

Predicting Probation Supervision Violations

Martin Hildebrand

Private Practice, De Bilt, The Netherlands

A. Michiel Hol

Adaptief.nl, Amsterdam, The Netherlands

Jacqueline Bosker

Dutch Probation Service, Utrecht, The Netherlands

The task of risk assessment is a central feature of probation work and a core activity of probation officers. Risk assessment forms the basis for subsequent interventions and management of offenders so that the likelihood of reoffending is reduced. A primary difficulty for probation workers is the ability to predict the risk of probation violations which could facilitate prevention. The main objective of the present study was to investigate the value of the 61-item Dutch diagnostic and risk assessment tool Recidivism Assessment Scales (RISc) with respect to predicting probation supervision violations of male probationers ($N = 14,363$). Because all RISc assessments included in the study were completed before the start of the supervision period, they could not have been influenced by behavior of the offenders or other circumstances during this period. It was found that the predictive accuracy of the RISc, with regard to supervision violation, was supported. All RISc subscales and the total score significantly predicted probation supervision violation. The AUC demonstrating the strength of the relationship of the RISc total score ($AUC = .70$) is satisfactory. Logistic regression analyses resulted in a fitting model, demonstrating that a selection of only 17 items from the total of 61 RISc items was sufficient to predict probation violation while preserving predictive accuracy ($AUC = .73$). For one of the possible cut-off sum scores used to select groups at high risk for probation violation, it was shown that is possible to double the percentage of correctly identified future violators when compared to the base rate of probation violation.

Keywords: offender assessment, probation, recidivism assessment scales (RISc), supervision violation, prediction

Risk assessment has been commonly used to classify offenders or forensic patients in order to place them in suitable levels of institutional security. The concerns underlying these assessments relate to institutional adjustment and safeguarding against absconding or escape. Assessments are also conducted to assist in decisions regarding when and under what conditions an offender might be released from prison into the community, and are used to determine the appropriate treatment interventions and the level of supervision by probation services required to maintain an offender safely in the community. For effective supervision, not only the risk of recidivism is relevant, but also the risk of noncompliance must be taken into account.

The task of assessing risk is a central feature of probation work and a core activity of the fully qualified probation officer. Risk assessment forms the basis for subsequent interventions and management of offenders so that the likelihood of reoffending is reduced. Andrews, Bonta, and Hoge (1990; see also Andrews & Bonta, 2003) cogently argued that effective interventions to reduce recidivism requires the targeting of appropriate risk factors in offenders. Drawn from the risk/need/responsivity (RNR) model (Andrews et al., 1990), the essence of the risk principle is that treatment is most effective when delivered proportionally to the level of risk of the offender. Thus, higher risk cases should receive more intensive services, whereas lower risk cases should receive less intervention. Risk level is defined as the overall probability of criminal offending that is determined by both the number and severity of risk factors. The need principle refers to the type of treatment targets and suggests that interventions should be geared toward those factors that are most closely related to the risk of criminal offending (i.e., criminogenic needs). Examples of criminogenic need domains include problematic family and marital relationships, substance abuse, emotional instability and procriminal attitudes (e.g., Andrews & Bonta, 2003). The responsivity principle, finally, concerns the delivery of treatment programs in a style and mode that is consistent with the competency and learning style of the offender. The latter principle emphasizes the importance of patient characteristics and conditions that promote or impede positive change. Meta-analyses of the of-

This article was published Online First May 28, 2012.

Martin Hildebrand, Consultant Forensic Psychologist, Private Practice, De Bilt, The Netherlands; A. Michiel Hol, Consultant Psychometrician, Adaptief.nl, Amsterdam, The Netherlands and Former Policy Worker, Dutch Probation Service; Jacqueline Bosker, Senior Policymaker, Dutch Probation Service, Utrecht, The Netherlands.

We thank Martine Wiekeraad for her assistance in retrieving and screening of the research data. We thank Gideon Mellenbergh and the anonymous reviewers for valuable comments and suggestions concerning earlier versions of this article. The views expressed are those of the authors and are not necessarily shared by the Dutch Probation Service.

Correspondence concerning this article should be addressed to Martin Hildebrand, Hessenweg 123-B, 3731 JG De Bilt, The Netherlands. E-mail: mhildebrand@casema.nl

fender recidivism literature (Bonta, Law, & Hanson, 1998; Gendreau, Little, & Goggin, 1996) clearly revealed that dynamic “need” variables correlate both with general and violent recidivism as well as or even better than static factors. This further stresses the importance of targeting these factors for recidivism-reducing strategies, for example community supervision programs.

The Current Study

Predicting the risk of supervision violation or noncompliance provides an important insight into administrative needs as well as policy-making decisions toward more effective and save use of community supervision. Community supervision is considerably less disruptive to the lives of the persons under such supervision than commitment to correctional facilities would be (Gray, Fields, & Maxwell, 2001). For offender rehabilitation to be consistent with the RNR model, knowledge of the offender’s risk level and criminogenic needs is essential (e.g., Andrews & Bonta, 2010; Ogloff & Davis, 2004) and risk and needs assessment should therefore be integrated with rehabilitation efforts (Wong, Gordon, & Gu, 2007). As policymakers (i.e., Dutch Probation Service, Ministry of Justice, 2009) in the Netherlands became increasingly convinced of the effectiveness of the “what works” ideas, it became clear that a national diagnostic and risk assessment tool was needed, one that fulfilled the twin aims of allowing more integrated working between prison and probation staff while also ensuring that all probation officers assessed risk against the same criteria and in the same way. The decision to develop a new tool, entitled the *Recidive Inschattingen Schalen* [Recidivism Assessment Scales] (RISc; Adviesbureau Van Montfoort & Reclassering Nederland, 2004) was taken after an extensive examination of existing risk and needs assessment systems led to the conclusion that none of them fully met prison and probation business requirements in the Netherlands. The aim of the RISc is to deliver a common, efficient and effective offender risk and needs assessment system that enables the prison and probation service to achieve targets for reduction in reoffending/reconviction rates, and for increased protection to the public. The RISc was promoted as a tool to help probation officers accurately (and consistently) assess the risk of reoffending and dangerousness for each offender. It was intended to support the quality of presentence report writing and the design of individual supervision plans as well as the level of supervision. It was also intended to ensure that probation officers’ judgments are comprehensive and evidence based. The RISc is based on the Offender Assessment System (OASys; Home Office, 2002; Howard, Clark, & Garnham, 2006) of the probation and prison service of England and Wales, and both instruments are highly comparable. RISc shows considerable similarities to the Level of Service Inventory—Revised (LSI-R; Andrews & Bonta, 1995) as well since OASys took its conceptual basis from the LSI-R (now LS/CMI). Once developed, the internal consistency and reliability of the RISc has been checked in other studies (van der Knaap, Leenarts, Born, & Oosterveld, 2012; van der Knaap, Leenarts, & Nijssen, 2007). Analyses further showed that the RISc has acceptable (AUC = .70) predictive validity regarding to general recidivism (van der Knaap & Alberda, 2009).

A primary difficulty for probation workers is the ability to monitor and detect probation violations. Predicting noncompliance and knowledge of risk factors related to noncompliance may help probation officers to respond effectively to a risk of noncompliance and help offenders to complete their sentence. This exploratory study focused on the characteristics (risk factors) of male probationers who showed noncompliant behavior during their probation. Specifically, the study investigates the value of the Dutch diagnostic and risk assessment tool RISc with respect to predicting probation supervision violations of male probationers.

Until now, the RISc is used for the assessment of criminogenic needs, responsivity, and the risk of recidivism. Some risk factors that are related to recidivism, however, are also linked to probation violation. In a meta-analysis of offender treatment attrition, Olver, Stockdale, and Wormith (2011) concluded that dropout from psychological treatment was related to risk factors (e.g., antisocial personality disorder, criminal history) as well as responsivity issues (e.g., low motivation, poor engagement), and dropout was significantly associated with recidivism. Therefore, it would be relevant to investigate whether the RISc, or a subset of items in the RISc, can be used to accurately predict the risk of noncompliance. Such a prediction can improve sentence planning. In case of a high risk of noncompliance, specific interventions should be taken to reduce this risk and help an offender to complete probation.

Method

Participants

The initial sampling frame comprised data on supervision within the framework of conditional release pending trial, conditional non-prosecution, conditional or suspended sentence and supervised release with a known outcome by 30,901 adult male offenders in the Netherlands in the period January 1, 2007 until July 1, 2010.¹ Data on supervision violation were obtained from the Dutch Probation Service. Cases were excluded from the study when no RISc was administered (14.5%); when a RISc could not reliably be matched to the supervision violation data (0.3%); when the RISc was administered using a slightly different version with some minor differences with regard to scoring instructions (3.4%); when the RISc was completed too long (more than one year) before the start of the supervision, or when the RISc was completed after the start of the supervision; it was undesirable that RISc assessments of probationers included in this study were in any way influenced by their behavior or circumstances

¹ Actually, data of 38,934 offender supervisions of adult males were available with some offenders having multiple supervisions within the research period. Keeping all the data of persons that occur more than once in the database would violate the assumption of independence that is made in the statistical analysis. Therefore, of persons occurring more than once, only one case was selected randomly. This procedure led to a database of 30,901 cases, all being unique offenders.

during supervision (33.3%).² In addition, 2.0% of the cases were excluded because the RISC assessment did not meet the minimum standards of data completion required for profiling the risks and needs of offenders according to the scoring instructions. This led to a final research group of 46.5% of all conditional sentences completed by adult male offenders in the Netherlands in the research period with a known outcome and consists of a total of 14,363 male offenders. The mean age of the 14,363 participants was 33.4 years (Median = 32, Mode = 20, $SD = 11.6$; range = 18 to 94 years).

As the selection process involved the rejection of a rather large number of cases, the 16,538 unselected cases and the finally selected research group ($N = 14,363$) were compared on available offender characteristics. Analyses show some indication that the research group can be described as a representative sample of the total group. Although supervisions that were ended between January 2007 and July 2010 were terminated significantly more often due to supervision violations (31.4%) in the research group than in the rest of the group (27.1%), $\chi^2 = 69.9$, $df = 1$, $p < .01$, its effect size could be regarded as small, Cramer's $V = .05$. Furthermore, a t test with group membership (research vs. no research group) as independent variable showed no significant difference in mean age (33.4 vs. 33.5 years; $t = 1.08$, $df = 30,899$, $p = .279$). Unfortunately, no other relevant variables could be examined to support the assumption of the representativeness of the research group.

Measures

RISC. The main body of the RISC consists of 61 scored questions across 12 domains of criminogenic needs (see Table 1)—the first two sections, which cover Offense History and Current Offense and Offense Pattern being combined into a “Prior and Current Offenses” scale. While the “Prior and Current Offenses” scale covers offending information, the other scales focus on Accommodation, Education and Employment, Income and Finances, Relationships with Partner, Family, and Relatives, Peer Relationships, Drug Use/Abuse, Alcohol Use/Abuse, Emotional Wellbeing, Thought Pattern, Behavior, and Social Skills, and Attitude (see the Appendix for sample items of the RISC scales).

The majority of the 61 RISC items are scored 0 (*no problems*), 1 (*some problems*) or 2 (*significant problems*), but some of the items use a dichotomous 0 (*no problems*) / 2 *yes*, (*significant problems*) response scale. The RISC is scored by summing the item scores within each domain or scale, with higher scores corresponding to increased criminogenic needs. In general, missing items are scored 0 when the minimum requirements of data completion, according to the RISC manual, are fulfilled.³ A criminogenic need is said to be present when the offender scores above a certain threshold of the maximum unweighted score available for the risk factors. The raw total score of the RISC varies from 0 to 122. Raw (unweighted) domain scores are converted into weighted scores, recognizing that not all the criminogenic needs are equally correlated with the likelihood of reconviction. For ordinal-level risk classification (low risk, medium-low risk, medium-high risk, high risk), (weighted) cutoff scores are offered.

Outcome variable. Supervision violation or noncompliance is defined as the termination of a conditional sentence because of an unacceptable breach of the conditions, for example, refusal to participate in treatment, repeated nonattendance to appointments with the probation officer, or making contact with a person with

whom contact is prohibited. When this happens, the probation service reports the breach to the public prosecutor who then decides whether or not to end probation supervision and transposes the conditional sentence into an unconditional sentence, most of the time a prison sentence. A second cause for termination of a conditional sentence is a conviction for a further offense. In this case, the public prosecutor also decides whether or not probation supervision can be continued.

Administration

The RISC assessments in our study were conducted by a large pool of at least 1,407 local probation officers at the probation agencies in the Netherlands.⁴ Completing a full RISC assessment takes about 5 to 7 hours. This includes collecting and reading an offender's file (e.g., criminal records and probation services files), conducting the offender's interview, completing the computerized RISC, formulating a sentence plan and consulting a senior probation officer to discuss the results of the RISC. As a general rule, probation officers take a 4-day training program on administering the instrument that covers relevant interview techniques, response categories, item meanings, and quality assurance issues.

Data Analyses

Reliability of the total RISC, as well as of its constituent domains, was assessed by calculating Cronbach's α . According to Bland and Altman (1997), Cronbach's α of .70–.80 is considered satisfactory for a reliable comparison between groups. However, for clinical purposes, a minimum of .90 is advised (Bland & Altman, 1997). In addition, mean interitem correlations were calculated. According to Clark and Watson (1995), a mean interitem correlation of .15–.20 is desirable for scales that measure broad characteristics, while values of .40–.50 are required for scales tapping narrower ones.

In order to examine differences between the scores on the RISC scales for the two groups under study (offenders violating probation supervision vs. nonviolators), t tests were calculated. Additionally, Cohen's d (Cohen, 1988), which provides an indication of the magnitude of the difference between two sample means in relation to the distribution of the scores within the samples, was computed. Typically, an effect size of .20 is considered a small effect, .50 a moderate effect, and .80 is considered a large effect (Cohen, 1988).

In keeping with practices by other researchers (e.g., Rice & Harris, 1995) of risk assessment measures, receiver operating characteristic (ROC) analysis was used to evaluate the ability of

² When multiple RISC assessments were available that satisfied this criterion, as well as the other criteria, only the assessment that was closest in time before the start of the probation supervision period was included in the study.

³ When a substantial part of the RISC items of scales is missing, simple alternative procedures are used to obtain a total score on the RISC. When too much information is missing, the RISC assessment is regarded as invalid and no total score can be obtained.

⁴ The exact number is unknown; two different probation officers may appear under the same name in the database (e.g., John Jones and Jack Jones could both appear as J. Jones). The database does not contain specific identification numbers for individual probation officers.

Table 1
 Scored RISC Scales (Unweighted Sum Scores)

	RISC scales	No. scored questions	Score range	Criminogenic need cutoff
1–2	Prior and Current Offenses	8	0–16	8+
3	Accommodation	4	0–8	2+
4	Education and Employment	7	0–14	4+
5	Income and Finances	4	0–8	3+
6	Relationships (partner, family, relatives)	5	0–10	4+
7	Peer Relationships	4	0–8	3+
8	Drug Use/abuse	6	0–12	2+
9	Alcohol Use/abuse	5	0–10	2+
10	Emotional Wellbeing	5	0–10	5+
11	Thought pattern, Behavior, Social Skills	8	0–16	4+
12	Attitude	5	0–10	4+
	RISC total	61	0–122	

Note. RISC = Recidivism Assessment Scale.

the RISC total score and domain scores to predict the offender's violation of probation supervision. ROC analysis computes an area under the curve (AUC) statistic by plotting the sensitivity of a tool against its specificity (Mossman, 1994). The AUC score can range from 0 (*perfect negative prediction*) to 1 (*perfect positive prediction*); an AUC of .5 indicates that the tool is not able to predict any better than chance (Douglas, Skeem, & Nicholson, 2011). As a rule of thumb, AUCs between .65 and .70 are considered moderate, AUCs of .70 or above satisfactory, and measures above .75 typically indicate good predictive accuracy (e.g., Hosmer & Lemeshow, 2000; Quinsey, Harris, Rice, & Cormier, 1998).

Logistic Regression Analysis

A logistic regression analysis was conducted to assess whether there is potential to create a shorter and more accurate instrument to predict supervision violation. To determine a prediction model, all 61 RISC items were included in a backward stepwise logistic regression analysis, with supervision violation as the dependent variable. In backward stepwise regression, at first, all variables are entered in the model and subsequently it is tested step by step whether variables can be removed. This method is often used in both explorative and predictive research (Menard, 1995). Given the availability of a large number of cases, the statistical power was large in the present study. Although generally considered as an advantage, large sample sizes may result in almost any statistic being significantly different from 0, while differences may be meaningless in practice (Murphy & Myors, 2004, p. 5). Generally, the significance level in a stepwise procedure is set at .10 and sometimes even to .20 to prevent failure to find existing relationships (Menard, 1995, p. 55). Given the very large sample size in the present study a stringent significance level of .01 was used for variable removal to avoid ending up with variables that have little or no predictive value.

Testing a model on the data that were used to estimate the model will almost certainly overestimate its performance (Mosteller & Tukey, 1977). Therefore, the research group was divided in a calibration sample and a test sample. The calibration sample was created by randomly selecting two-thirds of all cases in the research group; this calibration sample was used for the logistic regression analysis. The other one-third of the cases was used as a test sample, that is, to

assess the quality of the prediction model (Breiman, Friedman, Olshen, & Stone, 1984).

For the polytomous RISC items, a repeated contrast was used in the logistic regression analysis. In the repeated contrast, each category of an independent variable except the first (the reference category) is compared with the previous category, which is also referred to as backward difference coding. Such a contrast facilitates the interpretation of the category coefficients (Chen, Ender, Mitchell, & Wells, n.d.; Menard, 1995). Since all polytomous RISC items are ordinal and scored in the same direction, *B* coefficients are all expected to be positive using this contrast.

Missing values. As stated before, in general, missing items are scored 0 when the minimum requirements of data completion, according to the RISC manual, are fulfilled. A reason to score missing items as 0 when using the RISC in practical probation work is that value imputation methods could result in labeling an offender as having a high recidivism risk, just because a substantial number of item responses is missing. Although this strategy can be justified in practical work, we preferred to use a more sophisticated missing value handling method in the logistic regression analysis, because numerous studies have shown that these methods lead to better results (Schafer & Graham, 2002).

Logistic regression requires a complete item response pattern for each case that is entered in the analysis. Therefore, normally only persons with complete data can be included in the analysis. Generally, this leads to a substantial decrease in sample size and it could lead to biased model estimates. To overcome the problem of missing data, multiple imputation (MI) was used to fill in the missing item scores in the calibration sample that was used for the logistic regression analysis (Azur, Stuart, Frangakis, & Leaf, 2011; Schafer & Graham, 2002). In MI, multiple complete data sets are created where each missing value is replaced by a simulated value. The simulated values are based on the item scores of a given individual and on the observed relations in the data for the other participants (Azur et al., 2011; Schafer & Graham, 2002). All 61 RISC items, as well as the dependent variable, were included in the imputation model; although inclusion of the dependent variable may seem illegitimate, its inclusion in the imputation model is essential (Allison, 2001).

Table 2
RISc Unweighted Sum Scores According to Probation Violation Status (N = 14,363)

Scale	Probation violation (<i>n</i> = 4,512)		No violation (<i>n</i> = 9,851)		Cohen's <i>d</i>	Point-biserial correlation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
1–2	Prior and Current Offenses	7.3	3.8	5.1	3.8	0.58	.26*
3	Accommodation	2.0	2.4	1.0	1.7	0.49	.22*
4	Education and Employment	6.4	3.8	4.4	3.8	0.53	.24*
5	Income and Finances	3.2	2.4	1.9	2.1	0.58	.26*
6	Relationships (Partner, Family, Relatives)	4.1	2.5	3.4	2.5	0.28	.14*
7	Peer Relationships	2.8	2.1	1.9	1.9	0.45	.21*
8	Drug Use/Abuse	4.1	3.7	2.3	3.3	0.51	.24*
9	Alcohol Use/Abuse	3.0	3.2	2.2	3.0	0.26	.11*
10	Emotional Wellbeing	3.8	2.6	3.2	2.7	0.23	.09*
11	Thought Pattern, Behavior, Social Skills	8.4	3.5	6.7	3.7	0.47	.21*
12	Attitude	3.9	2.5	2.9	2.4	0.41	.19*
	RISc total	48.9	19.1	35.0	18.8	0.73	.32*

Note. RISc = Recidivism Assessment Scale. Point-biserial correlation between violation (yes/no) and scale score.

* $p < .01$.

With regard to the MI procedure, the method of multiple imputation using chained equations (MICE) was used. Using MI in a stepwise logistic regression analysis involves fitting the model under consideration to all imputed data sets and combining all model estimates across imputed data sets at each variable selection step (White, Royston, & Wood, 2011). Wood, White, and Royston (2008) proposed and tested some pragmatic methods to simplify this procedure for large datasets with many predictor variables. In the current study, their proposed method W2 was used.⁵ In this method all (*m*) imputed data sets are analyzed as one data set consisting of $m \times n$ cases in a weighted logistic regression analysis. In this method, the weight applied to all observations is $w = (1 - f)/m$, where *f* is the average fraction of missing data across all variables, and is calculated as the total number of missing values across all variables (*p*), including the dependent, divided by $p \times n$. After a model has been selected by the weighted stepwise procedure on $m \times n$ cases, this model is fitted to all imputed data sets to obtain pooled estimates using Rubin's rules. This is done by a second logistic regression analysis where the items selected by the stepwise method are now included using the enter method (Wood et al., 2008, p. 3232).

To determine the number of needed imputed data sets (*m*), the rule of thumb described by White et al. (2011) was used: *m* should be at least equal to the percentage of incomplete cases. Since the calibration set (*n* = 9,686; see the Results section) contained 51% of incomplete cases, 51 imputed data sets were generated, which were analyzed as one very large dataset in the weighted stepwise logistic regression analysis. Although 51% of the cases in the calibration set had at least one missing response, on item level for most (*n* = 52) of the 61 items the percentage of missing values was below 5%.⁶ For nine items the percentage of missing values ranged between 8% and 16%. Because the research was restricted to probation supervisions with a known outcome, the variable supervision violation never had a missing value.

Evaluation of the logistic regression model. Evaluation of the model was done on the complete cases that were available in the test sample. Model fit was assessed by the Hosmer-Lemeshow goodness-of-fit statistic \hat{C} , after computing predicted probabilities

for the test sample with the model obtained from the calibration sample (Hosmer & Lemeshow, 2000, p. 148).

Classification accuracy of the logistic regression model was assessed using a ROC analysis. The AUC value was calculated for the predicted probabilities by the logistic regression model as well as for the unweighted sum score of the items that were included in the regression model, because in practice sum scores are often used (for reasons of simplicity). In addition, Goodman and Kruskal's coefficient λ (Everitt, 1977) was calculated; λ represents the proportional reduction of falsely identified probationers when the prediction model is used.

To give an idea about the clinical utility of the model, for each possible cut-off criterion of the sum score percentages of correctly and falsely identified probationers, as well as the sensitivity and specificity, will be calculated. An α level of .01 was used for all statistical tests that were conducted in this study.

Results

Reliability

Cronbach's α for the total RISc was .93. Using Cronbach's α cut-off points of .70 for adequate scores and .80 for high scores, five of the scales (Education and Employment; Drug Use/Abuse;

⁵ Given equal percentages of missing data across variables, Wood and colleagues (2008) found that their W2 method performed almost equally well compared to their W3 method considering a simulation scenario using a binary model. In a simulation scenario using a linear model where the amount of missing data was increased to 50% for two variables, the W3 method performed better. Because the percentage of missing data in the present study was below 5% for most items and never exceeded 16%, we decided to use the W2 method which is much easier to implement.

⁶ For the RISc item Motivation to address drinking, missing values were replaced by a score 0 when the scores of 4 other variables made it clear that the individual did not have a drinking problem. In this way the percentage of missing values could be reduced from 49.7% to 2.0% for this particular item.

Alcohol Use/Abuse; Thought Pattern, Behavior, Social Skills; and Attitude), as well as the RISC total score, had high levels of reliability, and a further three scales (Accommodation; Peer Relationships; Emotional Wellbeing) had adequate reliability. Prior and Current Offenses, Income and Finances, and Relationships (Partner, Family, Relatives) had nonadequate reliability. All scales, except Prior and Current Offenses, Income and Finances, Relationships, Emotional Wellbeing, and Thought Pattern, Behavior, Social Skills yielded a mean interitem correlation value of $\geq .40$ that is required for the reliable use of “narrow” (i.e., specific) scales.

Probation Supervision Violation Analyses

Comparison of mean scores. Table 2 presents mean scores on the RISC and on its constituent domains (criminogenic needs) for the two groups under study (probation violation vs. no violation). Because additional inspection of histograms resulted in the conclusion that all scales, except scale 11 and the total score, had skewed frequency distributions, for these scales the groups were compared using the Mann–Whitney *U* test (Gibbons, 1985). Both the Mann–Whitney *U* tests, as additional *t* tests on scale 11 and the RISC total sum score, showed that violators scored significantly higher on all RISC scales ($p < .01$). The effect sizes varied from .23 (Emotional Wellbeing) to .73 (RISC total score), with effect sizes of most of the scales in the .41–.58 range. Point-biserial correlations between RISC scores and violation of probation supervision were significant for all RISC scales (see Table 2).

Predictive validity of the RISC—AUC scores. The RISC total score as well as the 11 subscales significantly predicted violation of probation supervision, with AUC values ranging from .56 to .70 (see Table 3). The AUC demonstrating the strength of the relationship of the RISC total score (AUC = .70) can be qualified as acceptable. Modest predictive validity (AUCs .65 and .66) was found for Prior and Current Offenses, Education and Employment, Income and Finances, and Drug Use/Abuse. The AUCs of the other scales (i.e., Accommodation; Relationships; Peer Relationships; Alcohol Use/Abuse; Emotional Wellbeing; Thought Pattern, Behavior, Social Skills; and Attitude) were $\leq .63$.

Logistic Regression Analysis

Of the 14,363 cases in the research group, 67% ($n = 9,686$) were randomly assigned to the calibration sample to establish a prediction model in the logistic regression analysis; 33% ($n = 4,677$) was assigned to the test sample that was used to test the predictive quality of the model.

Concerning multicollinearity among the 61 items that were used to start the backward stepwise method, all tolerance values exceeded the .20 criterion, indicating no serious multicollinearity problem (Menard, 1995, p. 66).⁷ The weighted stepwise logistic regression analysis on the 51 imputed datasets resulted in a model consisting of 17 RISC items, which were subsequently entered in a second logistic regression analysis to obtain pooled estimates using Rubin’s Rules. The pooled logistic regression coefficients that were obtained using a repeated contrast are shown in Table 4. Note that contrasts for ordinal variables have no effect on and no implications about the model

fit or on the statistical significance of the categorical ordinal variable as a whole. Contrast results may however suggest appropriate recoding of variables by interpreting the direction and significance of individual coefficients (Menard, 1995, pp. 51–52).

Because a repeated contrast was used for the ordinal polytomous items, it was expected that all *B*-parameters would be positive. However, for the items Work experience and Employment track record (item 4.4) and Frequency of drug use in the past (item 8.1b) the rounded value of the second parameter was found to be 0 and not significant, indicating that the last two categories of these items could be treated as one category. For example, item 4.4 currently has the following three ordered categories: probationer has always had a job, probationer mostly has a job, probationer has never worked. Although combining the last two categories into one would lead to a loss of information, the results suggest that this item could do with the following two categories: probationer has always had a job, probationer mostly has a job or has never worked.

The second parameter was found to be negative and not significant for the items Current accommodation situation (item 3.2), History of close relationships (with partner) in adulthood (item 6.2), and Quality of current relationship with partner, family, and other relatives (item 6.3). Although normally a negative second parameter would indicate that the order of the second and third category of these items should be reversed, these results should only be considered as an indication that these two categories could also be treated as one, since the second parameter did not differ significantly from 0.

Interestingly, for item Mental health problems (item 10.2) both parameters are negative, which suggests that this item should be recoded. In the model, a low score on this item corresponds with a higher risk on probation violation. The first parameter of item 10.2 was not significant, indicating that the first and second category could be taken together. This means that a score in the third category indicates a lower risk on probation violation than a score in the first two categories. This result is in contrast with our expectations. When this item is entered individually in a logistic regression, in contrary to the above, the results showed that the second and third category could be taken together, and, as expected, a low score on the item indicates a lower risk on violation. Apparently, some kind of interaction with the other variables in the model results in a reverse effect for this item.

For the rest of the items the *B* parameters were all positive. However, for the items Number of convictions as a juvenile (item 1.5), Type of drug use in the past (item 8.1a), and Insight and attitude toward self and criminal behavior (item 12.4) both positive parameters were not significant. An additional analysis where the categories of these items were coded with a simple contrast showed that it would be most sensible to take the first two categories of these items together.

For the items Current financial situation (item 5.2), Dominant behavior (item 11.3), Self control (item 11.4), and Problem handling (item 11.6) the second parameter was not significant showing that it would be most sensible to take the last two categories of these items together. Table 4 further shows that the single *B*-parameters of the

⁷ Tolerance statistics were computed on complete cases ($n = 4,700$) for the 61 items in the calibration sample, i.e. not on the imputed data sets.

Table 3

Predicting Violation of Probation Supervision: Areas Under ROC Curves With Level of Statistical Significance for RISC Scales, Unweighted Sum Scores (N = 14,363)

	Scale	ROC-AUC	99% CI
1-2	Prior and Current Offenses	.66*	.65-.67
3	Accommodation	.63*	.62-.64
4	Education and Employment	.65*	.64-.66
5	Income and Finances	.66*	.64-.67
6	Relationships (partner, family, relatives)	.58*	.57-.60
7	Peer Relationships	.63*	.62-.64
8	Drug Use/abuse	.65*	.63-.66
9	Alcohol Use/abuse	.57*	.55-.58
10	Emotional Wellbeing	.56*	.55-.58
11	Thought pattern, Behavior, Social Skills	.63*	.62-.64
12	Attitude	.62*	.60-.63
	RISC total	.70*	.69-.71

Note. RISC = Recidivism Assessment Scale; ROC-AUC = area under the receiver operating characteristic curve; CI = confidence interval.

* $p < .01$.

other dichotomous items were all significant and in the expected direction.

Model fit. The quality of the model was assessed by applying the model to the data of the test sample ($n = 4,677$). For the 17 items included in the model, the test sample contained 2,988 complete cases (64%). The percentage of complete cases is higher than in the calibration sample, because in the test sample complete cases are needed for only 17 (instead of the 61) items that were used in the stepwise logistic regression analysis. The model obtained from the calibration sample was used to obtain predicted probabilities for probation violation of complete cases in the test sample ($n = 2,988$). The Hosmer-Lemeshow goodness-of-fit statistic yielded a value of $\hat{C} = 10.5$ ($df = 8$, $p = .23$). The null hypothesis that the model fits could not be rejected. Hence, the test supports the fit of the model on the test sample. It is concluded that the model shows acceptable data fit; generally in large samples only slight deviations from perfect fit could lead to significant rejection of the model (Kramer & Zimmerman, 2007). Moreover, inspection of observed and expected frequencies within each decile of risk showed no peculiarities.

Predictive Validity

The predictive value of the logistic regression model in the test sample was acceptable (AUC = .73, $p < .01$; 99% CI [.70, .75]), when the predicted probabilities of violation were used in the analysis. The value of Goodman and Kruskal's coefficient λ was .12 when predicted probabilities of the logistic regression model were used to predict group membership (violation vs. nonviolation). This value means that the total proportion of falsely identified probationers could be reduced by 12% when the model is used.

Clinical Utility

In practice, professionals (e.g., probation workers, psychologists) often work with unweighted sum scores of the items that were selected in a prediction model. To show the potential of practical use of the model, the clinical utility was assessed using the sum score of the prediction model, consisting of 16 instead of 17 items; it was decided not to use item 10.2 (Mental health problems) in the sum score, because this item led to an unexpected result which could not

yet be explained (see the logistic regression analyses paragraph). Elimination of item 10.2 led to an additional 24 complete cases in the test sample, therefore $n = 3,012$.

Table 5 shows all possible sum scores of the 16 item model, which are used as cut-off scores: for each possible cut-off score the percentage of correctly identified probationers is given.

For example, a cut-off score of 20 means that probationers with a sum score of 20 or higher are classified as future violators and probationers with a score lower than 20 are classified as future nonviolators. A cut-off score of 20 to predict group membership corresponds with a value of .10 for Goodman and Kruskal's coefficient λ , which indicates that the total proportion of falsely identified probationers could be reduced by 10% when the sum score is used. The predictive value of the model, as measured by the AUC, when the sum score was used instead of the probabilities predicted by the logistic regression model, was .71 ($p < .01$; 99% CI [.68, .73]).

Table 5 also shows the total percentage of correctly identified probationers (columns 1 and 2), the percentages of probationers that were correctly identified as violator (correct positives; column 3) or as nonviolation (correct negatives; column 4), as well as the percentages of probationers that were falsely identified as violator (false positives; column 5) or as nonviolation (false-negatives; column 6). The last two columns show the sensitivity and the specificity of the 16 item instrument. The percentage of correctly identified probationers is the highest for a cut-off score of 20, namely 71.1%. For all other cut-off scores the percentage of correctly identified probationers is lower.

Table 5 (bottom) also shows the percentage of probationers that could be classified correctly without using an instrument, that is, just by using the base rate of probation violation, which is 32.2% for the 3,012 probationers. To maximize the percentage of correctly identified probationers with regard to violation *without* a model or without other knowledge, it would be most successful to identify every probationer as a nonviolation given this base rate, because most of the probationers/offenders (67.8%) will not violate the probation requirements. Using the model could increase the percentage of correctly identified probationers to 71.1%. Although one can argue that this increase is rather small (from 67.8%

Table 4
Pooled Logistic Regression Estimates Obtained From 51 Imputed Datasets Based on the Calibration Sample (n = 9,686)

	Predictor (RISc item)	B	SE B	p	Exp(B)
1.5	Number of convictions as a juvenile				
	(1)	.16	.07	.013	1.18
	(2)	.10	.01	.295	1.11
1.7	Previous noncompliance with probation conditions	.28	.07	.000*	1.33
1.8	Severity of current or previous charges	.20	.06	.001*	1.22
2.11a	Offenses are part of a pattern	.17	.06	.003*	1.19
3.2	Current accommodation situation				
	(1)	.32	.08	.000*	1.38
	(2)	-.03	.11	.779	0.97
4.4	Work experience and employment track record				
	(1)	.41	.06	.000*	1.51
	(2)	-.00	.07	.974	1.00
5.2	Current financial situation				
	(1)	.21	.06	.001*	1.23
	(2)	.09	.07	.188	1.09
5.4	Gambling addiction or other addiction (that eats into the main source of income)	.31	.06	.000*	1.37
6.2	History of close relationships (with partner) in adulthood				
	(1)	.08	.08	.322	1.08
	(2)	-.25	.10	.015	0.78
6.3	Quality of current relationship with partner, family, and other relatives				
	(1)	.21	.06	.001*	1.23
	(2)	-.02	.06	.707	.98
8.1a	Type of drug use in the past				
	(1)	.14	.09	.144	1.15
	(2)	.12	.09	.165	1.13
8.1b	Frequency of drug use in the past				
	(1)	.22	.09	.010	1.25
	(2)	.00	.08	.988	1.00
10.2	Mental health problems				
	(1)	-.12	.06	.057	.89
	(2)	-.24	.07	.001*	.79
11.3	Dominant behavior				
	(1)	.19	.06	.002*	1.20
	(2)	.08	.08	.311	1.09
11.4	Self control				
	(1)	.28	.07	.000*	1.33
	(2)	.10	.06	.117	1.10
11.6	Problem handling				
	(1)	.28	.09	.002*	1.32
	(2)	.03	.06	.639	1.03
12.4	Insight and attitude towards self and criminal behavior				
	(1)	.14	.08	.083	1.15
	(2)	.18	.07	.010	1.20
	Constant	-.96	.07	.000*	.38

Note. RISc = Recidivism Assessment Scale.
 * $p < .01$.

to 71.1%), it should be emphasized that using the model facilitates the selection of groups that contain more violators than nonviolators, which enables to distribute prevention measures much more efficiently. For example, it can be derived from Table 5 that 69.4% of the probationers in the group of persons with a sum score equal or higher than 25 was a violator, i.e.

$$\frac{\% \text{ Correct positives}}{\% \text{ Correct positives} + \% \text{ False positives}}$$

whereas the total group of probationers in Table 5 consists of a minority (32.2%) of violators. For the group selected with a cut-off score of 28, the percentage of violators is even 85.7%, although it

should be noted that this score group contains only a very small number of cases.

Discussion

The main purpose of this study was to examine whether the RISc, the Dutch diagnostic and risk assessment tool developed to help probation officers to accurately and consistently assess the risk of reoffending and dangerousness for each offender, is a valuable tool to predict probation supervision violations of male probationers. Also, the reliability of the RISc scales was examined.

A first general conclusion is that the majority of these scales reach levels of reliability that are within generally acceptable

Table 5

Association Between Cut-Off Sum Score(S) of 16 Items Selected in the Logistic Regression Model and the Percentage of Offenders Correctly Identified as Whether or not Violating Probation Supervision in the Test Sample (n = 3,012)

Total score 16 items	Total correct	Correct +	Correct -	False +	False -	Sensitivity %	Specificity %
≥ 1	32.5	32.0	.5	67.3	0.1	99.6	0.8
≥ 2	33.8	31.9	1.9	65.9	0.3	99.2	2.8
≥ 3	36.0	31.7	4.3	63.5	0.4	98.7	6.3
≥ 4	38.7	31.5	7.2	60.6	0.7	97.8	10.6
≥ 5	42.6	30.9	11.7	56.1	1.2	96.2	17.2
≥ 6	45.9	30.1	15.8	52.1	2.1	93.6	23.3
≥ 7	49.0	29.2	19.8	48.0	3.0	90.7	29.2
≥ 8	52.4	27.9	24.5	43.4	4.3	86.7	36.1
≥ 9	55.7	26.9	28.8	39.0	5.3	83.5	42.4
≥ 10	58.4	25.2	33.2	34.7	7.0	78.3	48.9
≥ 11	61.3	23.6	37.7	30.1	8.6	73.3	55.6
≥ 12	64.4	22.3	42.1	25.8	9.9	69.2	62.0
≥ 13	66.4	20.6	45.8	22.1	11.6	63.9	67.4
≥ 14	67.3	18.6	48.7	19.1	13.6	57.7	71.8
≥ 15	68.1	16.5	51.6	16.2	15.7	51.2	76.1
≥ 16	69.4	14.8	54.6	13.2	17.3	46.1	80.6
≥ 17	69.8	13.1	56.7	11.1	19.1	40.8	83.6
≥ 18	70.0	11.5	58.5	9.3	20.7	35.7	86.3
≥ 19	70.5	10.0	60.5	7.4	22.1	31.2	89.1
≥ 20	71.1	8.4	62.7	5.1	23.8	26.0	92.5
≥ 21	70.6	6.8	63.8	4.1	25.4	21.2	94.0
≥ 22	70.2	5.4	64.8	3.0	26.7	16.9	95.5
≥ 23	69.7	4.2	65.5	2.3	28.0	12.9	96.6
≥ 24	69.6	3.4	66.2	1.6	28.8	10.5	97.7
≥ 25	69.2	2.5	66.7	1.1	29.7	7.7	98.4
≥ 26	68.6	1.5	67.1	0.7	30.6	4.7	99.0
≥ 27	68.4	1.0	67.4	0.5	31.2	3.1	99.3
≥ 28	68.3	0.6	67.7	0.1	31.5	2.0	99.8
≥ 29	68.1	0.4	67.7	0.1	31.7	1.3	99.9
≥ 30	68.0	0.2	67.8	0.1	32.0	0.6	99.9
<i>Categorizing every offender as a violator or nonviolator (without using instrument)</i>							
Violator	32.2	32.2	0	67.8	0	100	0
Nonviolator	67.8	0	67.8	0	32.2	0	100

Note. Correct + = Correctly identified as violation; Correct - = Correctly identified as no violation; False + = Falsely identified as violation; False - = Falsely identified as no violation. Because only one person had a sum score of 31, and there were no persons with a score of 32, these scores were not reported.

ranges. Second, the predictive accuracy of the RISC with regard to supervision violation was supported. Third, logistic regression analyses resulted in a fitting model, demonstrating that a selection of only 17 items from the RISC, which consists of 61 items, was sufficient to predict probation violation while preserving predictive accuracy. Moreover, the contrasts included in the logistic regression analysis showed that some response categories could be taken together, which could simplify the instrument.

Regarding reliability, Cronbach's α for the total RISC was .93 and fulfilled the criterion for clinical usefulness of the instrument; although the high α may be influenced by the large number of items in the RISC (Schmitt, 1996). In addition, most of the RISC scales have α coefficients equal to or greater than .70, with only three exceptions, and from the latter, two were close to acceptable levels.

With regard to the predictive accuracy of the RISC, it was found that all RISC subscales as well as the total score significantly predicted probation supervision violation. The AUC demonstrating the strength of the relationship of the RISC total score (AUC = .70) is satisfactory. The predictive accuracy for individual RISC

scales, however, was lower (AUCs between .56 and .66). As noted previously, AUCs in the .50s are considered to have little or no predictive accuracy, those in the .60s are considered weak, those approaching or above the .70s are moderate, and measures above .75 typically indicate good predictive accuracy (e.g., Hosmer & Lemeshow, 2000; Quinsey et al., 1998). It is generally acknowledged that the accuracy levels achieved for most current instruments, across a large variety of samples and outcome variables, generally fall in the range of .65 to .75. Thus, we may conclude that RISC predictive accuracies for probation supervision violations are similar to AUCs obtained by other major instruments (e.g., the HCR-20, LSI-R, OASys, VRAG) in the field of forensic risk assessment (e.g., Dahle, 2006; Quinsey et al., 1998).

The stepwise logistic regression analysis resulted in a fitting model consisting of 17 of the total of 61 RISC items while preserving predictive accuracy (AUC = .73). The model selected in this study indicated that both static (e.g., Number of convictions as a juvenile, Previous noncompliance with probation conditions, Frequency of drug use in the past) as dynamic items (e.g., Current financial situation, Dominant behavior, Problem handling) are

important in predicting future probation supervision violation (see Table 4). Although helpful in the identification of risk of supervision violation, the static items may be less useful in daily risk management. Risk management is the most important component in reducing the subsequent level of risk and the probability of another violation or offense. However, only dynamic measures are changeable. Future research should therefore primarily be aimed at identifying additional dynamic factors which are powerful predictors. Research should also be aimed at studying prevention measures that are effective in reducing violation risk for offenders that score high on static factors. Both research strategies could help to improve the management of offenders for public protection.

Using the model results in an increase from 67.8% to 71.1% in correctly identified offenders when compared to the situation where only the base rate can be used (see Table 5). Although this seems only a small increase in predictive accuracy, its clinical utility can be of value. The primary advantage of using the model is that it facilitates the selection of groups that contain a majority of future violators, while the total group consists of a minority of violators. Using the model enables a more efficient use of available risk reduction resources, which will be demonstrated in the following *example* that uses the sum score model presented in Table 5. In our example, we assume (a) that intensive supervision (instead of less intensive supervision) highly reduces the risk of probation violation, (b) that intensive supervision costs \$ 3,000 more per year than less intensive supervision per probationer, and (c) that the available budget for intensive supervision (per year) is \$ 324,000. With this budget, intensive supervision can be provided to 108 offenders. When the group of probationers with a sum score of 25 and higher is assigned to intensive supervision, 108 probationers will be eligible, that is, the number of correct positives ($n = 75$ or 2.5%) plus the number of false positives ($n = 33$ or 1.1%). In this group, for 69.4% ($75/108 \times 100$) of the probationers intensive supervision would be justified, because they would have violated probation supervision. If 108 offenders would have been assigned *randomly* to intensive supervision (because no instrument or knowledge was available), intensive supervision would be justified for only 35 probationers ($32.2/100 \times 108$; only 32.2% of the probationers would have violated the probation requirements). In other words, given the same budget, using the model of 16 RISC items can *double* correct assignments to intensive supervision, because an additional 40 (75 minus 35) persons could be assigned correctly. Accuracy in risk assessment can thus play a major role in the selection of groups of offenders that are at high risk of supervision violation and thus in preventing risks during probation.

The present study has several strengths and limitations. First of all, the large sample ($N = 14,363$) is considered a strength. Two important aspects of the large sample size are that (a) the sample represents a large part of the population of probationers in a 3.5 year period and (b) it results in large statistical power which enables the detection of small but important effects. A second important strength of the study is that all included RISC assessments were completed before the start of the supervision periods, which rules out the possibility that probation workers and consequently, the RISC assessments were influenced by knowledge of success or failure of the probation period.

With regard to the limitations of the present study, Menard (1995, p. 54), describing the debate about the use of stepwise techniques, points out that some authors would consider the use of

a stepwise procedure as inappropriate. Generally, it is agreed that stepwise procedures should not be used for the testing of theories, because it capitalizes on random variations and results may be difficult to replicate in other samples. It should be emphasized, however, that the risk of capitalization on random variations was reduced in the present study by calibrating and testing the model on two different random samples that were selected from the available research group. Interestingly, the model that was selected in the calibration sample fitted the data in the test sample as was shown by the Hosmer-Lemeshow goodness-of-fit statistic. However, it should be noted that the stepwise procedure in the present study was used as an explorative tool, and *not* for theory testing. Stepwise regression is considered a useful technique when the outcome studied is relatively new (Hosmer & Lemeshow, 2000, p. 116), which is the case for probation violation which is not yet studied on a large scale, especially not in Europe. One of the goals of the present study was to explore the potential to predict probation violation using variables that are part of the Dutch diagnostic and risk assessment tool RISC, which is widely used in The Netherlands. Because the stepwise procedure selected items from almost all subscales of the RISC, it seems that most factors that are important for predicting general recidivism are also important in predicting probation violation. However, the finding that the degree of endorsement of item 10.2 (Mental health problems) had a contrary effect on violation risk (only) in the presence of the other variables in the model was unexpected.

Although in 2007 the tasks of diagnosis (RISC assessments) and supervision are separated by the Dutch Probation Service, and RISC assessments are made by different probation officers than the ones that are responsible for the supervision (they work at different units), the officer who is responsible for the supervision does have knowledge of the RISC scores/outcome. The knowledge of probationers' scores on the measure could have affected the likelihood that the occurrence of violations was influenced in some way, for example by easier or earlier detection or prevention of violation. If so, this could be seen as a limitation of the study.

Another limitation may be the fact that the outcome variable, probation violations, may have different causes that can vary in terms of seriousness. Arguably, some types of probation violation (e.g., reoffending) are more serious than other types (e.g., technical violations) even if both kinds of noncompliance lead to termination of the conditional sentence. It would be interesting to distinguish between probation supervision violations that represent new offenses from those that are technical violations. Unfortunately, the database of the Dutch Probation Service does not contain information on different types of violation; until now, the type of noncompliance is not recorded in the database. Future research may address possible variations in predictive accuracy of different types of violation. It may be even more important to prevent certain types of violation, such as violent recidivism during probation, certainly when prevention resources are scarce. However, this requires collecting information on the type of noncompliance. Hopefully, the results of this research will convince the Dutch Probation Service to make such an effort so that future research can address possible variations in predictive accuracy of different types of violation.

Although the predictive value of the RISC could be equaled with a smaller subset of RISC items, given the results of the stepwise logistic regression procedure, it should be emphasized that the

model and its composition are not ultimate results. For example, in the future predictive accuracy may be improved by writing additional items specifically for the prediction of probation violation. Future research should therefore focus on building a theoretical framework around probation violation, finding competing models that fit the data better, at clarifying the occurrence of possible interactions between variables, and at identifying potential other factors that could further increase predictive accuracy. From a research perspective, it would also be interesting to compare the predictive accuracy from the RISC with other RNR instruments, such as the LS/CMI and OASys, when it comes to supervision violation.

In sum, one of the primary missions of probation, especially as the use of the sanction increased, is to protect the public (Petersilia, 1985a; Petersilia, 1985b). For supervision to be effective, it is important that the risk of noncompliance is taken into account. The findings reported in this study reveal that the RISC is a valuable tool to predict probation supervision violations of male probationers. For one of the possible cut-off scores used to select groups at high risk for probation violation, it was shown that is possible to double the percentage of correctly identified future violators when compared to the base rate of probation violation. We hope that the identification of factors associated with supervision violation leads to a more effective supervision and prevention tactics.

References

- Adviesbureau Van Montfoort & Reclassering Nederland. (2004). *RISC versie 1.0. Recidive Inschattings Schalen. Handleiding*. [RISC version 1.0. Recidivism Assessment Scales. Manual]. Utrecht, the Netherlands: Reclassering Nederland.
- Allison, P. D. (2001). *Missing data. Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-136*. Thousand Oaks, CA: Sage.
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19–52. doi:10.1177/0093854890017001004
- Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (5th ed.). Cincinnati, OH: Anderson.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*, 39–55. doi:10.1037/a0018362
- Andrews, D. A., & Bonta, J. L. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28*, 369–404. doi:10.1111/j.1745-9125.1990.tb01330.x
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*, 40–49. doi:10.1002/mpr.329
- Bland, J. M., & Altman, D. G. (1997). Statistic notes: Cronbach's alpha. *British Medical Journal, 314*, 572. doi:10.1136/bmj.314.7080.572
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123–142. doi:10.1037/0033-2909.123.2.123
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. I. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth.
- Chen, X., Ender, P., Mitchell, M., & Wells, C. (n.d.). Additional coding systems for categorical variables in regression analysis. In *Regression with SPSS (chap. 5)*. Retrieved November 25, 2011, from <http://www.ats.ucla.edu/stat/spss/webbooks/reg/chapter5/spssreg5.htm>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319. doi:10.1037/1040-3590.7.3.309
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal re-offense in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry, 29*, 431–442. doi:10.1016/j.ijlp.2006.03.001
- Douglas, K. S., Skeem, J. L., & Nicholson, E. (2011). Research methods in violence risk assessment. In B. Rosenfeld & S. Penrod (Eds.), *Research methods in forensic psychology* (pp. 325–346). Hoboken, NJ: Wiley.
- Everitt, B. S. (1977). *The analysis of contingency tables*. New York, NY: Wiley.
- Gendreau, P. P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offenders: What works! *Criminology, 34*, 575–607. doi:10.1111/j.1745-9125.1996.tb01220.x
- Gibbons, J. D. (1985). *Nonparametric statistical inference*. New York, NY: Dekker.
- Gray, M. K., Fields, M., & Maxwell, S. R. (2001). Examining probation violations: Who, what, and when. *Crime & Delinquency, 47*, 537–557. doi:10.1177/0011128701047004003
- Home Office. (2002). *Offender Assessment System: OASys: User manual*. London, UK: Home Office.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley. doi:10.1002/0471722146
- Howard, P., Clark, D., & Garnham, N. (2006). *An evaluation of the Offender Assessment System (OASys) in three pilots 1999–2001*. London, UK: National Offender Management Service.
- Kramer, A. A., & Zimmerman, J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine, 35*, 2052–2056. doi:10.1097/01.CCM.0000275267.64078.B0
- Menard, S. (1995). *Applied logistic regression analysis. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-106*. Thousand Oaks, CA: Sage.
- Ministry of Justice (2009). *A compendium of research and analysis on the Offender Assessment System (OASys) 2006–2009*. London, UK: Ministry of Justice.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783–792. doi:10.1037/0022-006X.62.4.783
- Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.
- Murphy, K., & Myers, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Erlbaum.
- Ogloff, J. R. P., & Davis, M. R. (2004). Advances in offender assessment and rehabilitation: Contributions of the risk-needs-responsivity approach. *Psychology, Crime & Law, 10*, 229–242. doi:10.1080/10683160410001662735
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology, 79*, 6–21. doi:10.1037/a0022200
- Petersilia, J. (1985a). Community supervision: Trends and critical issues. *Crime and Delinquency, 31*, 339–347. doi:10.1177/0011128785031003001
- Petersilia, J. (1985b). *Probation and felony offenders: Research in brief*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association. doi:10.1037/10304-000
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing pre-

dictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737–748. doi:10.1037/0022-006X.63.5.737

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. doi:10.1037/1082-989X.7.2.147

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350–353. doi:10.1037/1040-3590.8.4.350

van der Knaap, L. M., & Alberda, D. L. (2009). *De predictieve validiteit van de Recidive Inschattingsschalen (RISc)*. [Predictive validity of the Recidivism Risk Assessment Scales]. Den Haag, the Netherlands: WODC.

van der Knaap, L. M., Leenarts, L. E. W., Born, M., & Oosterveld, P. (2012). Reevaluating interrater reliability in offender risk assessment. *Crime & Delinquency*, 58, 147–163. Online first publication October 1, 2010. doi:10.1177/0011128710382347

van der Knaap, L. M., Leenarts, L. E. W., & Nijssen, L. T. J. (2007). *Psychometrische kwaliteiten van de Recidive Inschattingsschalen (RISc)*. *Interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit*. [Psychometric qualities of the Recidivism Assessment Scales (RISc). Interrater reliability, internal consistency, and congruent validity]. Den Haag, the Netherlands: WODC.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30, 377–399. doi:10.1002/sim.4067

Wong, S. C. P., Gordon, A., & Gu, D. (2007). Assessment and treatment of violence prone forensic clients: An integrated approach. *British Journal of Psychiatry*, 190, s66–s74. doi:10.1192/bjp.190.5.s66

Wood, A. M., White, I. R., & Royston, P. (2008). How should variable selection be performed with multiple imputed data? *Statistics in Medicine*, 27, 3227–3246. doi:10.1002/sim.3177

Appendix

Sample Items of the RISc

	RISc scale	Sample items
1–2	Prior and Current Offenses	<ul style="list-style-type: none"> - Number of convictions as a juvenile - Number of convictions as an adult - Severity of current or previous charges - Previous noncompliance with probation conditions - Over time, the offender’s criminal behavior is getting more and more serious
3	Accommodation	<ul style="list-style-type: none"> - Accommodation track record (whether there have been periods of homelessness, etc.) - Current housing - Suitability and permanency of current housing
4	Education and Employment	<ul style="list-style-type: none"> - Level of training and certificates obtained - Work experience and employment track record - Current work situation
5	Income and Finances	<ul style="list-style-type: none"> - Main source of income - Current financial situation - Gambling addiction or other addiction (that eats into the main source of income)
6	Relationships (Partner, Family, Relatives)	<ul style="list-style-type: none"> - Quality of current relationship with partner, family, and other relatives - Family member has criminal record - History of domestic violence
7	Peer Relationships	<ul style="list-style-type: none"> - Quality of relationship with friends and acquaintances - Manipulates friends and acquaintances - Sensation and thrill seeking, likes to take risks
8	Drug Use/Abuse	<ul style="list-style-type: none"> - Frequency of drug use in the past - Drugs are at the forefront in the person’s life - Criminal behavior and drug use are linked
9	Alcohol Use/Abuse	<ul style="list-style-type: none"> - Excessive alcohol use in the past - Current alcohol use is problematic - Criminal behavior and alcohol use are linked
10	Emotional Wellbeing	<ul style="list-style-type: none"> - Struggles to survive - Mental health problems - Self-destructive behavior
11	Thought Pattern, Behavior, Social Skills	<ul style="list-style-type: none"> - Impulsivity - Dominant behavior - Problem handling
12	Attitude	<ul style="list-style-type: none"> - Procriminal attitude - Attitude toward sanction - Willingness to change

Note. RISc = Recidivism Assessment Scale. For each RISc scale, three sample items are given. For Scales 1 and 2 five examples are given since the (scores on the) items from these two scales are combined into one scale.

Received November 30, 2011
 Revision received February 28, 2012
 Accepted March 2, 2012 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.