002), the fixed-effect weighted Cohen's *d* on sexual recidivism was 0.71, 95% confidence interval (CI) [0.63, 0.79], $n = 5{,}699$, based on 22 unique samples. The *Q* indicated that the variability across studies was not significant ($Q = 25.5$, $p = .23$). The I^2 (17.7%) indicated that the amount of variability in effect sizes among studies that could not be attributed to chance was small.

Moderators. The $Q_{between}$ analyses for the fixed-effect model revealed significant effects for two categorical moderators. Instruments developed within the study sample produced larger effect sizes than replications on new samples (Cohen's *d* respectively 0.94 and 0.67; $Q_{between} = 4.2$, $df = 1$, $p = .04$). And instruments scored on combined locations produced larger effect sizes than instruments used either in an institution sample or in a community sample (Cohen's *d* response 0.97, 0.61, 0.65; $Q_{between} = 6.4$, $df = 1$, $p = .04$).

The fixed-effect meta-regressions for the continuous moderators showed that older studies and studies including shorter follow-up periods produced larger effect sizes than more recently published studies and those with longer follow-up periods (respectively, $b_1 = -.02757$, $SE = .00964$, $Z = -2.859$, $p = .004$, $k = 23$; $b_1 = -.00161$, $SE = .00072$, $Z = -2.224$, $p = .026$, $k = 21$). No significant effect was found for sample size ($b_1 = .00002$, $SE = .00024$, $Z = .068$, $p = .946$, $k = 21$).

Predictive properties for violent (including sexual) recidivism. In the overall analysis for violent (including sexual) recidivism (14 unique samples), the fixed-effect weights of the individual studies (the inverse of the variance) varied between 2.21 and 500.00, with a median value of 38.52. The studies producing the largest weights, 250.00 and 500.00, had sample sizes of respectively 3,402 and 3,360 (recidivism rates were 7.8% and 6.9% respectively). These weights were more than twice as large as the next largest study weight (90.91) and more than 113 and 226 times larger than the smallest study weight, respectively. To reduce the influence of these studies, the variances were artificially increased so that their weights were at most twice as large as the next largest study weight. After adjusting the weights, we found significant variability among study effect sizes ($Q = 42.7$, $p < .01$; $I^2 = 70.0$). Cohen's *d* for the random-effects model was 0.43, 95% CI [0.29, 0.57], $N = 10{,}368$, based on 14 unique samples.

Moderators. The $Q_{between}$ analyses for the fixed-effect model revealed no significant categorical moderators. The fixed-effect meta-regression for the continuous moderators showed that studies with larger samples produced smaller effect sizes ($b_1 = -.00011$, $SE = .00002$, $Z = -5.937$, $p < .001$, $k = 14$) and that studies including longer follow-up periods produced larger effect sizes ($b_1 = .00243$, $SE = .00065$, $Z = 3.742$, $p < .001$, $k = 12$). No significant effect was found for year of publication ($b_1 = -.02093$, $SE = .01267$, $Z = -1.652$, $p = .099$, $k = 14$).

Predictive properties for any recidivism. The fixed-effect weighted Cohen's *d* on any recidivism was 0.64, 95% CI [0.56, 0.72], $N = 3{,}405$, based on 10 unique samples. The *Q* indicated the variability across studies was not significant ($Q = 9.8$, $p = .37$). The I^2 (7.6%) suggested low variability among studies.

Moderators. As shown in Appendix C3 in the online supplemental materials, the $Q_{between}$ analyses revealed no significant effects for categorical moderators. The fixed-effect meta-regression of the continuous moderators revealed no significant effects either; year of publication ($b_1 = -.00624$, $SE = .01468$, $Z = -0.425$, $p = .671$, $k = 10$), sample size ($b_1 = .00011$, $SE = .00020$, $Z = 0.545$, $p = .586$, $k = 10$), and follow-up time ($b_1 = .00004$, $SE = .00091$, $Z = 0.044$, $p = .965$, $k = 10$).

Incremental Validity

Only studies reporting incremental validity of dynamic risk assessment instruments over static risk assessment instruments were included in this meta-analysis. The average follow-up interval was 80.5 (range = 36–240) months for studies predicting sexual recidivism, 54.4 (range = 36–240) months for studies predicting violent (including sexual) recidivism and 73.6 (range = 9–240) months for studies predicting any recidivism. Observed recidivism rates were 13.0% for sexual recidivism (13 samples, $n = 3{,}747$), 11.0% for violent (including sexual) recidivism (11 samples, $n = 9{,}019$), and 33.6% for any recidivism (8 samples, $n = 2{,}511$). Table 2 presents the effect sizes for the incremental validity of dynamic risk assessment instruments developed for adult male sex offenders for each of the three recidivism types. Effect sizes of individual samples are presented in Appendix B2 in the online supplemental materials.

Incremental validity for sexual recidivism. The fixed-effect weighted hazard ratio on sexual recidivism was 1.08, 95% CI [1.06, 1.10], $N = 3{,}747$, based on 13 unique samples. The *Q* indicated that the variability across studies was not significant ($Q = 18.3$, $p = .11$). The I^2 (34.3%) indicated that the variability in effect sizes among studies that could not be explained by chance was moderate.

Table 2
Effect Sizes for Research Question 2: Incremental Validity of Dynamic over Static Risk Assessment Instruments by Recidivism Type

	Fixed-effect		Random-effect						
Recidivism type	Hazard ratio	95% CI	Hazard ratio	95% CI	Q	I^2 (%)	N	K	L
Sexual	1.08	[1.06, 1.10]	1.09	[1.06, 1.12]	18.3	34.3	3,747	19	13
Violent	1.06	[1.04, 1.08]	1.06	[1.04, 1.08]	11.8	15.0	9,019	15	11
Any	1.04	[1.03, 1.06]	1.05	[1.02, 1.08]	16.3*	57.1	2,511	11	8

Note. K = number of studies; L = number of unique samples; CI = confidence interval.
* $p < .05$.

Moderators. None of the categorical moderators was found to be significant in the $Q_{between}$ analysis. This also applied to the fixed-effect meta-regression of the continuous moderators; year of publication ($b_1 = -.00934$, $SE = .00514$, $Z = -1.815$, $p = .070$, $k = 13$), sample size ($b_1 = -.00004$, $SE = .00006$, $Z = -0.676$, $p = .499$, $k = 13$), and follow-up time ($b_1 = -.00006$, $SE = .00020$, $Z = -0.285$, $p = .776$, $k = 13$). Detailed information on the categorical moderator analysis is provided in Appendix C4 (available in the online supplemental materials).

Incremental validity for violent (including sexual) recidivism. The fixed-effect weighted hazard ratio on violent (including sexual) recidivism was 1.06, 95% CI [1.04, 1.08], N = 9,019, based on 11 unique samples. The Q indicated that the variability across studies was not significant ($Q = 11.8$, $p = .30$). The I^2 (15.0%) indicated that there was low variability among studies.

Moderators. None of the categorical moderators in the $Q_{between}$ analysis and none of the continuous moderators in the fixed-effect meta-regression were significant; year of publication ($b_1 = .00896$, $SE = .00105$, $Z = -0.858$, $p = .391$, $k = 11$), sample size ($b_1 = .00002$, $SE = .00002$, $Z = 1.483$, $p = .138$, $k = 11$), and follow-up time ($b_1 = -.00019$, $SE = .00016$, $Z = -1.190$, $p = .234$, $k = 11$). Details on the categorical moderator analysis are presented in Appendix C5 (available in the online supplemental materials).

Incremental validity for any recidivism. The random-effect weighted hazard ratio on any recidivism was 1.05, 95% CI [1.02, 1.08], N = 2,511, based on 8 unique samples. There was significant variability among study effect sizes ($Q = 16.3$, $p = .02$; I^2 = 57.1%).

Moderators. The $Q_{between}$ analyses for the fixed-effect model revealed no significant effects for the categorical moderators. Given the fact that there were only eight unique samples available, we did not conduct a meta-regression analysis for the three continuous moderators.

Predictive Validity of Change Scores

Only the studies reporting predictive validity of change scores on dynamic risk factors were included in this step of the analyses. The average follow-up interval was 68.6 months (range = 36–140) for studies predicting sexual recidivism, 51.3 months (range = 36–112) for studies predicting violent (including sexual) recidivism and 72.6 months (range = 36–112) for studies predicting any recidivism (sexual, violent, or any). Observed recidivism rates were 9.0% for sexual recidivism (6 samples, n = 2,043), 9.1% for violent (including sexual) recidivism (6 samples, n = 6,234), and 33.7% for any recidivism (4 samples, n = 1,466). Table 3 summarizes the effect sizes of the predictive validity of change scores on dynamic risk assessment instruments developed for adult male sex offenders on the three recidivism types. Effect sizes for the individual samples are presented in the Appendix B3 (available in the online supplemental material).

Predictive validity of change scores for sexual recidivism. In the overall analysis for sexual recidivism (6 unique samples), the fixed-effect weights of the individual studies (the inverse of the variance) varied between 13.16 and 142.86, with a median value of 22.22. The study producing the largest weight involved a sample of 572 men (recidivism rate was 6.2%) and had more than twice the weight of the next largest study (25.64) and more than 10 times the weight of the study with the smallest weight. To reduce the influence of this study, its sample size was artificially decreased to 208, so that its weight was not more than twice the next largest study weight. After adjusting this weight, the fixed-effect weighted d for change scores on dynamic risk assessment instruments in predicting sexual recidivism was 0.26, 95% CI [0.10, 0.42], N = 2,043, based on 9 studies representing 6 unique samples. The Q indicated that the variability across studies was not significant ($Q = 5.8$, $p = .33$). The I^2 (13.8%) indicated that the variability among studies that could not be explained by chance was small.

Table 3
Effect Sizes for Research Question 3: Predictive Validity of Change Scores Derived from Dynamic Risk Assessment Instruments

	Fixed-effect		Random-effect						
Recidivism type	Cohen's d	95% CI	Cohen's d	95% CI	Q	I^2 (%)	N	K	L
Sexual	.26	[.10, .42]	.27	[.10, .44]	5.8	13.8	2,043	9	6
Violent	.14	[.05, .24]	.15	[−.00, .29]	10.54	52.6	6,234	8	6
Any	.10	[−.01, .22]	.10	[−.01, .22]	1.04	<.00	1,466	6	4

Note. K = number of studies; L = number of unique samples; CI = confidence interval.

Moderators. The data provided sufficient information to run a moderator analysis only on location type, which showed no significant effect ($Q_{between}$ = 0.1, *df* = 1, *p* = .71). There was not enough information to run analyses on continuous moderators.

Predictive validity of change scores for violent (including sexual) recidivism. For violent (including sexual) recidivism, the fixed-effect weighted Cohen's *d* was 0.14, 95% CI [0.05, 0.24], *N* = 6,234 based on 6 unique samples. The *Q* indicated that the variability across studies was not significant (*Q* = 10.54, *p* = .06). The I^2 (52.6%) indicated that the amount of variability in effect sizes among studies that could not be explained by chance was moderate.

Moderators. The data provided sufficient information to run a moderator analysis only on location type, which showed no significant effect ($Q_{between}$ = *1.0, df* = 1, *p* = .31). There was not enough information to run any analysis on continuous moderators.

Predictive validity change scores for any recidivism. The fixed-effect weighted Cohen's *d* on any recidivism was 0.10, 95% CI [−0.01, 0.22], *N* = 1,466, based on 4 unique samples. The *Q* indicated that the variability across studies was not significant (*Q* = 1.04, *p* = .79). The I^2 (<0.00%) indicated that there was no variability among studies that could not be explained by chance.

There was not enough information to run any moderator analysis.

Predictive Validity of Change Scores Controlled for Static and Initial Dynamic Scores

For this part of the analyses, only studies that reported the predictive validity of change scores, controlled for both static and initial dynamic scores, were included.

The average follow-up time for the included studies was 87.2 (range = 36–147) months for studies predicting sexual recidivism, 57.4 (range = 36–112) months for studies predicting violent (including sexual) recidivism and 63.9 (range = 36–89) months for studies predicting any recidivism (sexual, violent, or any). Observed recidivism rates were 11.3% for sexual recidivism, (6 samples, *n* = 1.980), 10.5% for violent (including sexual) recidivism (5 samples, *n* = 4,168) and 29.7% for any recidivism (3 samples, *n* = 1,172). Table 4 summarizes the effect sizes of the predictive validity of change scores on dynamic risk assessment instruments corrected for both static and initial dynamic score on the three types of recidivism. Effect sizes of the individual samples are published in the Appendix B4 (available in the online supplemental materials).

Predictive validity of controlled change scores for sexual recidivism. For the analysis of sexual recidivism, the fixed-effect weights of the individual studies (the inverse of the variance) varied between 976.56 and 19.75, with a median value of 257.56. The study producing the largest weight had a sample size of 481 (recidivism rate was 3.95%) and had more than twice the weight of the next largest study (349.41) and 49 times more weight than the smallest study weight. To reduce the influence of these studies, the variance was artificially increased to ensure that these study weights would not exceed the next largest study weight by more than twice its size.

The fixed-effect weighted hazard ratio on sexual recidivism for the corrected change score on the dynamic risk assessment instruments was 0.91, 95% CI [0.87, 0.95], *N* = 1,980 based on 6 unique samples. Thus, changes on the dynamic factors added to the prediction of sexual recidivism above and beyond static and initial dynamic risk scores. The *Q* indicated that the variability across studies was not significant (*Q* = 2.4, *p* = .79). The I^2 (<0.00%) indicated that the variability among studies that could not be explained by chance was low. Not enough information was available to run moderator analyses.

Predictive validity of controlled change scores for violent (including sexual) recidivism. For the analysis of violent (including sexual) recidivism the fixed-effect weights of the individual studies (the inverse of the variance) varied between 1,736.11 and 198.37 with a median value of 625.00. The study producing the largest weight came from an effect with sample size of 414 (recidivism rate of 8.21%) and had more than twice the weight of the next largest study (625.00) and more than 8 times more weight than the smallest study weight. To reduce the influence of this study, its variance was artificially increased so that its study weight did not exceed the next largest study weight by factor 2.

The fixed-effect weighted hazard ratio on violent (including sexual) recidivism for the corrected change score on the dynamic risk assessment instruments was 0.93, 95% CI [0.90, 0.97], *N* = 4,168, based on 5 unique samples. Thus, changes on dynamic risk factors added to the prediction of violent (including sexual) recidivism above and beyond static and initial dynamic scores. The *Q* indicated that the variability across studies was not significant (*Q* = 3.5, *p* = .48). The I^2 (<0.00%) indicated that there was no variability among studies that could not be explained by chance. There was not enough information to run any moderator analysis.

Predictive validity of controlled change scores for any recidivism. For the analysis of any recidivism the fixed-effect weights of the individual studies (the inverse of the variance) varied between 3,460.21 and 452.69 with a median value of 1,111.11. The study producing the largest weight had a sample size of 392 (recidivism rate of 21.2%) and more than twice the weight of the next largest study (1,111) and more than 7 times more weight than the smallest study weight. To reduce the influence of

Table 4
Effect Sizes Research for Question 3: Predictive Validity of Change Scores Corrected for Static and Initial Dynamic Scores

	Fixed-effect		Random-effect						
Recidivism type	Hazard ratio	95% CI	Hazard ratio	95% CI	*Q*	I^2 (%)	*N*	K	L
Sexual	.91	[.87, .95]	.91	[.87, .95]	2.4	<.00	1,980	6	6
Violent	.93	[.90, .97]	.93	[.90, .97]	3.5	<.00	4,168	5	5
Any	.95	[.93, .98]	.95	[.93, .98]	1.5	<.00	1,172	3	3

Note. K = number of studies; L = number of unique samples; CI = confidence interval.

this study, its variance was increased artificially to ensure its weight did not exceed the next largest study weight by factor 2.

The fixed-effect weighted hazard ratio on any recidivism for corrected change scores on dynamic risk assessment instruments was 0.95, 95% CI [0.93, 0.98], N = 1,172, based on 3 studies representing 3 unique samples. As such, changes on the dynamic risk factors added to the prediction of any recidivism above and beyond static and initial dynamic risk scores. The Q indicated that the variability across studies was not significant (Q = 1.5, p = .47). The I^2 (<0.00%) indicated that there was no variability among studies that could not be explained by chance. Not enough information was available to run any moderator analysis.

Discussion

This meta-analysis is the first to examine the predictive properties of dynamic risk assessment instruments designed to assess recidivism risk in adult male sex offenders. We evaluated their value in predicting sexual, violent (including sexual), and any (general) recidivism in terms of overall predictive properties, incremental predictive validity over static risk assessments, and the predictive validity of change scores on these instruments.

Our findings show that dynamic risk assessment instruments indeed significantly contribute to the prediction of sexual, violent, and any recidivism. In line with Hanson and Morton-Bourgon's (2009) observations, effect sizes were largest for the type of outcome that the instruments were primarily designed to assess, in this case sexual recidivism. The effect sizes we found for sexual and violent (including sexual) recidivism (d = 0.71; CI [.63, .79] and d = 0.43; CI [.29, .57], respectively) are similar to the mean effect sizes reported by Hanson and Morton-Bourgon (2009) for predominantly static actuarial instruments (d = 0.67; CI [.63, .72] and d = 0.51; CI [.47, .56], respectively). The effect size for any recidivism we found in this study (d = .64; CI [.56, .72]) is higher than the mean effect size for the actuarial overall static risk assessment instruments reported by Hanson and Morton-Bourgon (d = .52, CI [.48, .56]). The relative small effect size in our meta-analysis for violent (including sexual) recidivism is noteworthy and coincided with great variability in effect sizes among studies predicting violent recidivism and with great variability in the specific (dynamic) instruments used. The latter finding suggests that some dynamic risk assessment instruments may be better at assessing risk for violent sex offending than others. The current study aimed to investigate the value of dynamic risk assessment in general. For those studies that included more than one dynamic risk assessment instrument, the results of the various instruments were averaged. This makes it impossible to compare and evaluate the predictive properties of individual dynamic risk assessment instruments. Future meta-analyses could take this as their focus.

The incremental validity of dynamic over static risk assessment was established for all outcome measures. However, effect sizes tended to be small, which suggests that static and dynamic instruments overlap, at least when it comes to their predictive value. This outcome underpins the question that has been raised by Ward and Beech (2015), whether dynamic risk factors perhaps measure correlates of underlying propensities, like static risk factors do, but in different ways. That is, the possibility should be considered that dynamic risk factors measure clinical features that are associated with the psychological propensities that actually cause recidivism, instead of assessing these propensities themselves.

We also found significant effects for change scores. Statistically correcting for both static and initial dynamic risk scores, change scores significantly predicted all three types of recidivism, indicating that offenders who showed larger positive changes (reflecting a reduction in dynamic risk scores) recidivate at lower rates than those who show a smaller change in dynamic risk factors. Much like our findings regarding overall and incremental predictive validity, effects were the largest for sexual recidivism. However, all effect sizes were relatively small. Thus, only a small part of change in recidivism was explained by changes in dynamic risk factors. This raises questions on the nature of the variables currently included in dynamic risk assessments. If dynamic risk factors are predominantly correlates of psychological propensities, causally related to sexual (re)offending, then treatment of these dynamic risk factors would still, at best, 'treat the symptoms' and not the causes. This might be considered a less effective treatment approach and therefore could provide a basis for the small effect sizes we found for change scores. Nevertheless, positive change on dynamic risk does indicate that the risk level of a particular sex offender is reduced, and this finding warrants a continued focus on the role of change scores in future research.

We could not carry out moderator tests for several of our main analyses because of a lack of a sufficient number of studies in each category. However, of the moderator analyses that were possible, results suggest that the impact of publication bias was negligible. Nevertheless, some other moderator variables did show a significant impact on effect sizes. For the overall predictive properties for sexual recidivism of dynamic risk assessment instruments we found that original (vs. replication) studies, earlier published studies, studies with a shorter average follow-up interval, and studies using both institution and community samples, were associated with larger effect sizes. For violent (including sexual) recidivism, the opposite was found for average follow-up intervals (shorter follow-up intervals were associated with smaller effect sizes), and studies with a larger sample size had smaller effect sizes. We found no significant moderator effects in the analyses focusing on incremental validity, and we were not able to perform moderator analyses for change scores because of a lack of a sufficient number of studies in each category.

The present study is the first meta-analysis to specifically focus on the predictive properties of dynamic risk assessment instruments developed for adult male sex offenders and adds, we believe, valuable new knowledge on the assessment of sex offender recidivism risk. The meta-analysis was based on both published and unpublished studies from various countries and settings that used a range of risk assessment instruments. Nevertheless, some limitations should be acknowledged. For example, our meta-analysis was based on a modest number of unique study samples. Especially the analyses using change scores are based on a relatively small number of studies. Also, the fact that most samples came from studies conducted in Europe and North America limits generalizability of the findings to other, non-western adult sex offender populations. Despite these limitations, we believe the findings of this meta-analysis are promising and hopefully will inspire and encourage researchers to design and conduct further studies on dynamic risk assessment (instruments).

The findings that dynamic risk assessment instruments significantly predict recidivism, add additional predictive value to static risk assessment, and that change scores significantly predict recidivism, supports the conclusion that dynamic risk assessment instruments designed for use in adult sex offenders can be a useful aid in optimizing sex offender treatment in a way that is consistent with Andrews and Bonta's (2010) risk and need principles. The instruments can be used to guide the referral of offenders to appropriate treatment approaches and levels, to help aim the focus of the treatment toward criminogenic need factors, and to evaluate the success of the treatment.

Although significant, the predictive validity of intra-individual change scores was not at the order of magnitude that would allow them to be used alone, independently, for the estimation of the decrease in recidivism risk after treatment (this research field's "Holy Grail," according to Serin, Lloyd, Helmus, Derkzen, & Luong, 2013). However, we believe change scores of individual sex offenders have additional value and should be taken into consideration in decision making on supervision and continuation of judicial measures. Furthermore, it could be argued that the somewhat small incremental validity of dynamic over static risk assessment instruments does not justify the additional efforts and costs of using them simultaneously. However, given the importance of making the correct decisions in treatment referral and treatment goals, and to make any effort to further decrease the number of victims of sexual offending behavior, we believe the (additional) use of dynamic risk assessment instruments is both valuable and justified.

Obviously, our findings and conclusions are based on currently available instruments. Because the development and use of dynamic risk assessment instruments is still a relatively recent phenomenon, further efforts to improve the value and scope of dynamic risk assessment instruments can be expected to change and add to their value and contribution. Uncovering causal pathways between risk factors and sexual offending behavior may provide more insight into the constructs underlying the risk factors assessed by static and dynamic instruments and may lead to the inclusion of more "psychologically meaningful" risk factors (Mann, Hanson, & Thornton, 2010). On a more practical level, the measurement of individual (dynamic) risk factors may be improved upon, because this momentarily relies on self-report and clinical observation judgment. The development of implicit measures to assess, for instance, hostility toward women or sexual preoccupation in sex offenders, may reduce noise and result in or add to a more reliable assessment. Finally, research on the network structure of dynamic risk factors could provide additional insights and facilitates the development of more valid and clinically useful measures. Network analysis (Borsboom & Cramer, 2013) could help find answers to questions such as how dynamic risk factors interrelate and which dynamic risk factors form key factors in the reduction of the risk of recidivism. Also, this kind of analysis can be used to address questions about whether (some) dynamic risk factors tend to change simultaneously and whether change in (some) factors is conditional upon change in certain other factors. The answers to these and related questions can bring us a step closer to a more accurate and effective assessment of sex offender recidivism risk, and to our primary goal of building a safer society.

References

Entries marked with * indicate studies included in one or more of the meta-analyses.

Allan, A., & Dawson, D. (2002). *Developing a unique risk of violence tool for Australian indigenous offenders* (Report No. CRC 6/00–01). Canberra, Australia: Criminology Research Council. Retrieved from http://www.criminologyresearchcouncil.gov.au/reports/200001-06.pdf

*Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: Journal of Research and Treatment, 19,* 347–367.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis, Matthew Bender.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52,* 7–27. http://dx.doi.org/10.1177/0011128705281756

*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

*Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2012). How useful are psychometric scores in predicting recidivism for treated sex offenders? *International Journal of Offender Therapy and Comparative Criminology, 56,* 420–446. http://dx.doi.org/10.1177/0306624X11403125

*Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law, 19,* 85–110. http://dx.doi.org/10.1080/1068316X.2011.607820

*Beech, A. R., Erikson, M., Friendship, C., & Ditchfield, J. (2001). *A six-year follow-up of men going through probation-based sex offender treatment programmes* (Research Findings 144). London, England: Home Office Research, Development and Statistics Directorate. Retrieved from www.homeoffice.gov.uk/rds/pdfs/r144.pdf

*Beech, A., & Ford, H. (2006). The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending. *Psychology, Crime & Law, 12,* 685–701. http://dx.doi.org/10.1080/10683160600558493

*Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse, 14,* 155–167.

*Beech, A. R., Mandeville-Norden, R., & Goodwill, A. (2012). Comparing recidivism rates of treatment responders/nonresponders in a sample of 413 child molesters who had completed community-based sex offender treatment in the United Kingdom. *International Journal of Offender Therapy and Comparative Criminology, 56,* 29–49. http://dx.doi.org/10.1177/0306624X10387811

*Beggs, S. M. (2008). *Treatment outcome, risk assessment, and recidivism among sexual offenders against children* (Unpublished doctoral dissertation) University of Canterbury, Christchurch, New Zealand. Retrieved from http://ir.canterbury.ac.nz/bitstream/10092/1824/1/thesis_fulltextpdf.pdf

*Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender version. *Sexual Abuse: Journal of Research and Treatment, 22,* 234–251. http://dx.doi.org/10.1177/1079063210369014

*Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology, 79,* 182–192. http://dx.doi.org/10.1037/a0022900

*Blacker, J. E. (2009). *The assessment of risk in intellectually disabled sexual offenders* (Unpublished master's thesis). University of Birmingham, Birmingham, Alabama. Retrieved from http://etheses.bham.ac.uk/326/1/blacker09forpsyD.pdf

*Blacker, J., Beech, A. R., Wilcox, D. T.,& Boer, D. P. (2011). The assessment of dynamic risk and recidivism in a sample of special needs sexual offenders. *Psychology, Crime & Law, 17,* 75–92. http://dx.doi.org/10.1080/10683160903392376

Boccaccini, M. T., Murrie, D. C., Mercado, C., Quesada, S., Hawes, S., Rice, A. K.,& Jeglic, E. L. (2012). Implications of Static-99 field reliability findings for score use and reporting. *Criminal Justice and Behavior, 39,* 42–58. http://dx.doi.org/10.1177/0093854811427131

Bonta, J.,& Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). New York, NY: Routledge.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Borsboom, D.,& Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9,* 91–121. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185608

*Craig, L. A., Thorthon, D., Beech, A.,& Browne, K. D. (2007). The relationship of statistical and psychological risk markers to sexual reconviction in child molesters. *Criminal Justice and Behavior, 34,* 314–329. http://dx.doi.org/10.1177/0093854806291416

Dawes, R. M., Faust, D.,& Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674. http://dx.doi.org/10.1126/science.2648573

Edens, J. F., Penson, B. N., Ruchensky, J. R., Cox, J., & Smith, S. T. (2016). Interrater reliability of Violence Risk Appraisal Guide scores provided in Canadian criminal proceedings. *Psychological Assessment, 28,* 1543–1549. http://dx.doi.org/10.1037/pas0000278

*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: Journal of Research and Treatment, 24,* 5–28.

*Eher, R., Rettenberger, M., Gaunersdorfer, K., Haubner-MacLean, T., Matthes, A., Schilling, F., & Mokros, A. (2013). Über die Treffsicherheit der standardisierter Risikoeinschätzungsverfahren Static-99 und Stable-2007 bei aus einer Sicherungs maβnahme 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. http://dx.doi.org/10.1007/s11757-013-0212-9

*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

Elliott, I. A., & Beech, A. R. (2013). A U.K. cost-benefit analysis of circles of support and accountability interventions. *Sexual Abuse: Journal of Research and Treatment, 25,* 211–229. http://dx.doi.org/10.1177/1079063212443385

French, S., & Gendreau, P. Reducing prison misconduct: What works! *Criminal Justice and Behavior, 33,* 185–218, 2006. http://dx.doi.org/10.1177/0093854805284406

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, 9,* 865–891. http://dx.doi.org/10.1177/0093854809338545

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. http://dx.doi.org/10.1037/0022-006X.66.2.348

Hanson, R. K., & Harris, A. J. R. (2000a). Where should we intervene? Dynamic predictors of sex offense recidivism. *Criminal Justice and Behavior, 27,* 6–35. http://dx.doi.org/10.1177/0093854800027001002

Hanson, R. K., & Harris, A. J. R. (2000b). *The Sex Offender Need Assessment Rating (SONAR): A method for measuring change in risk levels* (Corrections Research User Report 2000–1). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada. Retrieved from http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/sx-ffndr-nd/index-eng.aspx

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

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research user report 2007–05). Ottawa, ON, Canada: Public Safety Canada. Retrieved from http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf

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

Hanson, R. K., & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis* (User report 2007–01). Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness. Retrieved from https://www.ccoso.org/sites/default/files/import/accuracy-of-risk-assessments.pdf

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. http://dx.doi.org/10.1037/a0014421

*Harkins, L., Thornton, D., & Beech, A. (2009, October). *The use of dynamic risk domains assessed using psychometric measures to revise relative risk assessment using RM 2000 and Static 2002*. Presentation at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

*Hart, S. D., Jackson, K., Healey, J., & Watt, K. A. (2008, July). *Validation of the Risk for Sexual Violence Protocol in adult sexual offenders*. Presentation at the annual meeting of the International Association of Forensic Mental Health Services, Vienna, Austria.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117,* 167–178. http://dx.doi.org/10.1037/0033-2909.117.1.167

*Helmus, L., Babchishin, K. M., & Blais, J. (2012a). 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. http://dx.doi.org/10.1177/0306624X11414693

Helmus, L., Babchishin, K. M., & Hanson, R. K. (2013). The predictive accuracy of the Risk Matrix 2000: A meta-analysis. *Sexual Offender Treatment, 8,* 1–24.

*Helmus, L., & Hanson, R. K. (2012b, 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.

Helmus, L., 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. http://dx.doi.org/10.1177/1524838012462244

*Helmus, L., 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,* 136–157. http://dx.doi.org/10.1080/13552600.2013.870241

*Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Washington, DC: U.S. Department of Justice. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/217618.pdf

*Kropp, P. R. (2001, April). *The Risk for Sexual Violence Protocol (RSVP)*. Paper presented at the founding conference of the International Association of Forensic Mental Health Services, Vancouver, BC, Canada.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1,* 451–476. http://dx.doi.org/10.1007/s11292-005-3541-7

*Lasher, M. P., McGrath, R. J., & Cumming, G. F. (2013, November). *The Vermont Assessment of Sex Offender Risk-2 and the Sex Offender Treatment Intervention and Progress Scale: A model of static and dynamic risk assessment*. Paper presented at the 32nd Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.

*Lofthouse, R. E. (2013). *Risk factors for offending behaviour in adults with an intellectual disability* (Unpublished doctoral dissertation). Bangor University, Bangor, United Kingdom. Retrieved from http://e.bangor.ac.uk/4928/2/R_Lofthouse_thesis%5B1%5D.pdf

*Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., Boer, D. P., & Haaven, J. L. (2013). 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, 26,* 394–403. http://dx.doi.org/10.1111/jar.12029

*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.

*Looman, J., Abracen, J., Serin, R., & Marquis, P. (2005). Psychopathy, treatment change, and recidivism in high-risk, high-need sexual offenders. *Journal of Interpersonal Violence, 20,* 549–568. http://dx.doi.org/10.1177/0886260504271583

*Looman, J., & Goldstein, J. (2014, October). *Incremental validity of the STABLE-2007 with the Static-99R*. Presentation at the 33st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency, 52,* 77–93. http://dx.doi.org/10.1177/0011128705281747

*Lussier, P., Deslauriers-Varin, N., & Râtel, T. (2010). A descriptive profile of high-risk sex offenders under intensive supervision in the province of British Columbia, Canada. *International Journal of Offender Therapy and Comparative Criminology, 54,* 71–91. http://dx.doi.org/10.1177/0306624X08323236

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: Journal of Research and Treatment, 22,* 191–217. http://dx.doi.org/10.1177/1079063210366039

Marlowe, D. B., Festinger, D. S., Lee, P. A., Dugosh, K. L., & Benasutti, K. M. (2006). Matching judicial supervision to clients' risk status in drug court. *Crime & Delinquency, 52,* 52–76. http://dx.doi.org/10.1177/0011128705281746

Martin, S. L., Macy, R. J., & Young, S. K. (2011). Health and economic consequences of sexual violence. In J. W. White, M. P. Koss, & A. E. Kazdin (Eds.), *Violence against women and children. Vol. 1: Mapping the terrain* (pp. 173–195). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12307-008

McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2011). *A model of static and dynamic sex offender risk assessment* (Document No. 236217). Washington, DC: United States Department of Justice. Retrieved from http://www.ncdsv.org/images/VDC_modelOfStaticAndDynamicSexOffenderRiskAssessment_10-2011.pdf

*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: Journal of Research and Treatment, 24,* 431–458. http://dx.doi.org/10.1177/1079063211432475

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. http://dx.doi.org/10.1037/11281-000

*Olver, M. E. (2003). *The development and validation of the Violence Risk Scale: Sexual Offender Version (VRS:SO) and its relationship to psychopathy and treatment attrition* (Unpublished doctoral dissertation). University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

*Olver, M. E., Nicholaichuk, T. P., Kingston, D. A., & Wong, S. C. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology, 82,* 312–324. http://dx.doi.org/10.1037/a0035340

*Olver, M. E., Nicholaichuk, T. P., & Wong, S. C. P. (2013). The predictive and convergent validity of a psychometric battery used to assess sexual offenders in a treatment programme: An 18-year follow-up. *Journal of Sexual Aggression, 20,* 1–24.

*Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38,* 113–126. http://dx.doi.org/10.1177/0093854810389534

*Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19,* 318–329. http://dx.doi.org/10.1037/1040-3590.19.3.318

*Rettenberg, 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.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

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

*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. Retrieved from http://psycnet.apa.org/PsycINFO/2007-99018-302

Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: An international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology, 11,* 597–630. http://dx.doi.org/10.1007/s11292-015-9241-z

*Scoones, C. D., Willis, G. M., & Grace, R. C. (2012). Beyond static and dynamic risk factors: The incremental validity of release planning for predicting sex offender recidivism. *Journal of Interpersonal Violence, 27,* 222–238. http://dx.doi.org/10.1177/0886260511416472

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior, 18,* 32–53. http://dx.doi.org/10.1016/j.avb.2012.09.002

*Smeth, A. (2013). *Evaluating risk assessment among sex offenders: A comparative analysis of static and dynamic factors* (Unpublished doctoral dissertation). Ottawa, Ontario, Canada: Carleton University. Retrieved from https://curve.carleton.ca/system/files/theses/27517_1.pdf

Smid, W. J., Kamphuis, J. H., Wever, E. C., & Van Beek, D. J. (2014). A quasi-experimental evaluation of high-intensity inpatient sex offender treatment in the Netherlands. *Sexual Abuse: Journal of Research and Treatment, 28,* 469–485. http://dx.doi.org/10.1177/1079063214535817

*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

*Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse, 14,* 139–153.

Thornton, D. (2013). Implications of our developing understanding of risk and protective factors in the treatment of adult male sexual offenders. *International Journal of Behavioral and Consultation Therapy, 8,* 3–4. http://dx.doi.org/10.1037/h0100985

*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. http://dx.doi.org/10.1177/1079063213511120

*Wakeling, H., Beech, A. R., & Freemantle, N. (2013). Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. *Psychology, Crime & Law, 19,* 233–252. http://dx.doi.org/10.1080/1068316X.2011.626413

*Wakeling, H. C., Freemantle, N., Beech, A. R., & Elliott, I. A. (2011). Identifying predictors of recidivism in a large sample of United Kingdom sexual offenders. *Psychological Services, 8,* 307–318. http://dx.doi.org/10.1037/a0025516

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

*Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: Journal of Research and Treatment, 19,* 449–465.

Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior, 32,* 172–204. http://dx.doi.org/10.1177/0093854804272889

Received June 3, 2016
Revision received November 8, 2016
Accepted January 11, 2017 ■