UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The development and validation of an actuarial risk assessment tool for the prediction of first-time offending

Assink, M.; van der Put, C.E.; Stams, G.J.J.M.

DOI
10.1177/0306624X14558204

Publication date
2016

Document Version
Final published version

Published in
International Journal of Offender Therapy and Comparative Criminology

Link to publication

Citation for published version (APA):
Assink, M., van der Put, C. E., & Stams, G. J. J. M. (2016). The development and validation of an actuarial risk assessment tool for the prediction of first-time offending. *International Journal of Offender Therapy and Comparative Criminology*, *60*(7), 847-864. https://doi.org/10.1177/0306624X14558204

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:26 Jul 2022

International Journal of
Offender Therapy and
Comparative Criminology
1–18
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0306624X14558204
ijo.sagepub.com

The Development and Validation of an Actuarial Risk Assessment Tool for the Prediction of First-Time Offending

Mark Assink[1], Claudia E. van der Put[1], and Geert Jan J. M. Stams[1]

Abstract

For prevention purposes, it is important that police officers can estimate the risk for delinquency among juveniles who were involved in a criminal offense, but not in the role of a suspect. In the present study, the Youth Actuarial Risk Assessment Tool for First-Time Offending (Y-ARAT-FO) was developed based solely on police records with the aim to enable Dutch police officers to predict the risk for first-time offending. For the construction of this initial screening instrument, an Exhaustive Chi-squared Automatic Interaction Detector (Exhaustive CHAID) analysis was performed on a data set that was retrieved from the Dutch police system. The Y-ARAT-FO was developed on a sample of 1,368 juveniles and validated on a different sample of 886 juveniles showing moderate predictive accuracy in the validation sample (area under the receiver operating characteristic curve [AUC] = .728). The predictive accuracy of the Y-ARAT-FO was considered sufficient to justify its use as an initial screening instrument by the Dutch police.

Keywords

actuarial risk assessment, screening, first-time offending, juvenile delinquency, CHAID analysis

The most important goal of prevention strategies in the area of juvenile delinquency is to prevent onset of delinquent behavior rather than treating juvenile delinquents

[1]University of Amsterdam, The Netherlands

Corresponding Author:
Mark Assink, Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15780, 1001 NG, Amsterdam, The Netherlands.
Email: M.Assink@UvA.nl

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

(DeMatteo & Marczyk, 2005). At least four different reasons can be put forward to the rationale of these prevention strategies. First, juvenile delinquency is a serious societal problem with detrimental physical and mental health effects for both victims and offenders (e.g., Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002; McGuw & Iacono, 2005; Piquero, Daigle, Gibson, Leeper Piquero, & Tibbetts, 2007). Second, a considerable amount of literature shows that signs of chronic antisocial behavior, which are predictive of later delinquency, can already be recognized early in the juvenile's life (Dishion & Patterson, 2006; Loeber, Farrington, Stouthamer-Loeber, & White, 2008) and can therefore be targeted by early prevention efforts. Third, as children and adolescents grow older, problem behavior tends to become more stable, which makes it more difficult to change (Bernazzani, Cothe, & Tremblay, 2001). Fourth, because delinquency is associated with significant monetary costs to society, investing in early prevention efforts can have substantial financial benefits (e.g., Nores, Belfield, Barnett, & Schweinhart, 2005).

Prevention programs are often characterized by broad-based and systemic efforts targeting juveniles at high risk for delinquent behavior. To identify these high risk children and adolescents, it is important that valid and reliable screening instruments be available, that can predict the likelihood of becoming a delinquent. To date, numerous risk assessment tools have been developed and reported on in scientific literature, but most of these instruments were designed for estimating the likelihood of recidivism among juveniles who already have committed one or more offenses. To our knowledge, no valid and reliable instrument is yet available for predicting the onset of general delinquency among juveniles. Therefore, the current study describes the development of a risk screening instrument for predicting the risk for onset of general delinquency among juveniles. The psychometric quality of this instrument will be considered by examining its predictive validity.

In the Netherlands, the police is an important link in the chain of youth care because police officers do not only deal with juvenile offenders, but also with a substantial number of juvenile non-offenders who are in some way involved in a criminal offense, but not in the role of a suspect. For example, from a large sample of juveniles who came into contact with the Dutch police in 2007 (N = 9,531), we could derive that 66.4% (n = 6,331) of these juveniles was an offender (i.e., having the role of a suspect) and 33.6% (n = 3,200) of these juveniles was a non-offender (i.e., having any role other than that of a suspect). Although the juveniles in the latter group have never been recorded by the Dutch police as being suspected of an offense, they may be at risk for future delinquency. Previous research has shown that juveniles involved in a criminal offense as a witness or a victim are at elevated risk for delinquency (e.g., Hurt, Malmud, Brodsky, & Giannetta, 2001; Loeber, Kalb, & Huizinga, 2001; Patchin, Huebner, McCluskey, Varano, & Bynum, 2006). These juveniles should have the opportunity to benefit from preventive strategies that increase their well-being in the long term by preventing the onset of delinquency.

For this preventive effort to be effective, it is essential that the police be able to identify which juvenile non-offenders are at high risk for delinquency, so that they can be timely referred to other specialized agencies in the chain of youth care for a more

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

thorough assessment and, if necessary, treatment to counteract the risk for first-time offending. Hence, the aim of the present study was to examine whether a risk screening instrument based on official police records could be developed, enabling Dutch police officers who lack significant clinical experience to make an initial screening of the risk for onset of general delinquency among juvenile non-offenders. In the development of this instrument, we only used information available in operational police systems so that an automatic assessment process would be possible within the limited time and resources available to Dutch police officers.

Today, different methods are being used in the field of forensic risk assessment. One of the dominant contemporary approaches to risk assessment is actuarial prediction (Singh, 2012). Actuarial (or statistical) risk assessment instruments estimate the likelihood of future delinquency through the assignment of numerical values to factors that are empirically associated with delinquency. A statistical algorithm is then used to calculate a probabilistic estimate of future delinquency from a total test score. In this actuarial approach, each individual is appraised using the same criteria so that individuals can be directly compared with others who have had the same tool administered regardless of who conducted the assessment (Singh, Grann, & Fazel, 2011).

A large body of literature showed that the actuarial method for risk assessment performs as well as clinical methods or is even more accurate (e.g., Aegisdottir et al., 2006; Dawes, Faust, & Meehl, 1989; Fazel, Singh, Doll, & Grann, 2012; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Meehl, 1954). However, actuarial models have been criticized, for instance, because purely actuarial predictions do not sufficiently take into account individual differences (Sreenivasan, Kirkish, Garrick, Weinberger, & Phenix, 2000). In addition, Scott and Resnick (2006) stated that actuarial instruments are rigid, lacking sensitivity to change and cannot be generalized to populations other than the sample that was used to construct the instrument. Furthermore, it should be noted that actuarial methods are less suitable for making decisions about effective treatment.

Despite these critical comments, an actuarial approach to risk assessment seemed most appropriate to us in the present study, for several reasons. First, Dutch police officers will only conduct an *initial* screening of risk with the aim to identify those juveniles who are at high risk for onset of delinquency and who are in need of further assessment by a more specialized agency in the chain of youth care. Second, an actuarial approach to risk screening reduces unwarranted disparities in decision making (e.g., racial and gender biases) relative to professional judgment approaches (Young, Moline, Farrell, & Bierie, 2006). Third, an actuarial risk assessment instrument was preferred above other instruments because of the ease of use, the relative low costs for training and materials, and the possibility of a rapid risk screening. These are important issues, because Dutch police officers do not have the time, resources, and expertise to engage in thorough clinical assessment, meaning that only information on juveniles derived from the Dutch police system is available to police officers. Finally, previous research suggests that developing a risk screening tool using only data derived from a police system can be a fruitful approach. For an elaborate discussion on the potential of data mining in the criminal justice context, see the work of Berk (2012).

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

In sum, the aim of the present study was to develop a valid and reliable actuarial risk screening instrument for predicting the risk for onset of general delinquency among juvenile non-offenders. This instrument will be further referred to as the Youth Actuarial Risk Assessment Tool for First-Time Offending (Y-ARAT-FO). In developing this instrument, the same procedure was used as in the development of the Youth Actuarial Risk Assessment Tool (Y-ARAT), which is a risk screening instrument for the prediction of general offense recidivism among juvenile offenders and which was also solely based on Dutch police records (van der Put, 2013). The construction and validation of the Y-ARAT-FO comprised several steps. First, we examined the extent to which police records were related to delinquency. Second, we examined whether an actuarial risk screening instrument could be developed with a sufficiently high predictive value, using only police records with a history of 10 years. Third, we examined whether the instrument was suitable for predicting specific types of delinquent behavior.

Method

Sample

The sample consisted of 2,254 juveniles between the ages of 12 and 18 years (M = 15.5, SD = 1.7), who were registered in official Dutch police records in 2007 because they were involved in an offense, but not in the role of a suspect. These juveniles were selected at random from all juveniles who came into contact with the Dutch police in 2007 in the police regions "Hollands-Midden" and "Rotterdam-Rijnmond." To construct and validate the model, the sample was split randomly into a construction sample (60%, n = 1,368) and a validation sample (40%, n = 886). The size of the full sample (N = 2,254) was sufficiently large for splitting this sample into a construction and validation sample and to perform an Exhaustive Chi-squared Automatic Interaction Detector (Exhaustive CHAID) analysis in which the total group is divided into a number of smaller subgroups. No significant differences were found between the construction and validation sample in terms of gender ($\chi^2(1) = .035$, $p = .829$), country of birth ($\chi^2(1) = .000$, $p = 1.000$), and age ($t(2,252) = .514$, $p = .607$).

Data Collection

The random sample of juveniles registered in official police records was drawn from two regional computer systems of the Dutch police. Juveniles who were involved in an incident that took place in 2007 were selected, meaning that the incident in 2007 was taken as the index incident. The juveniles involved in these incidents had any role other than that of a suspect (i.e., victim, witness, reporter of an offense, missing person, a juvenile attracting police attention, or a juvenile having any role not otherwise defined by the Dutch police) and had never been recorded by the police as a suspect of an offense. The records of these juveniles were retrieved from the police system for a period of 10 years prior to the date on which the index incident took place (i.e., from

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

1997 to 2007). Data about delinquent behavior were also retrieved from the system in the form of official records. Onset of delinquency was defined as a juvenile being suspected by the police of committing an offense within a period of 3 years after the index incident took place (i.e., from 2007 to 2010). The juvenile was recorded in the police registration system as having the role of a suspect in these offenses. In the Netherlands, a juvenile being registered as a suspect means either that the juvenile was caught by the police in the act of committing an offense (after which the juvenile was arrested) or that a juvenile was summoned to the police station because the police was convinced that the juvenile had committed an offense. In both situations, there is a temporal deprivation of liberty for the juvenile.

Prior to the construction of an assessment instrument, it is important to determine which predictor variables should be included in the analysis. The designated predictor variables were based on both information from literature reviews of risk factors for the onset of juvenile delinquency (e.g., Loeber, 1990; Smith, 2004; Youth Justice Board, 2005) as well as expertise of the Dutch police gained in the development of ProKid, which is a risk screening instrument for juvenile offenders younger than 12 years of age (Abraham, Buysse, Loef, & Van Dijk, 2011). Information on police records of co-occupants at the juvenile's living address was also retrieved. Table 1 presents the predictor variables that were established from the police records.

Analyses

We developed the risk screening instrument by conducting a tree classification method designated as Exhaustive CHAID analysis. Tree classification methods such as Chi-squared Automatic Interaction Detector (CHAID) are very useful for gaining insight into profiles of youth with a high and low probability for delinquency (Steadman et al., 2000; Thomas & Leese, 2003). The purpose of CHAID is to create homogeneous groups based on the value of a specific outcome variable (in the present study, onset of delinquent behavior among juveniles) by splitting cases into two or more groups on the basis of several predictor variables (Biggs, De Ville, & Suen, 1991; Kass, 1980). This technique is particularly useful if a study is exploratory rather than confirmatory, if the analysis involves relations between a number of independent variables and a single dependent variable, if these independent variables interact with each other, and if there is no strong theory about the relative importance of the independent variables in predicting the dependent variable (Boslaugh, Kreuter, Nicholson, & Naleid, 2005). A limitation of this analysis is that the time period within which a juvenile commits the first offense cannot be taken into account. However, we preferred CHAID analysis above logistic regression because the results are visually presented and therefore easily interpretable, which is of high importance for practical use by, for instance, police officers or youth care workers without substantial clinical expertise. Although CHAID and Exhaustive CHAID are very similar algorithms, the latter was preferred in the current study because Exhaustive CHAID performs a more thorough merging and testing of predictor variables than the regular CHAID algorithm (Biggs et al., 1991).

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Table 1. Background Characteristics and Types of Police Records: Descriptives and Association With Delinquent Behavior (Total Sample; N = 2,254).

Categorical independent variables	*M*	*SD*	Range	ϕ
Male (0 = no; 1 = yes)	0.547	0.498	0-1	.173***
Born outside the Netherlands (0 = no; 1 = yes)	0.069	0.253	0-1	.041

Continuous independent variables	*M*	*SD*	Range	r_b
Current age	15.499	1.660	12-18	−.155***
Age at first incident (all roles other than suspect)	14.109	1.986	4-17	**−.222*****
Number of incidents (all roles other than suspect)	2.177	2.018	1-25	**.256*****
Number of incidents (involved as victim)	0.550	0.799	0-7	−.039
Number of incidents (involved as witness)	0.191	0.472	0-4	−.121***
Number of incidents (involved as witness of violence)	0.067	0.277	0-3	−.033
Number of incidents (involved as aggrieved person or reporter of an offense)	0.208	0.490	0-4	−.012
Number of incidents (recorded by the police, not having a specific role)	0.021	0.155	0-2	.089**
Number of incidents (involved in all roles other than suspect), type of incident:				
Non-violent property offense	0.311	0.562	0-4	−.020
Violent property offense	0.043	0.215	0-2	.057
Public order offense without violence	0.559	1.034	0-13	**.274*****
Public order offense with violence	0.039	0.221	0-2	−.015
Sex offense without violence	0.029	0.185	0-2	.017
Sex offense with violence	0.042	0.229	0-3	−.032
Other offense without violence	0.921	1.279	0-21	.160***
Other violent offense	0.235	0.559	0-6	.066*
Number of incidents in which weapons were involved at the juvenile's living address (the juvenile does not need to be involved in this incident)	0.012	0.167	0-3	−.036
Number of incidents involving domestic violence at the juvenile's living address (the juvenile does not need to be involved in this incident)	0.096	0.610	0-9	−.042
Number of incidents of sexual offenses at the juvenile's living address (the juvenile does not need to be involved in this incident)	0.087	0.949	0-17	−.032
Number of incidents of child abuse at the juvenile's living address (the juvenile does not need to be involved in this incident)	0.004	0.067	0-2	−.008
Number of incidents in which a co-occupant at the juvenile's living address was a suspect	4.169	27.581	0-481	.077*
Number of incidents of child abuse in which a co-occupant at the juvenile's living address was involved (in any role)	0.026	0.275	0-7	.026
Number of incidents of neglect in which the juvenile and/or a co-occupant at the juvenile's living address was a victim	0	0	0-0	—[a]
Number of incidents of conflicts in which a co-occupant at the juvenile's living address was a victim	0.342	3.443	0-122	.039
Number of incidents of domestic strife in which the juvenile and/or a co-occupant at the juvenile's living address was a victim	0.051	0.529	0-21	.056

Note. Delinquent behavior was defined as a juvenile being suspected by the police of committing an offense within a period of 3 years after the index incident took place. Correlations ≥.20 and ≤−.20 are in boldface to highlight the strongest associations. ϕ = phi-coefficient; r_b = biserial correlation.
[a]The biserial correlation could not be calculated, because this variable was a constant.
*p < .05. **p < .01. ***p < .001.

The CHAID algorithm as applied in the current study involved dividing the total group of juveniles into a number of subgroups on the basis of the independent variables most strongly associated with delinquent behavior. The procedure for this

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

analysis comprises a number of steps in which the first step is to divide the total group into a number of subgroups on the basis of the variable most strongly associated with delinquent behavior. In the next step, the subgroups are split again on the basis of the variable that is second most strongly associated with delinquent behavior. This stepwise procedure is repeated until there are either no more variables that have a significant association with delinquent behavior or until the subgroups have reached a minimum size (n = 25 in the current study). The result is a visual tree model in which each terminal node represents a "risk group" in which juveniles have similar police records and thus a similar risk for committing a first offense. When interpreting the tree model from top to bottom, one can identify how the juveniles in each risk group score on the predictor variables that are part of the instrument. To build the CHAID tree model, the total group of juveniles was randomly divided into two groups; about 60% of the sample was used to build the model (construction sample) and about 40% of the sample was used to validate the model (validation sample).

The sensitivity, specificity, false positive and false negative rates, positive and negative predictive value, overall accuracy, and diagnostic odds ratios were examined in the validation sample and at different cutoff scores to assess the predictive validity of the instrument. Several researchers recommend reporting area under the receiver operating characteristic curve (AUC) values as the preferred statistic of predictive or diagnostic accuracy (e.g., Mossman, 1994; Rice & Harris, 2005), and therefore, AUC values were also calculated in the present study. Moreover, by reporting AUC values, the results of our study can be more easily compared with results of other risk assessment studies.

Results

Prevalence of Offenses

An overview of the prevalence of offenses within the 3-year period after the index incident can be found in Table 2 for both the construction and validation sample. In total, the percentage of juveniles that started offending was 15.79% (n = 216) and 15.91% (n = 141), respectively. No statistically significant differences were found in the prevalence of (different types of) offenses between the construction and validation sample.

Risk Factors for Delinquency

Table 1 shows the association between the variables that were retrieved from the police system and delinquent behavior. The variables most strongly associated with delinquent behavior were (a) age at first recorded incident in which the juvenile was involved (not in the role of a suspect), (b) total number of recorded incidents in which the juvenile was involved (not in the role of a suspect), and (c) number of recorded public order offenses without violence in which the juvenile was involved (not in the role of a suspect).

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Table 2. Prevalence of Different Offense Types Within 3 Years After the Index Offense.

	Total sample (*N* = 2,254)		Construction sample (*n* = 1,368)		Validation sample (*n* = 886)		
	n[b]	%	*n*[b]	%	*n*[b]	%	$\chi^2(1)$[a]
Total number of juveniles that committed an offense	357	15.84	216	15.79	141	15.91	.000
Specific type of offense:							
Violent offense	59	2.60	34	2.50	25	2.80	.125
Property offense[c]	203	9.01	127	9.28	76	8.58	.246
Public order offense[c]	148	6.57	84	6.14	64	7.22	.859
Sexual offense[c]	4	0.18	2	0.15	2	0.23	.000
Other offense[c]	198	8.78	118	8.63	80	9.02	.065

[a]Chi-square tests with Yates' correction were conducted to determine significant differences between the construction and validation sample. No significant chi-square test results were found.
[b]Number of unique juveniles in the sample that committed the offense.
[c]Either violent or non-violent in nature.

Development of the Y-ARAT-FO

To develop the Y-ARAT-FO, we conducted an Exhaustive CHAID analysis. The variables from Table 1 were included as independent variables, and the dependent variable was whether or not juveniles were registered in the Dutch police system as a suspect in the 3-year period after the index incident. Figure 1 presents the CHAID output for the validation sample. To validate the CHAID model, the tree nodes that were derived by the CHAID algorithm in the construction sample were specified a priori in the validation sample, rather than allowing the algorithm to run new analyses using the same variables. When the CHAID algorithm ended, the following five variables were part of the CHAID model: (a) the total number of previously recorded incidents in which the juvenile had any role other than that of a suspect, (b) the gender of the juvenile, (c) the total number of recorded incidents in which a co-occupant at the juvenile's living address was a suspect, (d) whether or not the juvenile was born in the Netherlands, and (e) the current age of the juvenile. As can be seen in Figure 1, the tree classification diagram consists of 10 end nodes that represent 10 different "risk groups". In each risk group, juveniles have similar scores on the variables that comprise the Y-ARAT-FO and thus a similar risk for committing a first offense. The risk for delinquent behavior ranged from .05 in the lowest risk group to .43 in the highest risk group, meaning that 5% of the juveniles in the lowest risk group and 43% of the juveniles in the highest risk group will become an offender. For each new case that Dutch police officers are confronted with, the predicted probability of offending equals the proportion of offenders in one of the risk groups (i.e., one of the gray-shaded end nodes in Figure 1), into which each case is classified depending on the sequence of the variables that are part of the tree classification diagram.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

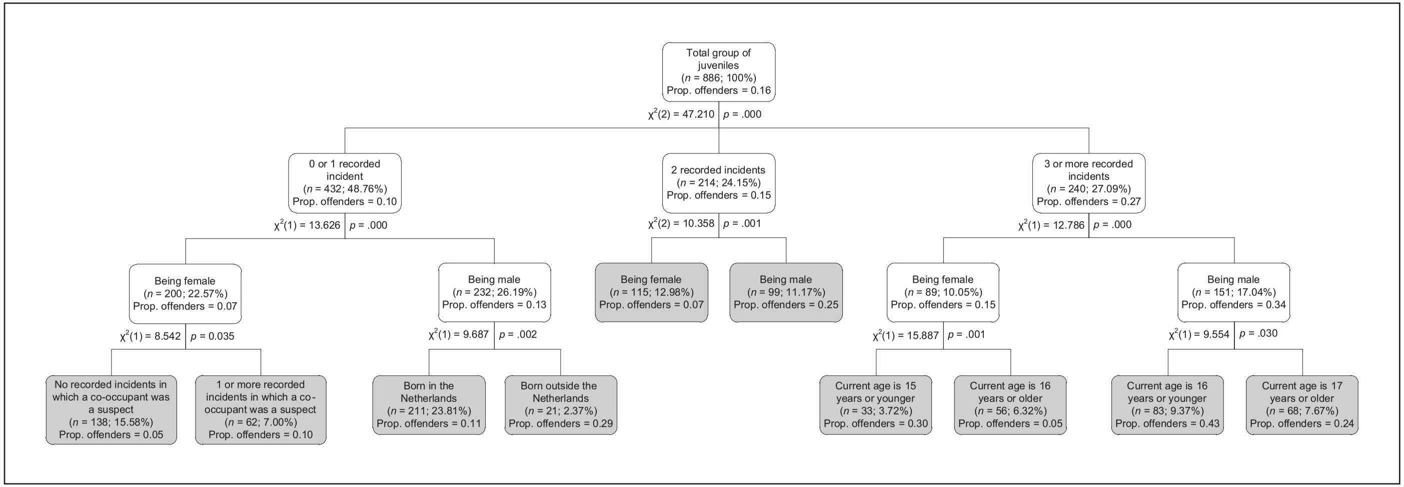

Figure 1. Classification tree for the validation sample.

Note. The 10 gray-shaded terminal nodes represent the "risk groups" in which juveniles have similar scores on the variables that comprise the Y-ARAT-FO, and thus a similar risk for committing a first offense. Y-ARAT-FO = Youth Actuarial Risk Assessment Tool for First-Time Offending; *n* (%) = number of juveniles that are classified in each tree node and the percentage of the total group of juveniles; Prop. offenders = proportion of offending juveniles in each tree node.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Predictive Power of the Y-ARAT-FO

To assess the predictive power of the Y-ARAT-FO, AUC values were calculated for both the construction and validation sample. The AUC values were .724 (p < .001, 95% confidence interval [CI] = [.700, .748]) in the construction sample and .728 (p < .001, 95% CI = [.697, .757]) in the validation sample, and were not statistically different (Z = .004, p = .903). The sensitivity, specificity, false positive and false negative rates, predictive values, overall accuracy, and diagnostic odds ratio of the Y-ARAT-FO are presented in Table 3. These test performance outcomes were all calculated for the performance of the Y-ARAT-FO in the validation sample. Because these measures are dependent on a cutoff score above which the test result of the Y-ARAT-FO is considered positive, the test performance outcomes were calculated for different cutoff scores. These cutoff scores equal the risk for offending behavior of the different "risk groups" that were produced by the CHAID analysis in the construction sample (i.e., the proportion of juveniles that start offending of every end node in the tree diagram that was produced in the construction sample). The sensitivity of the test refers to the proportion of juveniles who start offending and test positive, while the specificity refers to the proportion of juveniles who will not start offending and test negative. Ideally, a test should have high sensitivity and high specificity. However, sensitivity and specificity are inversely related, meaning that when sensitivity is high, specificity is low, and vice versa.

In Table 3, the false positive rate is the chance of testing positive among juveniles who do not start offending, and the false negative rate is the chance of testing negative among juveniles who do start offending. The positive predictive value refers to the probability that a juvenile scoring above a particular cutoff score is correctly identified as someone who will start offending, whereas the negative predictive value refers to the probability that a juvenile scoring below a particular cutoff score is correctly identified as someone who will not start offending. The overall accuracy refers to the proportion of offending and non-offending juveniles correctly classified by the instrument. Finally, the diagnostic odds ratio is a measure for the discriminative power of the test, which is base rate resistant and refers to the ratio of a true positive result relative to the odds of a false positive result. A high diagnostic odds ratio indicates a good test performance, and unlike the AUC value, it can take into account different cutoff values.

Predictive Power of the Y-ARAT-FO for Predicting Specific Types of Offenses

Finally, it was examined whether the Y-ARAT-FO can be used for the prediction of specific types of delinquent behavior (see Table 4). The Y-ARAT-FO showed an acceptable predictive accuracy for the prediction of violent (AUC = .702), property (AUC = .737), public order (AUC = .751), and other offenses (AUC = .731) in the validation sample. A somewhat less optimal predictive accuracy was found for the prediction of property offenses in the construction sample (AUC = .682), but this did

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Table 3. Estimates (and 95% CI) of Performance Measures for Different Cutoff Scores of the Y-ARAT-FO.

Cutoff score (>)[a]	Sensitivity (95% CI)	Specificity (95% CI)	False positive rate (95% CI)	False negative rate (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Overall accuracy (95% CI)	Diagnostic odds ratio (95% CI)
.032	.950 [.900, .980]	.176 [.149, .205]	.824 [.797, .852]	.050 [.014, .086]	.179 [.152, .209]	.949 [.898, .979]	.299 [.269, .329]	4.084 [1.867, 8.936]
.080	.894 [.831, .939]	.320 [.286, .354]	.681 [.647, .714]	.106 [.056, .144]	.199 [.169, .232]	.941 [.904, .966]	.411 [.378, .443]	3.943 [2.259, 6.883]
.089	.872 [.806, .923]	.391 [.355, .427]	.609 [.574, .644]	.128 [.073, .165]	.213 [.180, .249]	.942 [.909, .965]	.467 [.434, .500]	4.380 [2.614, 7.338]
.112	.830 [.757, .888]	.466 [.429, .502]	.534 [.498, .570]	.170 [.108, .209]	.227 [.192, .266]	.935 [.905, .958]	.524 [.491, .557]	4.250 [2.677, 6.748]
.124	.660 [.575, .737]	.717 [.683, .749]	.283 [.251, .316]	.340 [.262, .379]	.306 [.255, .361]	.918 [.892, .939]	.708 [.678, .738]	4.903 [3.343, 7.192]
.213	.482 [.397, .568]	.816 [.786, .843]	.184 [.156, .212]	.518 [.435, .555]	.332 [.268, .401]	.893 [.867, .915]	.763 [.735, .791]	4.134 [2.831, 6.038]
.225	.369 [.289, .454]	.886 [.861, .908]	.114 [.091, .137]	.631 [.552, .666]	.380 [.298, .467]	.881 [.856, .903]	.804 [.778, .830]	4.537 [3.011, 6.836]
.355	.326 [.250, .410]	.906 [.883, .926]	.094 [.073, .115]	.674 [.596, .707]	.397 [.307, .492]	.877 [.851, .899]	.814 [.788, .839]	4.669 [3.039, 7.175]
.370	.255 [.186, .336]	.937 [.917, .953]	.063 [.046, .081]	.745 [.673, .775]	.434 [.325, .548]	.869 [.844, .892]	.828 [.804, .853]	5.092 [3.150, 8.230]

Note. The presented measures are pertaining to the performance of the Y-ARAT-FO in the validation sample. Y-ARAT-FO = Youth Actuarial Risk Assessment Tool for First-Time Offending; 95% CI = 95% confidence interval.

[a]If a test score on the Y-ARAT-FO (i.e., the probability of committing a first offense) is greater than the cutoff score, the test result is considered positive; otherwise it is considered negative.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Table 4. AUC Values (and 95% CI) for the Prediction of Specific Types of Offenses.

	Construction sample (n = 1,368)	Validation sample (n = 886)	
Specific type of offense	AUC (95% CI)	AUC (95% CI)	Z^a
Violent	.779 [.756, .801]***	.702 [.671, .732]***	−1.015
Property[b]	.682 [.656, .706]***	.737 [.707, .766]***	1.283
Public order[b]	.735 [.711, .759]***	.751 [.721, .779]***	0.333
Other offenses[b]	.751 [.727, .774]***	.731 [.701, .760]***	−.471

Note. AUC = area under the receiver operating characteristic curve; 95% CI = 95% confidence interval.
[a]Hanley–McNeil tests were conducted to determine significant differences between AUC values in the construction and validation sample. No significant test results were found.
[b]Either violent or non-violent in nature.
***$p < .001$.

not differ significantly from the predictive accuracy for property offenses in the validation sample ($Z = 1.283$, $p = .200$).

Discussion

The aim of the current study was to examine whether a valid and reliable actuarial risk screening instrument could be developed for predicting the risk for onset of delinquent behavior among juvenile non-offenders, which is based solely on information derived from the Dutch police system and can be used by Dutch police officers without substantial clinical experience. The Youth Actuarial Risk Assessment Tool for First-Time Offending (Y-ARAT-FO) was developed by conducting an Exhaustive CHAID analysis, which yielded a risk classification scheme that can be used to classify new cases into 1 of 10 different risk groups from which the predicted probability for delinquent behavior can be derived. The Y-ARAT-FO consists of the following five variables that together determine the risk for onset of delinquent behavior: (a) the total number of previously recorded incidents in which the juvenile had any role other than that of a suspect, (b) the gender of the juvenile, (c) the total number of recorded incidents in which a co-occupant at the juvenile's living address was a suspect, (d) whether or not the juvenile was born in the Netherlands, and (e) the current age of the juvenile. Because the data that are needed as input for the Y-ARAT-FO can be extracted electronically from the computer database of the Dutch police, the risk value can be calculated quickly and consistently for large groups of juveniles without the need to retrieve or examine additional information about these juveniles.

The observed AUC value of the Y-ARAT-FO was .728 in the validation sample and is equivalent to a Cohen's d of .858 (see the formulas of Ruscio, 2008), which can be interpreted as a large effect (Cohen, 1988). This AUC value compares favourably with AUC values of other actuarial risk assessment instruments. For instance, Schwalbe (2007) reported an average AUC value of risk instruments for predicting general

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

recidivism of .640 (SD = 0.042), and Fazel et al. (2012) found a median AUC value of .66 (interquartile range = .58-.67) for instruments assessing the risk for general offending. Because the current AUC value equals a large effect size and meets the lower bound of .70 as reported by Swets (1988) and Hosmer and Lemeshow (2000), we consider the predictive validity of the Y-ARAT-FO as acceptable. In addition, we believe that the Y-ARAT-FO showed an acceptable predictive accuracy for the prediction of violent, property, public order, and other offenses. However, the predictive accuracy for specific types of offenses may be improved by developing separate models, and this is recommended for future studies.

The choice of a suitable cutoff score depends largely on the consequences of false positive and false negative test results. When the aim is to identify the largest proportion of juveniles who are at risk for becoming delinquent, the test should be highly sensitive. However, if further assessment is costly to society or may have unwanted (psychological) side effects, the test should be highly specific. Because sensitivity and specificity are inversely related, a cutoff score should be chosen that best meets the desired ratio between false positive and false negative test results. If the aim is to minimize the total number of erroneous decisions, a cutoff value should be chosen that maximizes both sensitivity and specificity. Erroneous decisions include both false positives (i.e., a juvenile not at risk for delinquency is referred to further assessment) and false negatives (i.e., a juvenile at high risk for delinquency is not referred to further assessment). A true high risk juvenile not referred to further assessment may be a more negative outcome than a low risk juvenile who is referred to further assessment. In the first case, a juvenile in need of further assessment is not referred to a youth worker with clinical experience and/or an appropriate behavioral intervention. The last case has serious implications for both available resources (i.e., time and money) and for juveniles and their parents who will be assessed too often. Actuarial instruments have the advantage that the number of false positives and false negatives can be calculated for different cutoff values so that an informed choice can be made about the appropriate cutoff value that best meets the requirements of the situation in which the instrument is used.

Regarding the Y-ARAT-FO, the cutoff score at which the total number of false results is minimized was .370, with a total of 16.95% false results (calculated as [[false positives + false negatives]] / total number of test results] × 100). At this cutoff score, the percentage of false positives was 56.41 (calculated as [false positives / total number of positives] × 100), and the percentage of false negatives was 13.21 (calculated as [false negatives / total number of negatives] × 100). If a reduction in false negatives is desirable, it is possible to choose a lower cutoff score, but this will increase the percentage of false positives. For instance, a cutoff score of .124 yielded 69.88% false positives, 8.76% false negatives, and 29.02% false decisions in total. Although the percentages of false positives seem high at every cutoff score, the actual number of juveniles that will be referred for further investigation seems acceptable at several cutoff scores. For instance, a cutoff score of .225 yielded 60.43% false positives, but when looking at all juveniles in the sample, only 14.46% tested positive on the Y-ARAT-FO. We therefore believe that the number of false positives produced by the Y-ARAT-FO will not be very problematic for the capacity of youth welfare agencies.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

The Y-ARAT-FO can only be used by Dutch police officers as a preliminary screening instrument in the initial stage of risk assessment. If police officers come into contact with a juvenile non-offender because of involvement in an incident (but not in the role of a suspect), police officers can use the test result of the Y-ARAT-FO to support their decision about referring the juvenile for further assessment. It is important to point out that the Y-ARAT-FO should be used in addition to, and not instead of, the judgment of police officers because the current instrument is not able to perfectly predict future delinquent behavior. There are numbers of false positive and false negative test results at each cutoff score of the instrument, meaning that test results of the Y-ARAT-FO should not be regarded as a gold standard. A possible explanation for these false test results is that the Y-ARAT-FO is constrained by the availability of data that can be retrieved from the Dutch police system, meaning that other variables that have a significant influence on the risk for delinquency (e.g., mental health and cognitive functioning) are not taken into account in calculating the risk for onset of delinquent behavior. In practice, this can mean that when a police officer interviews a juvenile non-offender that tests positive on the Y-ARAT-FO, the optimal decision may be to refrain from referring the juvenile for further assessment if youth care is already involved in the juvenile's life. However, if the test result of the Y-ARAT-FO is negative, but a police officer has suspicions of cognitive or mental health problems, the optimal decision may be to refer the juvenile for further assessment. These examples illustrate that appropriate judgments of Dutch police officers remain important, even when instruments are available that support police officers in their decisions about referring juveniles for further assessment by more specialized youth care agencies.

Regarding the practical use of the Y-ARAT-FO, there is also an important ethical aspect that needs to be mentioned. False positive test results (i.e., low risk juveniles testing positive on the Y-ARAT-FO) may imply a considerable risk of stigmatizing juveniles, leading in turn to an adverse effect on juvenile's well-being and behavior. Therefore, the risk of stigmatizing juveniles should be weighed against the advantages of implementing the Y-ARAT-FO. It is interesting to point out here that false positives in criminal justice settings are often seen as more problematic than false negatives (e.g., because of ethical implications of putting someone in detention unnecessarily), whereas the reverse may be true for predicting the risk of first-time offending. Although a stigma may be associated with referring juveniles based on false positive test results, further assessment is in general not particularly intrusive or harmful. In contrast, false negatives imply that high risk juveniles who need further assessment (and perhaps treatment) will not be referred, which may be regarded as more problematic and unethical. However, it remains an ethical challenge to screen large groups of juvenile non-offenders without stigmatizing them by labelling the juveniles as possible offenders. This ethical consideration also underlines that the Y-ARAT-FO should be used to support a police officer's decision about referral of a juvenile for further assessment, and not as a single criterion. In this manner, possible negative aspects of the practical use of the Y-ARAT-FO can be reduced.

Several limitations of the current study should be noted. First, the sample used in the present study consisted of juveniles who were randomly selected from only two

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

adjacent police regions in the Netherlands. These regions slightly differ from other regions in the Netherlands in terms of ethnicity, cultural background, household income, and number of committed offenses (Centraal Bureau voor de Statistiek [CBS], 2013), which might affect the generalizability of the results to other Dutch police regions. However, the Y-ARAT-FO was based on rather common predictor variables, and it is therefore not to be expected that the current instrument is far less accurate when administered in other police regions. Second, although developing a risk screening tool based solely on police records does have several advantages, it must also be noted that the number of official recorded incidents is an underestimation of the actual base rate. The willingness of victims to report an incident and police discretion (e.g., the decision not to lay charges) are examples of factors that directly contribute to the underreporting of incidents in official records (Alvi, 2012), and this can indirectly influence the validity of an actuarial risk screening tool that is based on official records. Third, data about official documented convictions could not be used in defining the onset of delinquency, because convictions are not registered in the Dutch police system. We therefore measured onset of delinquency by using the proxy of committing an offense as suspected by the Dutch police. Consequently, a problem that may arise in this definition is the conflation of suspected and actual offending, which, in turn, can produce bias in assessing the onset of delinquency. Perhaps a more objective test result would be obtained when the dependent variable of the Y-ARAT-FO was based on an official outcome variable, such as whether or not a juvenile is convicted. However, a drawback of using official convictions is a possible underestimation of the actual delinquent behavior of juveniles, for instance, in juvenile cases where the available evidence does not meet the quality standards for obtaining a criminal conviction.

We believe that forensic risk assessment research is extended by the present study because it addresses risk assessment within early prevention of juvenile delinquency. To our knowledge, the Y-ARAT-FO is the first actuarial instrument that predicts the risk for onset of general delinquent behavior among juveniles. The Y-ARAT-FO can be administered by police officers without a clinical background, and it can easily be implemented in the Dutch police system. The instrument provides a quick, consistent, well interpretable and cost-effective initial screening of the risk for onset of juvenile delinquency. At a broader policy level, the Y-ARAT-FO may facilitate juvenile justice organizations and affiliated child and youth care organizations operating within early prevention in concentrating their time and resources more effectively on juveniles who have the highest risk for becoming a delinquent.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

References

Abraham, M., Buysse, W., Loef, L., & Van Dijk, B. (2011). *Pilots ProKid Signaleringsinstrument 12- geëvalueerd* [Evaluation of the pilots of the ProKid screening instrument]. Den Haag, The Netherlands: DSP groep.

Aegisdottir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . .Cohen, G. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341-382. doi:10.1177/0011000005285875

Alvi, S. (2012). The nature of youth crime. In S. Alvi (Ed.), *Youth criminal justice policy in Canada* (pp. 49-75). New York, NY: Springer. doi:10.1007/978-1-4419-0273-3_3

Berk, R. (2012). *Criminal justice forecasts of risk: A machine learning approach*. New York, NY: Springer. doi:10.1007/978-1-4614-3085-8

Bernazzani, O., Cothe, C., & Tremblay, R. E. (2001). Early parent training to prevent disruptive behaviour problems and delinquency in children. *The Annals of the American Academy of Political and Social Science*, *578*, 90-103. doi:10.1177/000271620157800106

Biggs, P., De Ville, B., & Suen, K. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, *18*, 49-62. doi:10.1080/02664769100000005

Boslaugh, S. E., Kreuter, M. W., Nicholson, R. A., & Naleid, K. (2005). Comparing demographic, health status and psychosocial strategies of audience segmentation to promote physical activity. *Health Education Research*, *20*, 430-438. doi:10.1093/her/cyg138

Centraal Bureau voor de Statistiek. (2013). *StatLine*. Retrieved from http://statline.cbs.nl/statweb/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

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

DeMatteo, D., & Marczyk, G. (2005). Risk factors, protective factors and the prevention of antisocial behaviour among juveniles. In K. Heilbrun, N. E. S. Goldstein, & R. E. Redding (Eds.), *Juvenile delinquency: Prevention, assessment, and intervention* (pp. 19-44). New York, NY: Oxford University Press.

Dishion, T. J., & Patterson, G. R. (2006). The development and ecology of antisocial behaviour. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (Vol. 3, 2nd ed., pp. 503-541). New York, NY: John Wiley.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24827 people: Systematic review and meta-analysis. *British Medical Journal*, *345*. doi:10.1136/bmj.e4692

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2,* 293-323. doi:10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30. doi:10.1037/1040-3590.12.1.19

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

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: John Wiley.

Hurt, H., Malmud, E., Brodsky, N. L., & Giannetta, J. (2001). Exposure to violence: Psychological and academic correlates in child witnesses. *Archives of Pediatrics & Adolescent Medicine*, *155*, 1351-1356. doi:10.1001/archpedi.155.12.1351

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*, 119-127. doi:10.2307/2986296

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva, Switzerland: World Health Organization.

Loeber, R. (1990). Development and risk factors of juvenile antisocial behaviour and delinquency. *Clinical Psychology Review*, *10*, 1-41. doi:10.1016/0272-7358(90)90105-J

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & White, H. R. (2008). *Violence and serious theft: Development and prediction from childhood to adulthood*. New York, NY: Routledge.

Loeber, R., Kalb, L., & Huizinga, D. (2001). *Juvenile delinquency and serious injury victimization* (Juvenile Justice Bulletin). Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

McGuw, M., & Iacono, W. G. (2005). The association of early adolescent problem behavior with adult psychopathology. *The American Journal of Psychiatry*, *162*, 1118-1124. doi:10.1176/appi.ajp.162.6.1118

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783-792.

Nores, M., Belfield, C. R., Barnett, W. S., & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis*, *27*, 245-261. doi:10.3102/01623737027003245

Patchin, J. W., Huebner, B. M., McCluskey, J. D., Varano, S. P., & Bynum, T. S. (2006). Exposure to community violence and childhood delinquency. *Crime & Delinquency*, *52*, 307-332. doi:10.1177/0011128704267476

Piquero, A. R., Daigle, L. E., Gibson, C., Leeper Piquero, N., & Tibbetts, S. G. (2007). Research note: Are life-course-persistent offenders at risk for adverse health outcomes? *Journal of Research in Crime & Delinquency*, *44*, 185-207. doi:10.1177/0022427806297739

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

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

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior*, *31*, 449-462. doi:10.1007/s10979-006-9071-7

Scott, C. L., & Resnick, P. J. (2006). Violence risk assessment in persons with mental illness. *Aggression and Violent Behaviour*, *11*, 598-611. doi:10.1016/j.avb.2005.12.003

Singh, J. P. (2012). The history, development, and testing of forensic risk assessment tools. In E. L. Grigorenko (Ed.), *Handbook of juvenile forensic psychology and psychiatry* (pp. 215-225). New York, NY: Springer Science+Business Media. doi:10.1007/978-1-4614-0905-2_14

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*, 499-513. doi:10.1016/j.cpr.2010.11.009

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016

Smith, D. (2004). *The links between victimization and offending*. Edinburgh, Scotland: Centre for Law and Society.

Sreenivasan, S., Kirkish, P., Garrick, T., Weinberger, L. E., & Phenix, A. (2000). Actuarial risk assessment models: A review of critical issues related to violence and sex-offender recidivism assessments. *The Journal of the American Academy of Psychiatry and the Law*, *28*, 438-448.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P., Robbins, P. C., Mulvey, E. P., . . .Banks, S. (2000). A classification tree approach to the development of actuarial violent risk assessment tools. *Law and Human Behavior*, *24*, 83-100. doi:10.1023/A:1005478820425

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285-1293. doi:10.1126/science.3287615

Thomas, S. D., & Leese, M. (2003). Invited editorial. A green-fingered approach can improve the clinical utility of violence risk assessment tools. *Criminal Behaviour and Mental Health*, *13*, 153-158. doi:10.1002/cbm.538

Van der Put, C. E. (2013). Youth Actuarial Risk Assessment Tool (Y-ARAT): The development of an actuarial risk assessment instrument for predicting general offense recidivism on the basis of police records. *Assessment*, *21*, 340-351. doi:10.1177/1073191113495710

Young, D., Moline, K., Farrell, J., & Bierie, D. (2006). Best implementation practices: Disseminating new assessment technologies in a juvenile justice agency. *Crime & Delinquency*, *52*, 135-158. doi:10.1177/0011128705281752

Youth Justice Board. (2005). *Risk and protective factors*. London, England: Author.

Downloaded from ijo.sagepub.com at Universiteit van Amsterdam on April 8, 2016