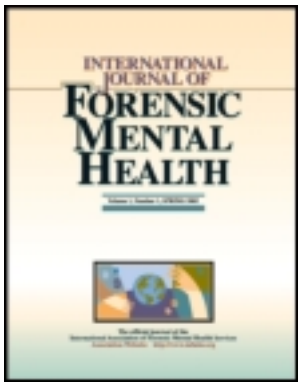
Assessment of offending during leave

Hilterman, E.L.B.; Philipse, M.W.G.; de Graaf, N.D.

Published in:
International Journal of Forensic Mental Health

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record



International Journal of Forensic Mental Health
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ufmh20

Assessment of Offending During Leave: Development of the Leave Risk Assessment in a Sample of Dutch Forensic Psychiatric Patients

Ed L. B. Hilterman [a] , Martien W. G. Philipse [b] & Nan Dirk de Graaf [c]
[a] Justa Mesura, Consultancy & Applied Research, Barcelona, Spain
[b] Pompestichting Institute for Forensic Mental Health, Nijmegen, The Netherlands
[c] Nuffield College, University of Oxford, Oxford, UK

Available online: 02 Sep 2011

To cite this article: Ed L. B. Hilterman, Martien W. G. Philipse & Nan Dirk de Graaf (2011): Assessment of Offending During Leave: Development of the Leave Risk Assessment in a Sample of Dutch Forensic Psychiatric Patients, International Journal of Forensic Mental Health, 10:3, 233-243

To link to this article: http://dx.doi.org/10.1080/14999013.2011.598601

INTERNATIONAL JOURNAL OF FORENSIC MENTAL HEALTH, 10: 233–243, 2011

ISSN: 1499-9013 print / 1932-9903 online
DOI: 10.1080/14999013.2011.598601

Routledge
Taylor & Francis Group

Assessment of Offending During Leave: Development of the Leave Risk Assessment in a Sample of Dutch Forensic Psychiatric Patients

Ed L. B. Hilterman
Justa Mesura, Consultancy & Applied Research, Barcelona, Spain

Martien W. G. Philipse
Pompestichting Institute for Forensic Mental Health, Nijmegen, The Netherlands

Nan Dirk de Graaf
Nuffield College, University of Oxford, Oxford, UK

The Leave Risk Assessment (LRA) is an actuarial risk assessment tool composed of both historical and treatment-related subscales, developed to assess the risk of serious reoffending by forensic psychiatric patients. This study examines the psychometric properties of the Leave Risk Assessment. The sample was drawn from the same population on which the tool was developed consisting of 195 Dutch forensic psychiatric patients; 78 who re-offended during leave and 117 non-reoffenders. The Leave Risk Assessment had moderate predictive validity (AUC = .84) as well as incremental predictive value over the HCR-20. The results show that the LRA can have a significant contribution in the decision-making process regarding authorized leave.

Keywords: risk assessment; actuarial assessment; context risk assessment; forensic patients

Probationary leave for forensic psychiatric patients is valued as an important means to evaluate the progress of treatment outside the closed environment of the forensic psychiatric hospital. Decisions to grant authorized leave (including conditional discharge) can have serious legal and ethical implications not only for patients but also for members of the community. Offenses committed by forensic patients during leave may have severe consequences for the victims and a significant impact on society, creating alarm and anxiety. As a response to recidivism during leave, political parties have requested increasingly repressive measures on patients, and these policies have led to substantial restrictions of rehabilitation possibilities.

While clinicians progressively integrate structured risk assessment in their daily practice to establish risk levels and create treatment plans (e.g., Douglas & Reeves, 2009), the risk factors applied in the decision-making for probationary leave by governing bodies responsible for these decisions remain often unclear. Callahan and Silver (1998) found variability in leave decisions were influenced by an interaction of system structures and individual characteristics of patients. Gobeil and Serin (2009) suggested that parole board members are not following guidelines as intended and vary in the type of information taken into account when forming these important decisions.

Among forensic patients, there have been few studies examining factors associated with decision to approve leave (McDermott, Scott, Busse, Andrade, Zozaya, & Quanbeck, 2008) and factors underlying these decisions (Gobeil & Serin, 2009). McDermott and Thompson (2006) argue that,

The authors wish to thank Anders Tengström and Makenzie Chilton for their comments on earlier versions of this article.

Editor's Note. I am sad to report that Martien W. G. Philipse passed away on April 18, 2011, after a lengthy illness. Martien was a regular attendee and presenter at academic and professional meetings, including IAFMHS conferences. He will be missed, but remembered fondly, by many of us.

Address correspondence to Ed L. B. Hilterman, Justa Mesura, Consultancy & Applied Research, C/ Palomar 35 3-3, 08030 Barcelona, Spain. E-mail: ehilterman@justamesura.com

compared to the use of structured assessment by clinicians, "the need for data-driven decisions in forensic systems tasked with making release decisions is even more critical" (p. 110). Further, the authors highlighted the importance of actuarial methods to assist government bodies with making serious and significant decisions. This actuarial assessment should take into account dynamic risk factors, since this is pivotal to measure the variability in the patients' *risk state*. The *risk state* has been defined as an individual's propensity to become involved in violence at a given time, based on changes in biological, psychological, and social variables in his or her life (Skeem & Mulvey, 2002). Measuring *risk state* focuses on the degree to which risk factors change within a person during a given period; more specifically, on the intra-individual variability in violence potential (Skeem & Mulvey, 2002). This measurement of intra-individual variability, through a combination of static and dynamic risk factors, is a key issue in improving the accuracy of risk assessment (Craig, Browne, Stringer, & Beech, 2005; Douglas & Skeem, 2005).

DUTCH HOSPITAL ORDER AND LEAVE POLICY

In The Netherlands, serious offenders who are not fully responsible for their crime can be sentenced to *terbeschikkingstelling* (TBS). TBS is a court-ordered treatment measure which serves to protect society through confinement of the mentally disordered offender in a forensic psychiatric hospital and by offering treatment to the offender (for a description of the TBS, see, Van Marle, 2002). Reintegration to the community is a gradual process, through successive levels of increased freedom, dependent on treatment progress. The first step is supervised leave in the community, followed by unsupervised leave, which can include up to six overnight stays outside the hospital. In the final phase, the patient will completely move out of the hospital to housing in the community.

In the Dutch forensic psychiatric system, as in other forensic systems (see Blaauw, Hoeve, Van Marle, & Sheridan, 2002), the hospital requests approval for leave trajectories to a governing body. In The Netherlands, the Ministry of Justice is responsible for the decisions. The leave application includes a section on treatment progress and risk assessment, in which the risk of recidivism is discussed.

After several serious offenses during authorized leave in 1996 and 1997, the Dutch Ministry of Justice sought the development of a risk assessment tool specifically regarding risk of recidivism during authorized leave. The main purpose of this tool was to facilitate the administrative decision making by the Ministry of Justice regarding leave applications from the forensic hospitals. This tool was seen as a priority, as the decision for authorized leave previously were made through unstructured assessments prepared by civil servants without any specific training or education in risk assessment.

The criteria for the development of the assessment tool were:

- The tool should be easy to use by the staff members who prepared the decision making on leave applications;
- Rating should require file information only, and should take no more than 30 minutes.

Since the objective was a straightforward conclusion regarding the risk of offending during authorized leave and not intervention planning, an actuarial approach was chosen. Actuarial assessments, however, are not without criticism. Hart, Michie, and Cook (2007) pointed out issues with actuarial assessments. Although actuarial prediction frames risk assessment in probabilistic statements, Hart et al. have cautioned that estimates from actuarial tools are not applicable to individual persons if the actuarial instrument is developed on the basis of group estimates. They analyzed the 95% confidence intervals (CI) of two well-known actuarial instruments, and found that the CIs overlapped considerably, with little distinction when the predictors are applied at the individual level. Others (e.g., Mossman & Sellke, 2007) have disputed these conclusions, arguing that a differentiation between "group risk" and "individual risk" is incorrect.

Both actuarial and non-actuarial risk assessment tools (e.g., Structured Professional Judgment, or SPJ, tools) aim to help us make better decisions than those we would make without them. However, it is imperative to choose the assessment approach that is most appropriate for the context. Governing bodies responsible for leave decisions, without direct contact with the patient nor the responsibility to develop treatment plans, have an obligation to make an independent decision regarding authorized leave and actuarial tools can offer a solution to facilitate this decision making. On the other hand clinicians who are in direct contact with the patient and are responsible for risk reduction through treatment, often obtain more useful information when they apply tools based on the methodology of Structured Professional Judgment (SPJ). For their decision making, governmental bodies use information supplied by the hospitals; however, they cannot rely solely on SPJ made by clinicians, and must independently determine the risk of offending while on leave. Further, SPJ estimates are also based on clinical judgments and are therefore not free from biased decision making (Murray & Thomson, 2010), neglecting gathered information in the decision making process (Hilterman & Chakhssi, 2002), and being influenced by the relation the clinician has with the patient (Dernevik, Falkstein, Holmquist, & Sandell, 2001; De Vogel & De Ruiter, 2004).

The LRA is an actuarial risk assessment tool (for the development of the LRA see Hilterman, 2000, 2001); the risk is calculated through an algorithmic procedure using regression weights, which were established by a combination of static and dynamic risk factors. This combination of risk factors is important to measure change in the patient's *risk*

state (Douglas & Skeem, 2005). A computer program with a database was developed to facilitate the use of the LRA where the user has to input the information for the several risk factors. An algorithm calculates the risk for each individual patient and reports this in four categories, low (0% – 24%), medium (25% – 49%), high (50% – 74%) and very high risk (75% – 100%), and also provides a specific percentage of the estimated risk based on logistic regression scores for individuals.

The initial results of the LRA seem very promising (Hilterman, 2000, 2001); however a study of the measures psychometric properties is needed to support its validity. The present study had the following aims: 1) to test interrater reliability, 2) to examine the association of the LRA with the HCR-20 risk assessment tool as they measure similar constructs, 3) to test the validity of the individual items of the LRA as predictors of the outcome variables, 4) to determine if the predictive accuracy of the LRA differs between the LRA construction sample and the validation sample, and 5) to compare LRA's accuracy in predicting serious and general offending during leave with the predictive validity of the HCR-20.

METHOD

Participants

Study participants included all TBS patients who committed a violent or felony offense during authorized leave in the years 1997 through 2003. The majority of the offenses committed during leave were not registered in a database or in any other way and we identified these reoffenders through an extensive search of patient files and case notes at the Ministry of Justice, hospital files, and information from individual employees from hospitals and the Ministry of Justice. All information was verified through contacts with forensic hospitals, probation officers and/or police officers. We identified a total of 79 reoffenders; due to insufficient information, one reoffender, who committed a less serious offense (described below), was excluded from the analyses. Serious criminal offenses were operationalized in accordance with the Dutch criminal law, serious offenses are defined as offenses with a maximum prison sentence of at least four years and were of a violent and/or sexual nature. Fifty individuals (63.3%) committed a serious offense, and 29 (36.7%) committed a less serious offense while on leave. In 3.8% of the offenses, the victim did not survive, 48.1% involved physical violence, and in 27.8% there were threats of violence.

The mean time between the start of leave and the offense was 134 days ($SD = 232.19$). However it is of note that 26.9% ($n = 21$) of reoffenders committed the offense on the first day of leave. The offenses were committed during supervised leave in 15.4% ($n = 12$) of cases, during unsupervised leave in 42.3% ($n = 33$) of cases, while the patient lived outside the hospital in 32.1% ($n = 25$) of cases, and while the hospital order was terminated conditionally in 10.3% ($n = 8$) of cases. The base rate of offending during leave in the population of TBS patients between July 1997 and December 2003 was 5.8%.

The 117 non-reoffenders were randomly selected from TBS patients who were on authorized leave and terminated the TBS between January 1998 and December 2003 and did not commit an offense during leave. The sub-sample of non-reoffenders represents 31.5% of the patients who terminated the TBS in the aforementioned period (1,392 TBS patients did not reoffend during leave in this period) and is representative of this population. Characteristics of reoffenders and non-reoffenders are presented in Table 1.

Materials

Leave Risk Assessment. In 1996, after several serious offenses committed by forensic psychiatric patients during leave, the Dutch Ministry of Justice took the initiative that led up to the development of the LRA (Hilterman, 2000, 2001). The sample used in the development of the LRA consisted of 47 patients who committed a serious offense during leave in the period 1988 to 1997, and 107 patients who, in the same period, did not commit any offense during leave or committed a minor non-violent offense ($n = 5$), such as theft. The retrospective development study had an explorative design with variables divided into different domains of static and dynamic risk factors.

On the basis of logistic regression analysis, risk factors predictive of reoffending during leave were identified. The result was a risk assessment tool with two separate subscales; the first was a static subscale with historical information measured before the treatment period, the LRA-Historical Information Subscale (LRA-HIS). The second subscale consisted principally of dynamic risk factors measured during the treatment period, the LRA-Treatment Related Information Subscale (LRA-TRIS). See Table 2 for the risk factors on both subscales. In the initial validation study, the Area under the ROC Curve (AUC) was .83 (CI 95%: .76–.89) for the LRA-HIS and .91 (CI 95%: .85–.97) for the LRA-TRIS. In a prospective pilot study (Hilterman & Chakhssi, 2002), the LRA showed good interrater reliability (LRA-HIS: $ICC = .97$, LRA-TRIS: $ICC = .83$, $n = 19$) and concurrent validity with the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997).

The coding of all but one of the LRA-TRIS risk factors was based on the available information regarding the behavior of the patient during the last year before the leave decision. An exception was made for the reallocation risk item, the most static risk factor of the LRA-TRIS, which included all reallocations (transfers) during the entire TBS period.

HCR-20. The HCR-20 (Webster et al., 1997) is a structured professional judgment (SPJ) risk assessment tool that comprises 10 historical (H), five clinical (C), and five risk management (R) items. The historical items evaluate previous antisocial and violent behavior and mental health history.

TABLE 1
Characteristics of Patients Who Did Not Reoffend, Who Committed a Less Serious Offense, and Who Reoffended Seriously During Leave

Variable	Non-reoffenders $N = 117$	Reoffenders		
		Not serious $N = 28$	Serious $N = 50$	Total $N = 195$
Demographic				
Mean age (at the time of the leave)	34.7	33.5	35.7	34.8
Female	6 (5%)	0 (0%)	1 (2.0%)	7 (4%)
Unemployed (at the time of the index offense)	47 (40%)	14 (50%)	28 (56%)	89 (46%)
Living with partner (at the time of the index offense)	24 (21%)	9 (32%)	8 (16%)	41 (21%)
Psychiatric				
Axis 1 disorder	21 (18%)	0 (0%)	9 (18%)	30 (15%)
Axis II disorder	78 (71%)[a]	24 (89%)	42 (88%)[b]	144 (78%)
Alcohol abuse	41 (35%)[c]	10 (36%)	26 (52%)[d]	77 (40%)
Drug abuse	25 (21%)[a,c]	12 (43%)[d]	21 (42%)[b]	58 (30%)
Mean intelligence score	98.9	100.3	97.2	98.7
Offenses				
Mean age of first conviction	23.9[d]	20.3[c]	21.0[c]	22.7
No prior convictions	63 (54%)[b]	6 (21%)[a]	15 (30%)[a]	84 (43%)
Mean number of prior convictions (≥ 1)	4.2[a]	6.2	8.9[b]	5.8
Previously convicted for sexual offense(s)	22 (19%)[a,c]	11 (39%)[d]	27 (54%)[b]	60 (31%)
Previously convicted for violent offense(s)	48 (41%)[c]	18 (64%)[d]	29 (58%)[d]	95 (49%)

Note. [a] < [b], $p < .05$. [c] < [d], $p < .01$ (two-tailed). The differences were investigated with the one-way F test or Chi-squared analysis. Sample sizes vary due to missing values.

The clinical items represent the clinical adjustment, and the risk management items assess the expected adjustment to future circumstances. In contrast to the H items, the ratings of the C and R items can fluctuate over time. All items of the HCR-20 are scored on a three-point scale from 0 to 2, yielding a total score ranging from 0 to 40. Having gathered the information for these items, a final risk judgment is then made indicating low, moderate, or high risk; no probabilistic estimates are assigned to these risk categories.

We incorporated the HCR-20 in this research because it is one of the most commonly used risk assessment tools in forensic psychiatry. We were not able to include the Psychopathy item of the HCR-20. Exclusion of this item is supported by several studies (De Vogel, De Ruiter, Hildebrand, Bos, & Ven, 2004; Douglas, Ogloff, Nicholls, & Grant, 1999; Kroner & Mills, 2001) that have found, when applied in research settings, the psychopathy item only minimally adds to the predictive accuracy to the HCR-20 scores. However, for clinical use the inclusion of the PCL-R is strongly recommended (Hare, 1991).

Procedure

All variables were coded using information from the TBS patient files at the Ministry of Justice. These files contain information on treatment progress, judicial background as well as information on violent and non-violent incidents during treatment. The agreement between the raters was calculated on the basis of 20 cases (10.3% of the total sample), which were each coded independently by two raters.

To compare the information on the risk state of the non-reoffenders during the rehabilitation process with the risk state of the reoffender group, we coded the information on the basis of leave requests and other reports, like advice of the hospital to the court on the prolongation of the TBS. To ensure continuity, the data gathered for the two groups was matched on the type of leave during which the reoffenders committed the reoffense. The selection of the type of leave for the non-reoffenders was random, only taking into account the equal distribution of the type of leaves between the two groups.

Three graduate students and two licensed professionals, a sociologist and a psychologist, both employed in forensic psychiatric research, rated the information. Before data collection, the raters received 40 hours of group training to become familiar with the files and the instruments. Raters were instructed to collect information on the basis of concrete descriptions of behavior and not on the basis of interpretations present in the files. The raters were blind to which group (reoffender versus non-reoffender) patients were in, as the files were screened and all information pertaining to the situation during or after leave was removed.

Statistical Analyses

To measure the interrater reliability the intraclass correlation coefficient (ICC) was used, with one-way random effects for single raters. The classification to translate the single measure ICCs into descriptors was: ICC $\geq .75$ = excellent; $.60 \leq$ ICC $< .75$ = good; $.40 \leq$ ICC $< .60$ = moderate; ICC $< .40$ = poor (Fleiss, 1986). To obtain the concurrent

TABLE 2
Risk Factors of Leave Risk Assessment-Historical Information Subscale and Treatment Related Information Subscale

LRA Historical Information Subscale Item	Coding
1. Judicial career before index offense	
1.1 Prior conviction	0 = none or one 1 = two or more
1.2 Age at first conviction	0 = 25 years or older 1 = younger than 25 years
1.3 Duration all prior detentions	0 = 120 days or inferior 1 = superior to 120 days
1.4 Prior conviction for sexual offense	0 = none 1 = one or more
1.5 Prior conviction for property offense	0 = none 1 = one or more
2. Indirect sexual abuse: sexual abuse of others younger than 16 years, in the educational environment of the subject before age of sixteen	0 = no 1 = yes
3. Relation with the victim of the index offense	0 = (ex-)partner or family relation 1 = no (ex-)partner or family relation
4. Type of index offense	0 = no sexual offense 1 = sexual offense
LRA Treatment Related Information Subscale	
1. Treatment compliance as reported in the leave request of the hospital	0 = total resistance: therapy is perceived as completely useless, there is no collaboration whatsoever with the planning or organization of the treatment: subject has passive and active resistance 1 = resistance 2 = no compliance, no resistance 3 = compliance 4 = total compliance: treatment is perceived as very useful, this is illustrated by regular initiatives for the benefit of the treatment: subject shows active involvement
2. Extent to which the patient takes responsibility for the index offense: rated on the basis of information extracted from the offense script. In this therapeutic intervention, the patient reflects upon the combination of events, behaviors, feelings and thoughts at the time before, during and after the index offense.	0 = totally not taken 1 = not taken 2 = not clear 3 = taken 4 = totally taken
3. Degree of norm violating behavior during the last year before leave	0 = no norm violation 1 = some not serious/not frequent 2 = serious or frequent norm violation
4. Deviant social network: Number of family, friends or acquaintances with history of criminal or anti-social behavior or known to stimulate such behavior who had contact with the patient in the two years before leave	0 = none 1 = one or two persons 2 = three to four persons 3 = five or more persons
5. Frequency of violations of leave conditions in the last two years before leave	Range from none up to four
6. Alcohol use during TBS: included both legal and illegal alcohol use in the year before leave was granted	0 = never 1 = one single time 2 = more than once
7. Reallocation: Number of transfers to another TBS-hospital during whole the TBS	0 = no reallocation 1 = one reallocation 2 = two or more reallocations

validity, Spearman's rho was used to measure the association of the LRA items with the items on the HCR-20.

In our analyses we used two dependent variables; the first was serious offending during leave, the second includes all other offenses committed in this context, i.e., general offending during leave. To establish which LRA risk items contribute to the prediction of offending during leave we used logistic regression analyses with a 95% confidence interval. This is a technique to estimate the probability that an event occurs.

To compare the effects of the risk factors on the outcome variable between the LRA construction study and the present validation study, interaction effects were added to the model. The interaction terms were the product of multiplications between the group variable (0 = *LRA construction study,* 1 = *LRA validation study*) and the individual risk factors of the

LRA. The b-coefficient of the interaction term represents the change on the log odds between the two studies. A significant interaction effect indicates that the predictive accuracy of the risk factor on the outcome changed significantly between the two studies.

The sensitivity and specificity (illustrated in Table 6) was calculated according to the method proposed by Quinsey, Harris, Rice, and Cormier (1998, p. 50). The predicted probability for each individual patient, with a range from 0.0 to 1.0, was calculated by applying the parameters of the logistic regression model. For this purpose the predicted probability based on LRA and HCR-20 was recoded into the original risk categories of the tools; four categories (low, moderate, high, and very high) for the LRA and three categories (low, moderate, and high) for the HCR-20.

Predictive power was estimated by means of the Receiver Operating Characteristic (ROC) curve analysis (Mossman, 1994). For the interpretation of the area under the ROC curve (AUC), we use the categorization proposed by Sjöstedt and Grann (2002): < .60 low accuracy; .60–.70 marginal accuracy; .70–.80 modest accuracy; .80–.90 moderate accuracy; and over .90 high accuracy. The AUCs, and 95% confidence intervals, were calculated using MedCalc version 8.1, as were the differences between the AUCs. All other analyses were conducted using SPSS 15.1.

RESULTS

Interrater Reliability and Concurrent Validity

The interrater reliability of the LRA-HIS was excellent (ICC = .93) and good for the LRA-TRIS (ICC = .62) and the LRA total score (ICC = .72). The interrater reliability was also good for the HCR-20 total score (ICC = .74) and the SPJ summary risk rating of high, moderate, or low risk (ICC = .61).

The LRA-HIS was moderately correlated with the H scale of the HCR-20 ($r_s = .46$, $p < .001$) and with the C scale and R scale of the HCR-20 ($r_s = .20$ for each, $p < .01$). The LRA-TRIS was strongly associated with the C and R scales, and the HCR-20 total score (respectively $r_s = .58$, $r_s = .56$ and $r_s = .60$, respectively, all p's < .001), but less so with the Historical items of the HCR-20 ($r_s = .27$, $p < .001$). Finally, the LRA total score correlated strongly with the HCR-20 total score ($r_s = .57$, $p < .001$).

The Accuracy of the LRA Risk Factors in Assessing Reoffending During Leave

Using univariate logistic regression analysis all the individual risk factors of the LRA-HIS were significant predictors of both serious and general offending during leave. In these analyses, most risk factors of the LRA-TRIS were significantly associated with both serious and general offending during leave; however, there were three exceptions. Responsibility taken for the index offense was the only risk item of the LRA-TRIS without any significant predictive value either to serious ($p = .053$) or general offending ($p = .116$) during leave. Alcohol use during TBS was a significant predictor for general offending during leave ($p = .002$), but it was not significant in regard to serious offending ($p = .325$). In contrast, the Deviant Social Network item, was significantly associated with serious offending during leave ($p = .014$), but was not associated with general offending ($p = .084$).

To measure the contribution of the individual risk factors of the LRA subscales in the assessment of serious and general offending during leave, two sets of multiple logistic regression analysis were performed with both the LRA-HIS and the LRA-TRIS. All four sets of analysis are summarized in Table 3. The LRA-HIS had a significant fit in predicting serious and general reoffending. The *Wald* values in Table 3 show that the Judicial History item was the strongest static predictor in the historical model for serious and general reoffending. The odds of committing a serious offense during leave increase 1.37 times with each step of the five-point scale on the Judicial History item. This means that a four-step change (from 1 to 5) increases the odds by 5.48 for serious reoffending and by 6.44 for committing a general offense during leave.

The LRA-TRIS also had a significant fit in the models assessing risk of serious and general reoffending. Treatment compliance reduced the risk of general offenses during leave significantly, while in assessing serious offenses it did not achieve significance. Alcohol use during the TBS did not significantly predict serious reoffending, but it was one of the most powerful risk factors in the assessment of the risk of general offending during leave. The Reallocation item proved to be an important risk factor in the assessment of serious offenses during leave; a patient who had been reallocated several times had a 5.1 times higher risk of committing a serious offense than a patient who never had been reallocated to another TBS-clinic. A multivariate analysis was conducted with the predictors of the static and the treatment-related models combined for both serious and general reoffending. It can be seen that, under influence of the treatment-related predictors in the model, the Judicial History item lost significance as a predictor of serious offense while on leave (see Table 3). Also, the predictive significance of the Treatment Compliance item increased for serious offense when the influence of the static predictors was included in the model. In relation to general offending during leave both Judicial History and Treatment Compliance items were significant predictors.

Comparison of the Predictive Accuracy Between the LRA Construction and the Validation Study

As shown in Table 4, there were no significant differences between the univariate predictive accuracy of the historical

TABLE 3
Summary of Multivariate Logistic Regression Models for LRA-HIS and LRA-TRIS Separate and Combined, Predicting Serious and General Offending During Leave

	Serious Offending						General Offending					
	LRA-HIS & LRA-TRIS separate			Entire LRA			LRA-HIS & LRA-TRIS separate			Entire LRA		
LRA-HIS Risk Factor	*B*	*Wald*	*Exp (B)*	*B*	*Wald*	*Exp (B)*	*B*	*Wald*	*Exp(B)*	*B*	*Wald*	*Exp (B)*
Judicial history	0.31**	6.43	1.37	0.13	0.77	1.13	0.47***	17.20	1.61	0.45***	11.66	1.57
Indirect sexual abuse	1.53**	5.81	4.61	1.86***	7.68	6.41	1.28*	3.50	3.60	1.34*	3.33	3.80
Relation to victim	0.66	1.97	1.93	0.97*	3.20	2.63	0.64	2.48	1.90	0.63	1.95	1.92
Sexual index offense	0.75*	3.82	2.11	1.10**	3.00	1.25	0.47	1.65	1.61	0.65	2.58	1.92
Constant	−2.87	31.68	.057				−2.42	30.74	.089			
LRA-TRIS	$\chi^2(4, N = 195) = 29.67, p < .001$						$\chi^2(4, N = 195) = 42.92, p < .001$					
Treatment compliance	−0.54*	3.59	.58	−0.68**	4.35	0.51	−0.62**	5.11	0.54	−0.72**	5.33	0.49
Responsibility taken for the index offense	0.06	0.05	1.06	0.06	0.04	1.06	0.08	0.12	1.08	0.15	0.34	1.16
Norm violating behavior	0.48*	2.75	1.62	0.45	2.03	0.64	0.34	1.61	1.41	.28	0.85	0.76
Alcohol use	−0.07	0.05	.93	−0.24	0.49	0.79	0.65**	5.54	1.91	0.72*	5.18	2.04
Reallocations	0.94***	9.29	2.55	1.11***	10.50	3.04	0.53*	2.99	1.70	0.58*	2.77	1.78
Deviant social network	0.43*	3.27	1.53	0.35	1.80	1.42	0.27	1.63	1.31	0.07	0.08	1.07
Violations leave conditions	0.12	0.40	1.13	0.18	0.74	1.20	0.09	.22	1.09	0.03	0.02	1.03
Constant	−0.62	0.43	0.54	−1.01	1.14	0.37	0.46	0.27	1.58	−0.72	0.65	0.49
	$\chi^2(7, N = 195) = 37.26, p < .001$			$\chi^2(11, N = 195) = 61.92, p < .001$			$\chi^2(7, N = 195) = 34.98, p < .001$			$\chi^2(11, N = 195) = 68.99, p < .001$		

Note. $N = 195$. $^{*}p < 0.10$; $p < 0.05$; $^{***}p < 0.01$.

risk items in the construction and validation studies for either serious or general offending during leave. Further, in the multivariate analysis, the effects of the historical risk factors of the LRA on serious and general offending while on leave showed no significant differences between the construction and the validation studies.

The univariate predictive accuracy of the treatment-related risk items Responsibility Taken for the index offense and Norm Violating Behavior for serious reoffending were, compared to the construction study, significantly decreased in the validation study. The parameters of the remaining treatment-related risk factors were not different between the two studies. Further, in the multivariate model on serious reoffending, the effect of Responsibility Taken for the index offense was significantly decreased ($p < .05$). The predictive accuracy of all other treatment-related risk items regarding serious reoffending were not significantly different between the construction and validation studies.

Predictive Validity Compared with the HCR-20

The two subscales of the LRA assessed serious and general offending during leave or conditional discharge equally well (see Table 5). The combination of both subscales in the LRA total score yielded a significant increase in predictive power compared to the HIS ($p < .05$) and TRIS ($p < .01$) subscales separately for both serious and general reoffending during leave. The LRA-HIS was not significantly different from the H scale score of the HCR-20 in assessing reoffending during leave. In contrast, the LRA-TRIS assessed risk of serious and general reoffending during leave significantly better than the Clinical and Risk management scales of the HCR-20 ($p < .05$ and $p < .01$, respectively). Moreover, the

TABLE 5
Areas Under Receiver Operating Characteristic Curves for LRA and HCR-20 ($N = 195$)

	Serious Offending	General Offending
LRA Historical Information Subscale	.76 (95% CI: .69–.82)	.77 (95% CI: .70–.82)
LRA Treatment Related Information Subscale	.75 (95% CI: .69–.81)	.73 (95% CI: .66–.79)
LRA total score	.84 (95% CI: .78–89)	.83 (95% CI: .77–.88)
HCR-20 total score	.69 (95% CI: .62–.75)	.70 (95% CI: .63–.76)
H-scale score	.68 (95% CI: .61–.74)	.70 (95% CI: .63–.77)
C-scale score	.66 (95% CI: .58–.72)	.64 (95% CI: .57–.71)
R-scale score	.62 (95% CI: .55–.69)	.62 (95% CI: .55–.69)
Final risk judgment	.69 (95% CI: .62–.75)	.65 (95% CI: .58–.71)

LRA total score had a significantly higher predictive validity compared to the HCR-20 total score ($p < .001$) and HCR-20 summary risk judgment ($p < .001$). Multivariate logistic regression analyses were conducted with the HCR-20 total score, the HCR-20 summary risk judgment and the LRA total score using the forward stepwise method. The LRA total score produced a significant fit for serious reoffending, $\chi^2(1, N = 195) = 57.41$, $p < .001$, and general reoffending, $\chi^2(1, N = 195) = 66.53$, $p < .001$. In both analyses, the HCR-20 total score and the final risk judgment did not produce a significant improvement to the model after the LRA was entered. Subsequently, we tested the incremental validity of the Clinical and Risk management scales of the HCR-20 on the LRA-TRIS, using the same procedure. The LRA-TRIS produced a significant fit ($\chi^2(1, N = 195) = 38.27$, $p < .001$ and $\chi^2(1, N = 195) = 33.96$, $p < .001$ for serious and general reoffending respectively), the Clinical ($p = .81$, $p =$

TABLE 4
Areas Under Curves (AUCs) of Receiver Operating Characteristic Analyses for the LRA Risk Factors of the LRA Construction Study and the Validation Study

	Historical Information Subscale				Treatment Related Information Subscale		
Risk factors	Construction study	Validation study	Difference between studies	Risk factors	Construction study	Validation study	Difference between studies
Judicial history	.73**	.69**	n.s.	Treatment compliance	.68**	.65**	n.s.
Prior conviction >1	.67**	.59	n.s.	Responsibility taken for index offense	.75**	.58	**
Former detention > 120 days	.66**	.63**	n.s.	Norm violating behavior	.78**	.67**	*
At least one prior property offense	.63*	.60*	n.s.	Alcohol use during TBS	.63*	.56	n.s.
At least one prior sex offense	.63*	.67**	n.s.	Reallocations	.62*	.62*	n.s.
Age first conviction < 25	62*	.57	n.s.	Deviant social network	.61*	.61*	n.s.
Indirect sexual abuse	.61*	.61*	n.s.	Violations leave conditions	.71**	.57	n.s.
Relation to victim	.64**	.59	n.s.				
Sexual index offense	.59	.62*	n.s.				

Note. $N = 349$. Differences between the LRA construction and validity study were calculated using logistic regression analysis with interaction terms (group * risk factor) on serious offending for construction and validation sample.
n.s., not significant; * $p < .05$; ** $p < .01$.

TABLE 6
Recidivists and Nonrecidivists During Leave in the Original 4 Risk Categories for LRA and 3 Risk Categories for HCR-20, Serious Versus General Recidivism

	Serious Recidivism		General Recidivism	
	Recidivists	Non-recidivists	Recidivists	Non-recidivists
LRA				
Low	11 (22.0%)	111 (76.6%)	11 (14.1%)	64 (54.7%)
Moderate	14 (28.8%)	24 (16.6%)	16 (20.5%)	35 (29.9%)
High	15 (30.0%)	8 (5.5%)	25 (32.1%)	14 (12.0%)
Very High	10 (20.0%)	2 (1.4%)	26 (33.3%)	4 (3.4%)
Total	50 (100%)	145 (100%)	78 (100%)	117 (100%)
HCR-20				
Low	23 (46.0%)	120 (82.8%)	17 (21.8%)	59 (50.4%)
Moderate	27 (54.0%)	25 (17.2%)	51 (65.4%)	53 (45.3%)
High	0 (0.0%)	0 (0.0%)	10 (12.8%)	5 (4.3%)
Total	50 (100%)	145 (100%)	78 (100%)	117 (100%)

Note. $N = 195$. Numbers in parentheses are column percentages for each instrument.

.99, respectively) and Risk Management scales ($p = .59$, $p = .88$, respectively) of the HCR-20 were not entered into the model because they did not contribute significantly to the improvement of the model.

Sensitivity and Specificity

Table 6 presents the predictive probabilities, classified in the original three and four risk categories for the HCR-20 and LRA respectively. The LRA classified 50% of the serious reoffenders as high or very high risk and 93% of the non-reoffenders were classified as low or moderate risk. The overall correct classification for the LRA for serious recidivism during leave was 82.1%, compared to 75.4% for the HCR-20. None of the reoffenders were classified as high risk on the HCR-20.

With regards to general offending while on leave, the LRA obtained an overall correct classification of 76.9% and the HCR-20 of 69.7%. According to the predictive probability the LRA identified 65.4% of the reoffenders as high or very high risk, whereas 84.6% of the non-reoffenders were classified as low or medium risk (i.e., true negatives). The predictive probability based on the HCR-20, recoded into the three risk categories, classified 50% of non-reoffenders correctly as true negatives and identified 12.8% of reoffenders as high risk (i.e., true positives), but classified the majority (53.3%) of the patients as moderate risk, 65.4% of the reoffenders and 45.3% non reoffenders respectively.

DISCUSSION

Practicing acquired skills during leave in community surroundings forms an important part in forensic psychiatry, showing and enhancing the progress of treatment. Clinicians seek to test the treatment progress of their patients in the outside environment. In this study we found that 5.75% of the patients who went on leave in the period July 1997 to December 2003, committed an offense while on leave, and at least 64% of these offenses resulted in extremely serious repercussions for the victims. Apart from the far-reaching effects for the victims, these offenses had serious consequences for the rehabilitation possibilities of the forensic psychiatric patients. In recent years the rehabilitation of forensic psychiatric patients has suffered, as a consequence of serious offenses during leave, under increasing repressive measures leading to pressing limitations.

To our knowledge the present study is the first one in The Netherlands in which the predictive validity of a risk assessment instrument, developed for assessing risk for reoffending during authorized leave or conditional discharge, was replicated in a sample of the same overall population. These data indicate that the LRA proves reliable to rate and has good concurrent validity with the HCR-20. Reliability and concurrent validity achieved in this study are comparable to the results obtained in an earlier prospective pilot study (Hilterman & Chakhssi, 2002).

With few exceptions, the individual items of the LRA-HIS were significantly related to serious or general offending during leave. Indirect Sexual Abuse was significantly associated with serious reoffense, but not with general offending, while Relation with the Victim of the index offense was associated with general recidivism but not with serious offense during leave. The same was true for almost all of the treatment-related risk items of the LRA, with the exception of Responsibility Taken for the index offense, which was not significantly associated with either serious or general offending while on leave. The many significant findings, on the other hand, highlight the important function dynamic risk items serve in assessing risk of reoffense.

Comparing the construction and the validity studies, there was no significant difference between the contribution of the LRA-HIS risk factors to the prediction of reoffense during leave. The same holds true for treatment-related risk factors, with the exception of the Norm Violating Behavior item in the univariate comparison and the item Responsibility Taken for the index offense in both univariate and multivariate comparisons. A recent study by Hildebrand (2006) found that the item Responsibility Taken for the index offense was a strong predictor (AUC = .82, $p \leq .001$) of absconding from the hospital, but was not significantly associated with recidivism during treatment. However, another study (Hildebrand, Hesper, Spreen, & Nijman, 2005) found that this item was a significant predictor of both violent and general recidivism. Hanson and Morton-Bourgon (2005) concluded, on the basis of their meta-analysis, that denial and low victim empathy had little or no relationship with recidivism. However, they also recognize that it is difficult to assess sincere remorse in criminal justice settings.

In this study, the LRA obtained moderate predictive validity. The combination of the two subscales in the LRA total score proved to have added predictive value compared to both subscales individually. It is important to note that the LRA-TRIS, a predominantly dynamic subscale performed equally well as the historical subscale in assessing reoffending during leave. This confirms that the combination of historical and dynamic subscales in actuarial risk assessment enhances the accurateness of the assessment (Craig et al., 2005).

In line with previous findings (De Vogel, Ruiter, Hildebrand, Bos, & Van de Ven, 2004; Douglas, Ogloff, & Hart, 2003; Hildebrand, Hesper, Spreen, & Nijman, 2005; Kroner & Mills, 2001), we found modest predictive validity for the HCR-20. An important result of this study is that the LRA performed better than the HCR-20 in the specific context of assessing offending during leave on a forensic psychiatric population. The LRA-TRIS showed strong incremental validity on the Clinical and Risk Management scales of the HCR-20. After controlling for the LRA total score, the HCR-20 had no predictive power in assessing serious and general offending during leave. However, the HCR-20 was used as an actuarial instrument rather than as an SPJ tool, by summing total scores (as has often been done in past research, but is different from its intended use), rather than using trained clinicians to generate a summary risk rating. Moreover, the psychopathy item was not included. Thus, while this method for scoring and interpreting the HCR-20 is similar to many past research studies, it may underestimate the true utility of the tool.

The LRA is not tested in other jurisdictions outside The Netherlands, however, the use of the LRA may have widespread applicability, as authorizations for leave are assigned to governing bodies in many other jurisdictions. This is an important area for future examinations to discover the potential contribution of the LRA as a standard tool in the authorization of leave decision-making process.

While not detracting from these findings, it is necessary to highlight several limitations of our study. It would have been preferable to test the LRA while controlling for the time the patients were at risk. Unfortunately, this was not possible because this information was not registered in the majority of the forensic hospitals. A prospective design would also have been preferable to the retrospective approach used, as this approach would have permitted interviews with patients during the data gathering, and the use of self-report and collateral information on reoffending rather than the exclusive reliance on hospital files.

The choice of graduate students and a licensed sociologist and psychologist as raters may be seen as a limitation, particularly in rating the HCR-20 risk summary. However, these raters are comparable to the intended users of the LRA, who are civil servants without special training or education in risk assessment. Vincent et al. (2001) found that trained research assistants were able to rate the HCR-20 from file information, and that these ratings predicted recidivism comparably to the ratings of clinicians. Further, Murray, Thomson, Cooke, and Charles (2011) found, in a recent comparison between experts in risk assessment, semi-experts and laypeople, that semi-experts were less subjected to attributional manipulations when rating dangerousness while laypersons and experts had similar difficulties. Additionally, there is a great deal of research with the HCR-20 in which raters of various levels of experience, training, professional background, and different methodology obtained comparable results in terms of reliability and predictive validity. In sum, the impact of rater experience is probably most evident in risk summary of the HCR-20, but may be modest.

Decision making on authorized leave implies serious and significant decisions that can have far reaching ethical, legal, and financial implications. In this study we demonstrated that the LRA can improve decisions made by governing bodies authorizing leaves. Having an evidence-based risk measure enhancing decisions will lead to more effective prevention of serious offenses during leave, subsequent negative consequences for victims, and more appropriate rehabilitation possibilities for forensic psychiatric patients. It is important that these decisions are empirically based, our results show that the LRA can be a positive contribution to this high priority area of government and public safety.

REFERENCES

Blaauw, E., Hoeve, M., Van Marle, H., & Sheridan, L. (2002). *Mentally disordered offenders: nternational perspectives on how to deal with criminally irresponsible offenders.* The Hague, The Netherlands: Elsevier.

Callahan, L. A., & Silver, E. (1998). Revocation of conditional release: A comparison of individual and program characteristics across four U.S. states. *International Journal of Law and Psychiatry, 21*, 177–186.

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression, 11*, 65–84.

Dernevik, M., Falkheim, M., Holmqvist, R., & Sandell, R. (2001). Implementing risk assessment procedures in a forensic psychiatric setting: Clinical judgment revisited. In D. Farrington, C. Hollin & M. McMurran (Eds.), *Sex and violence: The psychology of crime and risk assessment* (pp. 83–101). London, UK: Routledge.

De Vogel, V. & De Ruiter, C. (2004). Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *The Journal of Forensic Psychiatry and Psychology, 15*, 145–164.

De Vogel, V., Ruiter, C. de, Hildebrand, M., Bos, B., & Van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3*, 149–165.

Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*, 1372–1379.

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing the risk of violence in psychiatric outpatients: The predictive validity of the HCR-20 risk scheme. *Journal of Consulting and Clinical Psychology, 67*, 917–930.

Douglas, K. S., & Reeves, K. A. (2009). Historical-Clinical-Risk management-20 (HCR-20) violence risk assessment scheme: Rationale

application, and empirical overview. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 147–186). New York, NY: Routledge.

Douglas, K. S., & Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, & Law*, 11, 347–383.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York, NY: Wiley.

Gobeil, R., & Serin, R. (2009). Preliminary evidence of adaptive decision making techniques used by parole board members. *International Journal of Forensic Mental Health*, *8*, 97–104.

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

Hare, R. D. (1991). *The Revised Psychopathy Checklist*. Ontario, Canada: Multi-Health Systems.

Hart, S. D., Michie, C., & Cook, D. J. (2007a). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry, 190*(suppl. 949), 60–65.

Hildebrand, M. (2006). *Onderzoek Onttrekkingen en incidenten TBS 2000–2005 [Research into absconding and incidents during TBS, 2000-2005].* Utrecht: Expertisecentrum Forensische Psychiatrie.

Hildebrand, M., Hesper, B. L., Spreen, M., & Nijman, H. L. I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie: Een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R [The value of structured risk assessment and of the diagnosis psychopathy: A research to the reliability and predictive validity of the HCR-20, HKT-30 and PCL-R]*. Amsterdam, The Netherlands: Expertisecentrum Forensische Psychiatrie.

Hilterman, E. L. B. (2000). Tijdens de oefening terug naar af: Een onderzoek naar de predictie van ernstige recidive door TBS-gestelden [Back to square one: prediction of serious recidivism during leave by TBS patients]. *Tijdschrift voor Criminologie, 42*, 232–252.

Hilterman, E. L. B. (2001). Statische vergelijking van tijdens verlof recidiverende en niet- recidiverende TBS-gestelden [Static comparison of during leave recidivating and non-recidivating TBS patients]. *Proces, 80*, 121–127.

Hilterman, E. L. B., & Chakhssi, F. (2002, September). *The HCR-20, BSI & LRA: A comparison of three methods of prospective risk assessmen*t. Paper presented at the 12th annual conference of the European Association of Psychology and Law, Leuven, Belgium.

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471–489.

McDermott, B. E., Scott, C. L., Busse, D., Andrade, F., Zozaya, M., & Quanbeck, C. D. (2008). The conditional release of insanity acquittees: Three decades of decision-making. *Journal of the American Academy of Psychiatry and the Law*, *36*, 329–336.

McDermott, B. E., & Thompson, J. W. (2006). The review panel process: An algorithm for the conditional release of insanity acquittees. *International Journal of Law and Psychiatry*, *29*, 101–111.

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

Mossman, D., & Sellke, T. M. (2007, 5 July). Avoiding errors about 'margins of errors'. *British Journal of Psychiatry* Electronic letters sent to the journal available from: http:/bjp.rcpsych.org/cgi/letters/190/49/s60–5674

Murray, J., & Thomson, M. E. (2010). Applying decision making theory to clinical judgments in violence risk assessment. *Europe's Journal of Psychology, 2*, 150–171.

Murray, J., Thomson, M. E., Cooke, D. J., & Charles, K. E. (2011). Influencing expert judgment: Attributions of crime causality. *Legal and Criminological Psychology, 16*, 126–143.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*, 179–183.

Skeem, J., & Mulvey, E. (2002). Monitoring the violence potential of mentally disordered offenders being treated in the community. In A. Buchanan (Ed.), *Care of the mental lydisordered offender in the community* (pp. 111–142). New York, NY: Oxford Press.

Van Marle, H.J.C. (2002). The Dutch entrustment act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health, 1*, 83–92.

Vincent, G. M., Ross, D. J., Whittemore, K., Eaves, D., Hart, S. D., Ogloff, J. R. P., et al. (2001, April). *Using the HCR-20: File-based researcher ratings vs. file + interview- based clinician ratings*. Paper presented at the Founding conference of the International Association of Forensic Mental Health Services, Vancouver, BC, Canada.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *The HCR-20 scheme: Assessing risk for violence - version 2*. Burnaby, Canada: Mental Health, Law and Policy Institute, Simon Fraser University.