

Static and Dynamic Patient Characteristics as Predictors of Criminal Recidivism: A Prospective Study in a Dutch Forensic Psychiatric Sample

Martien W. G. Philipse,^{1,5} Maarten W. J. Koeter,² Cees P. F. van der Staak,³
and Wim van den Brink⁴

Published online: 28 April 2006

If clinicians in forensic psychiatry want to reduce risk of reoffending in their patients, they require insight into dynamic risk factors, and evidence that these add predictive power to static risk indicators. Predictors need to be evaluated under clinically realistic circumstances. This study aimed to validate dynamic and static variables as predictors of reconviction in a naturalistic outcome study. Data on static and dynamic risk factors were collected for 151 patients discharged from Dutch forensic psychiatric hospitals. Community follow-up was prospective, with a 5.5 year minimum. A prediction model was developed using Cox regression analysis. The magnitude of the predictive power of this model was estimated using receiver operating characteristic (ROC) analysis. The final prediction model contained four static and no dynamic predictors. The model's ROC area under the curve was .79 (95% CI .69–.89). Clinical risk ratings were non-predictive. Post hoc analyses exploring the influence of subgroups of patients did not yield better models. It is concluded that a small set of static predictors yielded a good estimate of future reconvictions; inclusion of dynamic predictors did not add predictive power.

KEY WORDS: violence risk assessment; criminal recidivism; clinical prediction; forensic psychiatry; dynamic risk factors.

Accurate assessments of patients' reoffending risk is of supreme importance in forensic psychiatric settings. Arguably, the extent to which risk is reduced during treatment and can be properly managed afterwards, are the prime markers for treatment success in such settings; accurate estimates of these parameters are therefore

¹ Department of Research and Assessment, Pompestichting Institute for Forensic Mental Health, Nijmegen, The Netherlands.

² Amsterdam Institute of Addiction Research, Overschiestraat 65, Amsterdam, The Netherlands.

³ Academic Center for Social Sciences, Radboud University of Nijmegen, Nijmegen, The Netherlands.

⁴ Department of Psychiatry, Academic Medical Center of the University of Amsterdam, Tafelbergweg 25, Amsterdam, The Netherlands.

⁵To whom correspondence should be addressed at Department of Research and Assessment, Pompestichting Institute for Forensic Mental Health, P.O. Box 31435, 6503 CK, Nijmegen, The Netherlands; e-mail: m.philipse@pompestichting.nl.

indispensable. They are also of vital importance to the community, which generally suffers the consequences of erroneous risk estimates. In establishing the level of reoffending risk, there has been a shift from exclusive reliance on clinical experience and expertise to the incorporation of knowledge and tools derived from empirical research findings. Over the years, research has consistently shown structured and empirically based methods to outperform unstructured clinical assessments (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Kroner & Mills, 2001; Mossman, 1994; Quinsey, Harris, Rice, & Cormier, 1998).

This finding lends support to the use of structured methods of risk assessment in forensic clinical practice. Meanwhile, an almost bewildering variety of such instruments has become available. Though most of these tools share some basic historic predictor variables, such as one or more indicators of past offending, or history of substance abuse, they nonetheless differ markedly with regard to the number and type of items included, scoring methods, and instructions for reaching a final assessment. Interestingly, these differences do not appear to relate in any systematic way to differences in predictive validity, which for most of these instruments is variable over studies. For the Violence Risk Appraisal Guide (VRAG, Quinsey et al., 1998), ROC areas under the curve (auc) for violent reoffending were reported as high as .86 (Harris & Rice, 2003) and as low as .60 (Kroner & Mills, 2001). For prediction of violent offending using the Historical/Clinical/Riskmanagement-20 (HCR-20, Webster, Douglas, Eaves, & Hart, 1997a), ROC-values ranging between .86 (De Vogel, 2005) and .62 (Kroner & Mills, 2001) were found. Similar variability is seen with other instruments, including the Level of Service Inventory—revised (LSI-R, Andrews & Bonta, 1995) and the Psychopathy Checklist—revised (PCL-r, Hare, 1991). Under ideal circumstances, most of such instruments, including the complex multiple iterative classification trees from the MacArthur risk assessment study (Monahan et al., 2001), seem able to produce an ROC auc of around .85. Available empirical evidence does not offer clues for allegiance to one particular instrument. Regarding the HCR-20, Webster et al. (1997a) state that the value of the instrument lies “in the general principles it espouses rather than in its detail” (p. 5). This may hold true for other risk assessment tools as well, in the sense that use of any structured risk assessment method offers predictive benefits, regardless of the exact content of that method. Monahan et al. (2001) observe and demonstrate that, if several different but equally (moderately) valid risk measures exist simultaneously, combined use of two or more of these may increase predictive validity.

With empirical data offering such leeway, one important factor determining the choice of risk assessment method from the clinical perspective, will be the question whether the instrument offers clues for treatment intervention. It has been argued that risk assessment and risk management are two separate tasks that should therefore be dealt with separately in research (Kraemer et al., 1997; Monahan et al., 2001). But such disengagement of the problem of risk assessment from clinical practice seems artificial, and begs the question where forensic mental health is concerned. In this field, risk assessments are conducted for the purpose of planning intervention. This means there will be a strong predilection for tools that include dynamic, changeable variables. Indeed, empirically based instruments that are purely

static essentially defy the entire notion of forensic mental health treatment, and are evidence that “statistical significance does not equal clinical relevance” (Grubin, 1999, p. 332).

This does, unfortunately, not change the fact that the clinical preference for dynamic risk factors does not sit comfortably with empirical findings thus far. Though instruments like the HCR-20 or LSI-R comprise several dynamic predictors, there is no clear evidence yet that these have incremental predictive validity given the historic (static) risk factors that are also included. Regression analysis by De Vogel (2005) showed that only a single predictor from each of the HCR-20’s three predictor domains (past-present-future) was needed to explain subsequent violent behavior, of which the historic predictor “Young age at first violent incident” was by far the strongest. Of the six dynamic predictors in the Sexual Violence Risk-20 (SVR-20, Boer, Hart, Kropp, & Webster, 1995) only one (“negative attitude towards intervention”) entered a similar regression analysis performed by Dempster & Hart (2002). Bonta, Law, and Hanson (1998) concluded from a meta-analysis that criminal history variables were most predictive of criminal behavior while clinical variables were least effective, and suggest that models of psychopathology should possibly not be a main point of reference in risk assessment. On the other hand, Monahan et al. (2001) found substance abuse, anger and violent fantasies to be consistently valid, dynamic predictors for future violence, that had incremental predictive power given several static predictors. In any case, the requirements of forensic mental health practice warrant continued efforts to determine to what extent dynamic variables can add predictive power to historical givens that are known to be predictive of future offending. The present study is one such effort.

Definition of what exactly constitutes a static or a dynamic variable tends to be a thorny issue. It has been addressed with great clarity, however, by Hanson & Harris (2000, 2001), and for the purpose of the present study we adopt their definition: dynamic variables are those that are amenable to deliberate intervention. Such variables may be stable, which means they only change slowly (e.g., substance dependence), or acute, in which case they may change very rapidly (e.g., mood). Changes on such variables should correspond systematically and demonstrably to changes in reoffending risk.

This final requirement recently led Gagliardi, Lovell, Peterson, and Jemelka (2004) to argue that risk factors that are only assessed once, can never be dynamic. The present authors do not concur with this view. The dynamism of a predictor variable is not determined by its measurement, but by its inherent possibility to change (i.e., its conceptual and clinical nature). Furthermore, forensic treatment settings aim to effect changes in behavior and pathology that will durably persist after treatment is ended; this is certainly the case in the setting where the present study was conducted. As soon as the patient is discharged the changes that were effected are hoped to endure, and as such indeed to be static. Gagliardi et al.’s assumption that dynamic characteristics of patients will go on changing randomly after discharge may not necessarily be false, but it is at odds with the assumptions on which much of forensic mental health treatment is based. Because our concern in the present study was with clinical realism, we chose to work from the latter assumptions, in which case a single measurement at the moment of discharge, the

method commonly applied in risk assessment studies, seems sufficient—especially where stable dynamic risk factors are concerned, as was the case. Repeated *pre-discharge* measurements would have the added benefit of allowing to test whether *changes* on the variable relate to subsequent reoffending.

Clinical choices with regard to risk assessment methods will not only be guided by content. Practical feasibility is another important factor. Amenities included in research projects will often be available to a far lesser extent, if at all, in everyday clinical practice. Gardner, Lidz, Mulvey, and Shaw (1996) note that the implementation of actuarial tools in clinical practice may simply be too costly. The problem is easily recognized when considering, for example, Hanson & Harris's (2001) study regarding the sex offender needs assessment rating (SONAR). This study employed researchers who were trained for a full week; who were supplied with an extensive manual; who completed their first week of file rating under direct supervision of the project leader; who met at regular intervals to avoid rater drift; and who spent 3–5 h on a single case. Under such circumstances, anything but excellent reliability findings would have been surprising. However, it is unlikely that such conditions can be structurally maintained when the SONAR is implemented in a treatment setting;—no matter how desirable this would be, limitations on resources will generally preclude it. Thus, reliability and validity findings from such research may not tell us very much about the validity of the instrument when applied clinically. It is likely that a considerable amount of “shrinkage” will occur because the way the instrument is applied does not answer to the high standards that were maintained in research. Indeed, Webster et al. (2002) expressed fears that some practitioners begin and end their study of the HCR-20 with a reading of the coding sheet. The authors of the MacArthur Risk Assessment study were aware of this problem too, and reduced their optimum, complex prediction model to a less powerful, but more practicable “clinically feasible” version, for instance by deleting the time consuming PCL-R rating from the procedure. (Monahan et al., 2001).

Another aspect of research methodology that increases the distance between research findings and clinical practice, is the fact that the majority of risk assessment studies has been conducted retrospectively. Rather than rating a patient in the present situation and following up his behavior in the future, present (violent) behavior is related to ratings of risk factors as documented in files from his clinical past. Though it may be argued that technically the two approaches will achieve the same goal, it is clear that only the former represents risk assessment as it appears clinically. By their very nature, retrospective studies allow only limited insight into an instrument's validity when used in clinical practice. Buchanan (1999) notes that the clinician has to deal with a great variety of situations, with fragmentary information, and with the fact that his patients may represent very different “samples”—aspects that are rarely fully reproduced in retrospective research designs. For such reasons, Harris, Rice, and Cormier (2002); considered prospective replication of the VRAG's predictive validity useful, even though this had already been demonstrated retrospectively, noting that the prospective design provided a “unique and rigorous test of the predictive accuracy” (p. 381). Generally, prospective research is preferred over retrospective setups in clinical prediction research (Andreasen, 2000; Grann, 1998).

Several authors have expressed doubts whether clinical risk assessment has been fairly judged in the research that is available (Buchanan, 1999; Hart, 1999; Litwack, 2001), and whether risk assessment instruments that are available are of value to routine clinical practice (Buchanan, 1999). Litwack (2002) encourages further study of “the underpinnings of clinical risk assessment” (p. 174). Even if such study does not yield the desired results, it will still help to identify points of divergence between clinical perceptions and convictions on the one hand and empirical fact on the other (see, for example, Sturidsson, Haggård-Grann, Lotterber, Dernevik & Grann, 2004), that can help to improve clinical processes. To the present authors it also seems obvious that, irrespective of results, empirical findings will be more persuasive to clinicians when they are directly involved in the research effort. As Margison et al. (2000) pointed out, evidence based practice requires practice based evidence.

The present study represents an attempt to derive dynamic predictors of re-offending risk directly from forensic clinical practice. Their incremental predictive value is weighed against historic risk factors. This was done in a longitudinal, prospective research setup. Hypothesized predictor variables, suggested by clinicians, were rated by the responsible treatment supervisor, while historical predictors were rated from files by clerical staff.

METHOD

Setting

The study was set in The Netherlands, in forensic inpatient settings for the execution of the judicial measure of *Terbeschikkingstelling* (TBS). TBS is a court-ordered treatment measure that can be imposed on perpetrators of serious violent or sex offences who were not fully accountable for their criminal acts due to a mental disorder at the time of the offence, and who are considered to be at risk to reoffend. TBS is imposed and extended or terminated by a judge, who takes into account the advice of a psychologist or psychiatrist expert witness, usually the patient’s treatment supervisor.

Participants

Data on static and dynamic patient characteristics were collected in seven (of the then nine) forensic psychiatric hospitals in The Netherlands between January 1, 1996 and December 31, 1998. Two hospitals did not participate in the study, in one case due to lack of discharges as the hospital was newly established, and in the other case due to lack of a coordinator to support the research effort on site.

Patients

Assessments of dynamic and static risk factors were completed for patients discharged from any of the seven hospitals either due to termination of the TBS-measure, or at the start of probationary leave. Patients who were transferred to

other forensic hospitals or prisons were excluded from the study, because they would not be at risk in the community. Probationary leave is the preferred mode of discharge in TBS. It means the patient lives and works outside the hospital with unrestricted liberties and minimum supervision, usually by a probation officer. Typically, contact intensity will be 1 h a week or less. The TBS-measure however still remains in force and if circumstances require re-hospitalization the patient can be readmitted.

Patients were not directly involved in the data collection process, so that according to Dutch medical research ethics patient consent was not needed. This was confirmed by the ethics board of one of the participating hospitals, where the project plan was submitted.

Nationwide, 180 patients met our discharge criteria during the data collection period (data provided by Dienst Justitiële Inrichtingen, Ministry of Justice, The Hague), of whom 151 were in hospitals that participated in the study, and thus entered our database. This means that the study covered 83.9% of our target population. After screening, 19 cases were discarded for the following reasons. (1) For 2 non-Dutch nationals who were repatriated directly after discharge, and 1 person who deceased shortly after discharge, it would be either unfeasible or meaningless to retrieve reconviction data. (2) Checklist ratings were sometimes returned with considerable time delay. It was decided to discard ratings received more than eight months after discharge ($n = 16$). Although this cut-off was to some extent arbitrary, it provided an optimum balance between inclusion of possibly less valid data on the one hand, and loss of reoffenders on the other. Of the remaining ratings, 70% were performed within 3 months after discharge. Cases that were omitted from the analyses did not differ significantly from the remaining cases with regard to treatment duration (319 vs. 256 weeks, ns) or age at discharge (33.6 vs. 34.8 years, ns). The omitted patients appeared to have relatively fewer diagnoses on axis 1 (7% had 2 codes or more, vs. 24.1% in the final sample), and relatively more diagnoses on axis 2 (7.1% only one diagnoses, vs. 19.9% in the final sample), but neither difference was significant. Finally, there also were no significant differences with regard to the offence type for which TBS was imposed.

The final sample thus comprised 132 cases, or 73.3% of our target population. Of these, 92.4% were male, 7.6% female. This is in accordance with the general TBS-population, which is over 90% male. TBS had been imposed for (mostly violent) property offences (e.g. robbery) in 12.9% of cases; for violent offences in 50.7% of cases; for sex offences in 19.7% of cases; and for arson in 16.7% of cases. At the beginning of treatment, 75.7% of patients were diagnosed with at least one personality disorder according to DSM-III or DSM-III-R (American Psychiatric Association 1980, 1987), mostly falling into the B-cluster (33.3%) or the Not Otherwise Specified category (32.6%). Also, 69.7% of patients had at least one Axis 1 disorder, with psychotic disorders (29.6%) and substance-related disorders (18.9%) being the most common primary diagnoses. The mean age at imposition of the TBS measure was 29.7 years (SD 8.3), while the mean age at discharge was 34.4 years (SD 8.1). Of the sample, 37.9% did not complete any education beyond primary school; 28.8% had lower-level secondary education, 10.6% medium-level secondary education, and 4.6% received higher professional or university education. The remaining

18.1% attended special schools for children with mental retardation or behavioral problems. Of the patients in the sample, 64.4% had received some form of mental health care in the past, which in 37.2% of the 132 cases included in-patient-care. Of the sample, 22.0% received some form of forensic mental health care in the past.

Length of stay in the hospital from which the patient was discharged varied from 1.2 to 11.4 years, with an average of 4.9 years. The vast majority of patients, 83.3% ($n = 110$), left the hospital on probationary leave. The remaining cases ($n = 22$) were discharged into the community because the court terminated the TBS-measure. This ruling was in accordance with hospital advice in only 4 cases, and went against hospital advice in the remaining 18 cases.

Raters

Forty-three different raters participated in the study. Of these, 34 were treatment supervisors: clinical psychologists or psychiatrists responsible for treatment planning as well as advising the court in TBS extension hearings. They were the preferred raters, because in daily practice they, too, are the ones who prepare risk reports, conduct risk assessments, and appear in court as expert witness. Similar to what happens in these situations, clinicians rated their own patients. Treatment supervisors provided over 90% of the ratings. The remaining nine raters were head nurses or psychotherapists, standing in for treatment supervisors who, for practical reasons, were unable to complete a particular rating.

Materials

Data on dynamic patient characteristics were collected using the Clinical Inventory of Dynamic Reoffending Risk Indicators (CIDRRI, Philipse, Koeter, van den Brink, & van der Staak, 2004). This rating scale was developed specifically for this study. It contains 47 statements on patient behavior, affect, skills, and clinical symptoms that were identified by clinicians from participating hospitals as pivotal to their risk management decision making. Examples of items are "Patient usually discusses others in negative terms"; "If patient encounters trouble, he calls in help"; "Patient actively uses opportunities to learn new things"; "patient is unable to imagine how others are feeling"; "Patient lacks essential social skills"; or "Patient is guided by current impulses." In terms of the definitions cited above, the items in the CIDRRI represent stable rather than acute dynamic risk factors. Items are rated on a six point scale, ranging from "not at all characteristic of the patient" to "very characteristic." At the end of the list the assessor is asked to rate the patient's risk of reoffending directly, again on a 6-point scale. No specific training or instruction is required: any clinician familiar with the patient can use the instrument.

The construction process of the CIDRRI, as well as its factor structure are described in detail in Philipse et al. (2004). CIDRRI items were condensed into six scales, as shown in Table 1. An earlier study showed the instrument to have acceptable interrater reliability, and to discriminate to an appreciable extent between patients just starting their treatment and those recently discharged (Philipse, Koeter, van der Staak, & van den Brink, 2005).

Table 1. Dynamic scales of the Clinical Inventory of Dynamic Reoffending Risk Indicators

Scale name and description	Cronbach α	single measure, ICC(95% CI) ^a
1. Empathic acceptance of responsibility for the offence. Patient acknowledges his responsibility for the offence and does so while truly aware of the impact of his actions on the victim(s)—11 items	.92	.52(.33–.67)
2. Self-reliance. Patient has sufficient skills to function acceptably in daily life without professional help—12 items	.87	.43 (.23–.60)
3. Anti-social narcissism. Patient has traits from narcissistic and anti-social personality disorders—10 items	.88	.67 (.52–.78)
4. Treatment compliance. Patient has engaged in the treatment process to the best of his abilities—5 items	.85	.57 (.40–.70)
5. Attainment of treatment goals. Treatment goals that are generally considered important, such as establishing an offence script or improving network conditions, have been achieved—7 items	.85	.56 (.38–.70)
6. Avoidance. Patient is inclined to stay away from others and to dislike intimacy—2 items	.60	.59 (.42–.72)

^aSingle measure intraclass correlation with 95% confidence interval.

Data on static risk factors were collected from hospital files by clerical workers, using an additional structured inventory called the File Checklist. It contained 41 items, most of which were directly derived from the contents of a national patient data management system that was developed in the early 1990's for use in forensic hospitals. The File Checklist covered demographic characteristics; DSM axes 1 and 2 diagnoses at admittance; data on intelligence, education and employment history; details of the TBS index offence; characteristics of TBS offence victims; criminal history; mental health history; irregularities during TBS; and personal circumstances at discharge.

Procedure

Data collection was coordinated by a research team, with a member in each participating hospital. These representatives kept track of discharges meeting inclusion criteria, and asked the responsible treatment supervisor to complete the CIDRRI. At the same time a clerical worker completed the File Checklist.

Outcome Measure

Reconviction data were coded dichotomously, as the presence or absence of any judicial sanction confirming the patient as the perpetrator of a new offence that involved sexual or other violence, including attempts at or threats of such violence. These offences are representative of the type of behavior that would generally be targeted in treatment and risk assessment in TBS. The criterion of a judicial sanction was added to assure that there was sufficient proof the subject had indeed committed the offence. Data on post-discharge reoffending were retrieved from the Centrale Justitiële Documentatie (Central Criminal Justice Files of the Ministry of Justice) on June 16, 2004.

Table 2. Overview of Static Risk Factors from the File Checklist that were Included in the Final Cox Regression Analyses.

DSM axis 1: Psychosis at time of admission (yes/no) (relates to HCR-20 items H6, "Major mental illness" and C3, "Active symptoms of MMI", and VRAG item 11, "Meets DSM-III criteria for schizophrenia")
DSM axis 1: substance use disorder at time of admission (yes/no) (relates to HCR-20 item H5, "Substance use problems", and VRAG item 3, "History of alcohol problems")
DSM axis 2:
A-cluster personality disorder (yes/no)
B-cluster personality disorder (yes/no)
C-cluster personality disorder (yes/no) (all three relate to HCR-20 item H9, "Personality disorder", and VRAG item 10, "Meets DSM-III criteria for any personality disorder")
Comorbidity of personality disorder and substance use disorder at time of admission (yes/no) (relates to the combination of HCR items H5 and H9, and VRAG items 3 and 10—see above)
Highest level of employment (7 categories) (relates to HCR-20 item H4, "Employment problems")
Any female victims in TBS index offence(s) (yes/no) (relates to VRAG item 9, "Any female victim")
Patient had been criminally violent before TBS-offence (yes/no) (relates to HCR-20 item H1, "Previous violence")
TBS-offence was first offence (yes/no) (relates to HCR-20 item H1)
Age at first conviction (years) (relates to HCR-20 item H2, "Young age at first violent incident; also related to VRAG item 7, "Age at index offence")
Number of institutional homes where patient lived before 18th year (3 categories) (relates to HCR-20 item H8, "Early maladjustment", and VRAG item 1, "Lived with both biological parents to age 16")
Number of times absent without leave during TBS (n) (relates to HCR-20 item H10, "Prior supervision failure", and VRAG item 6, "Failure on prior conditional release")
Arrangements for therapeutic aftercare after discharge have been made (yes/no) (relates to HCR-20 items R1, "Plans lack feasibility", R2, "Exposure to destabilizers", and R4, "Noncompliance with remediation attempts")
There is a regular source of income after discharge (yes/no) (relates to HCR-20 item R2, Exposure to destabilizers; and from LSI-r item 11, "Currently unemployed")

Statistical Analysis

Entering all 41 available static predictors into the analysis would compromise the meaningfulness of results due to chance capitalization. Instead, static predictors were selected for inclusion in the final analyses in two successive steps. First, we dropped all items from the File Checklist for which we did not find at least approximate equivalents in validated risk assessment instruments. The HCR-20 (Webster et al., 1997a), VRAG (Quinsey et al., 1998), and LSI-R (Andrews & Bonta, 1995) were used for reference, and fifteen items were retained; these are shown in Table 2. In the second step, univariate predictive validity of each variable was assessed in Cox regression analysis. Only items with univariate predictive validity at $p < .05$ were included in the final analysis.

In the final analysis, Cox regression (survival) analysis was used to develop the prediction model. Like other forms of regression analysis, it eliminates redundant predictors from the model and retains only items that each add significant predictive validity to the other predictors in the set. It has the advantage of also taking into account the variable times at risk among subjects.

Multivariate analysis was performed using forward stepwise entry of variables, with likelihood ratio significance testing. Significance levels for entering as well as remaining in the model were set at $p = .05$.

Predictors were entered in three successive blocks. First, the preselected static predictors were entered in Block 1, to establish the baseline multivariate static

prediction model. The six CIDDDRI-scales were then entered in Block 2. All scales were included irrespective of univariate significance, as the emphasis in this study was on finding dynamic predictors of risk. CIDDDRI scales that did not significantly add to the predictive value of the static predictor set or to other CIDDDRI scales, were removed from the model. Finally, separate constituent CIDDDRI items from discarded scales were entered in Block 3, but only if the item odds ratio was statistically significant at $p < .05$ in univariate Cox regression. This limitation was again set to prevent the introduction of an unfeasibly large number of variables into the analysis. The order in which variables were entered in successive blocks was deliberately chosen, guided by the rationale that historic predictors are an inevitable given, and always precede dynamic predictors in the here and now, providing the basis of any assessment.

To evaluate the predictive power of the ensuing final prediction model while avoiding effects of overfitting, an unweighted prediction score was computed by summing the rough scores on predictor variables. A Receiver Operating Characteristic (ROC) curve of sum scores was then plotted, and the area under the curve computed to establish predictive power. An ROC-curve shows the balance between false positive and false negative predictions at every cut-point in the assessment scale. The area under the curve (AUC) represents general predictive power, with 0.5 equalling non-prediction, 1.0 equalling perfect positive prediction, and 0.0 equalling perfect negative prediction.

RESULTS

Follow-up and Reconviction

At the point in time when reconviction data were retrieved, the last patient to enter the study had been discharged exactly 5.5 years ago, whereas follow-up for the first patient entering the study amounted to 8.5 years. Time at risk either to first reconviction or to the end of the study, corrected for periods of temporary readmission of some patients on probationary leave, varied from 71 days (2.4 months) to 3088 days (8.5 years), with a mean of 2272 days (6.2 years) and a median of 2493 days (6.8 years). The quickest relapse occurred 71 days after discharge, whereas the longest time-span between discharge and relapse was 2,177 days (5.9 years). The Kaplan–Meier Survival curve is shown in Fig. 1. A total of 26 dischargees (19.7%) were again convicted for a violent ($n = 21$) or sexual ($n = 5$) offence during follow-up. In 11 cases the new offence was similar to the offence for which TBS had been imposed.

Prediction of Reconviction

The direct risk assessment by clinicians on the 6-point scale at the end of the CIDDDRI did not have any predictive power with regard to actual reconviction for a violent or sexual offence (ROC AUC = .44; 95% CI .31–.56; ns).

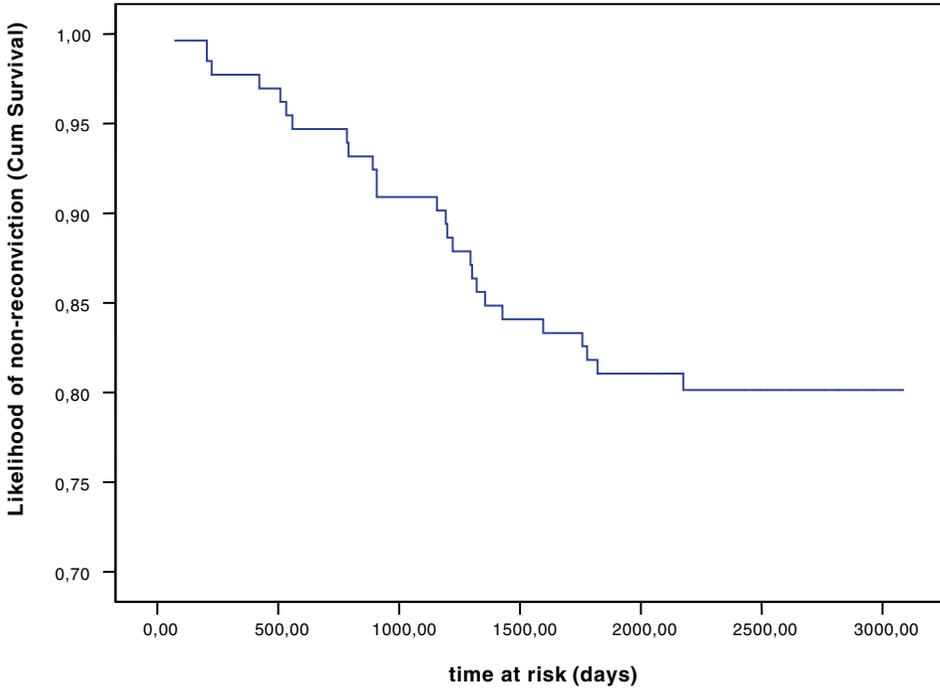


Fig. 1. Kaplan–Meier survivalcurve of non-reconviction for violent or sexual offence after discharge (n = 132).

Of the 15 static variables selected, only 6 proved significant predictors in univariate Cox regression analysis ($p < .05$), as shown in Table 3. Of these, four remained in the initial multivariate static prediction model.

None of the six dynamic scales added to the predictive power of the static model. Therefore, single CIDRR1 items were assessed for their univariate significance as a predictor ($p < .05$), to establish a selection for inclusion in the multivariate analysis. Only four items remained, these are again shown in Table 3. Interestingly, for all four items the direction of the relation between item scores and

Table 3. Overview of Risk Factors with Univariate Predictive Validity in Cox Regression Analysis

Static factors (file checklist)	β	p
DSM axis 1: Psychosis at time of admission	-1.530	.005
DSM axis 1: Substance use disorder at time of admission	0.930	.021
DSM axis 2: B-cluster personality disorder	0.843	.036
Comorbidity of personality disorder and substance use disorder at time of admission	1.149	.004
TBS-offence was first offence	-1.111	.017
Number of times absent without leave during TBS	0.467	.000
Dynamic items (CIDRR1)		
7. Avoids contact	-0.392	.023
8. Is unable to empathise with the victim's suffering	-0.271	.046
11. Completely denies his offence	-0.744	.018
36. Is unable to live on his own and take care of himself	-0.290	.023

Table 4. Prediction Model for Risk of Renewed Violent Offending Resulting from Multivariate Cox Regression Analysis with Forward LR Selection

Step	Predictor	β	Sig.	Odds ratio
1	Number of times absent without leave during tbs (count)	.422	.000	1.525
2	Comorbidity of any personality disorder with substance use disorder at time of admission (dichotomous score)	.981	.016	2.667
3	DSM axis 1: psychosis, at time of admission (dichotomous score)	-1.153	.037	.316
4	DSM axis 2: any cluster B personality disorder at time of admission (dichotomous score)	.818	.048	2.267

Note. Model $\chi^2(4) = 43.261; p < .001$.

reconviction ran counter to expectations: it appeared that higher levels of observed dysfunction were associated with *lower* risk of reconviction.

When these four dynamic items were entered stepwise into the analysis in addition to the already established static predictor set, none of them added significantly to the prediction of reconviction.

Thus, the final model, displayed in Table 4, was identical to the initial static prediction model. It shows that being absent without leave during TBS, comorbidity of axis 2 and substance use disorder at admission, and presence of cluster B personality disorder each increase the risk of reconviction, whereas presence of psychosis at admission reduces reconviction risk.

In order to assess the predictive power of this model, a simple sum score was computed. One point was added for each of the following: presence of cluster B personality disorder at admission; presence of comorbidity of personality disorder with substance related disorder at admission; and having been absent without leave at least once during TBS. One point was subtracted for presence of a psychotic disorder at admission. This unweighted score produced an ROC area under the curve for reconviction for renewed violent or sexual offending of .79 (95% CI .69–.89, $p < .001$), showing that the model, however succinct, possessed significant and substantial predictive power.

Post hoc Analysis

Because of the limited scope of the resulting prediction model, it was decided to conduct a post hoc analysis to establish whether findings might have been influenced by particular subgroups of patients, and a more extensive model might result if such a subgroup were deleted from the analysis.

To identify subgroups, reconviction rates were compared for patients with different TBS-offence types, different types of diagnoses, different modes of discharge and different modes of aftercare. It was found that patients with psychotic disorders at admission were reconvicted significantly less often than others ($\chi^2(1) = 9.201$, $p = .002$); patients with substance use disorders were reconvicted significantly more than others ($\chi^2(1) = 5.700$, $p = .017$), as were patients with cluster B personality disorders ($\chi^2(1) = 4.447$, $p = .035$). None of these findings provided new insights, as these characteristics are all part of the prediction model. However, it was also found that arsonists were reconvicted substantially more often than perpetrators of any

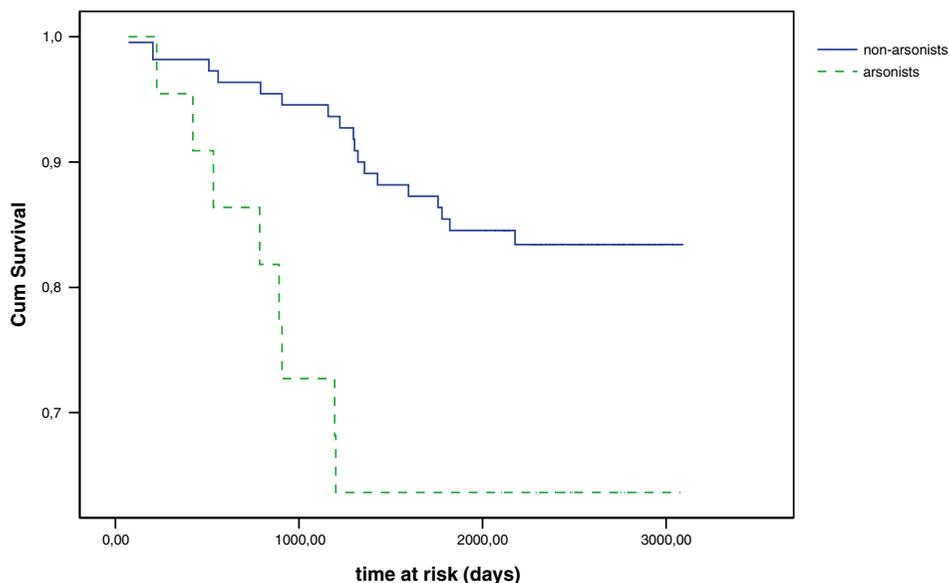


Fig. 2. Kaplan–Meier survivalcurves of non-reconviction for arsonists (convicted to tbs for arson, $n = 22$) vs. non-arsonists ($n = 110$).

other type of offence. Eight out of 22 arsonists were reconvicted (36.4%) as opposed to 18 out of 110 non-arsonists (16.4%) ($\chi^2(1) = 4.636, p = .03$). It is interesting to note that new offences by arsonists were of a violent or sexual nature rather than involving new arson. Kaplan Meier survival curves of arsonists versus non-arsonists are shown in Fig. 2.

On the basis of this finding, the Cox regression analysis was repeated for the sample excluding arsonists. This yielded a three factor prediction model that again lacked clinical salience. In this model two predictors from the main model, psychosis and Cluster B personality disorder, were replaced by a single CIDRRI item, 35: “Lacks essential social skills.” Like other dynamic variables in univariate analysis this one too had a “reversed” relation with reoffending: patients whose social skills were perceived as inadequate, were reconvicted less often than those who were seen as having good social skills. Absence without leave remained in the model with an odds ratio similar to that in the main model. Comorbidity of personality disorder with substance use disorder also remained, and had increased predictive power as compared to the main model.

An unweighted sum score was again computed by adding or subtracting the dichotomous scores on the three predictors. The 6-point rating on CIDRRI item 35 was dichotomized for this purpose by splitting scores between ratings 3 and 4. This new unweighted score, based on only three predictors, again had considerable predictive validity: the ROC AUC was .75 (95% CI .62–.88, $p < .01$). Nevertheless, the inclusion of a counterintuitive predictor (item 35) as well as the reduced ROC AUC left the main model for the full sample of patients unchallenged as the best prediction scheme derivable from the data.

DISCUSSION

The present study investigated whether clinically relevant, dynamic patient characteristics that are routinely regarded as reconviction risk indicators in forensic psychiatric settings, added significantly to the predictive power of static risk factors. The research was conducted under naturalistic conditions in a prospective setup. A four-predictor risk assessment model was found, comprising: absence without leave during TBS; comorbidity of axis 2 and substance use disorder at admission; Cluster B personality disorder; and psychosis at admission. This set predicted future reconvictions for sexual or violent offences with considerable accuracy, comparable to that of the VRAG, HCR-20 and PCL-r (Dolan & Doyle, 2000).

Regrettably, this model is fully static. Some dynamic factors were shown to have predictive power when considered separately, but proved redundant in a multivariate model including the static predictors. Interestingly, univariate Beta coefficients of these dynamic risk factors in all cases pointed in the opposite direction of that expected in clinical practice, suggesting substantial discrepancies between behavior observed on the outside by clinicians, and possible underlying drives and intentions of the patient. Such negative relations between clinically rated problem-levels and risk have been reported in several other studies as well (e.g., Harris et al., 2002; Hilton & Simmons, 2001).

As would be expected given these findings, direct clinical assessments of risk at the moment of discharge were completely unrelated to subsequent reconviction. Though this conclusion is unsurprising in the light of international research (Grove & Meehl, 1996), it had not been previously reproduced in the context of Dutch forensic psychiatry.

Implications for Clinical Risk Assessment

The dynamic variables included in this study were all directly derived from clinical practice in TBS (Philipse et al., 2004). They represented patient characteristics that were, and are, routinely used by clinicians to assess the risk of future reoffending, and to identify focus points for treatment. Though the data were gathered 6–9 years ago, the items that were tested continue to be cited by TBS clinicians in risk assessment reports today. Moreover, many of them are, in some form or another, also part of the definition of dynamic items in the HCR-20 or the LSI-R. The finding that none of these variables added any predictive power to a handful of static predictors is therefore worrying. It suggests that clinicians need to be very careful when considering dynamic patient characteristics, such as were investigated in this study, as part of a risk assessment. As Webster, Douglas, Eaves, and Hart (1997b) have noted, “historical variables deserve a position of primacy in any scheme used in attempts to assess violence potential in persons with psychiatric disorders” (p. 256).

These findings do not, in our view, change the fact that forensic mental health treatment without dynamic risk factors is essentially a contradiction in terms. Absence of dynamic predictors renders such treatment futile and reduces the role of the clinician to that of gatekeeper. Our findings do however reinforce the impression

that clinical assumptions regarding (durable) changeability of risk are overly optimistic. Stronger reliance on static, actuarial risk factors may be advisable, yet leaves the clinical treatment dilemma unsolved. Moreover, static risk indicators carry significant margins of error, too. False positive predictions resulting from exclusive reliance on such predictors will inevitably lead to long-term, wrongful detention of a considerable percentage of patients. Possibilities of investigating dynamic risk assessment have by no means been exhausted; the implications of static risk are too far-reaching to abandon the search for dynamic alternatives in a hurry. With regard to the present study, several ways forward are imaginable.

Earlier findings with regard to CIDRRI interrater reliability were sufficiently satisfactory to render it unlikely that predictive validity of dynamic factors was grossly underestimated due to reliability problems (Philipse et al., 2005). Nevertheless, dynamic risk factors are more complex than static ones, and more difficult to assess (Quinsey et al., 1998). Even when items like those in the CIDRRI are scored reliably, a scoring procedure designed to allow easy use in the busy everyday practice may incur the risk of only skimming the surface. Here lies another possible explanation for the counterintuitive direction of predictive relationships between dynamic items and outcome in univariate analysis (for instance, patients denying their offence are reconvicted less often). It may indicate that adequate risk management strategies are put into place for patients with obvious areas of dysfunction, while risk management for patients who project an image of adequate functioning relies too heavily on their apparent abilities. In this context, psychopathy may be of crucial influence. For instance, a patient may seem very contrite with regard to his offence, but may in fact be faking this sentiment because he knows it will enhance his chances of discharge. Seto & Barbaree (1999) detected such deceptive good behavior in psychopaths who received treatment.

Unfortunately no Dutch version of the PCL-r was available at the time of data collection, so that we were not able to test the effect of psychopathy on predictor-outcome relations. Though we can therefore not be sure that such mechanisms are at work in the present study, the CIDRRI relies exclusively on the clinician's individual perceptions of the patient, and is thus vulnerable to "impression management." In this respect CIDRRI performance might improve if items are operationalized in detail, and if judgments are based on systematic observation and formal psychological assessments, and involve multiple raters in the process.

One may also wonder whether dynamic characteristics would have performed better had they been rated by an assessor who does not have a treatment relation with the patient. Webster et al. (1997a) suggest that raters who are involved in the treatment process may introduce professional bias; for instance, they may be overly eager to see improvement. Clinicians' feelings towards patients have been shown to influence their risk assessments (De Vogel, 2005; Dernevik et al. 2001; Hilton & Simmons, 2001). However, De Vogel also studied whether the clinical and risk management domains of the HCR-20 had different predictive validity depending on who rated the items. Ratings by researchers who were uninvolved in the patient's treatment, were compared to those by treatment supervisors, who bear final responsibility for the treatment process. This study, conducted prospectively in a TBS setting, yielded no significant differences between the rater groups. Thus,

it remains the question whether a different choice of raters in the present study would have yielded significantly different results.

It should be noted with regard to the solutions suggested here, that a midway needs to be found between the ease of use of the CIDRRRI and the extreme demands on resources made by some actuarial methods. If detection of a handful of dynamic predictors implies loss of practical feasibility of the instrument, we are merely replacing one problem with another. However, policy makers and clinicians may also need to adapt to the idea that improved solutions to the risk assessment problem imply structural and extensive investments of time and money. Such investments will no doubt be made more willingly if the instrument at hand offers starting points for clinical treatment rather than a mere indication of the required security level.

Research Considerations and Study Limitations

The benefit of a hospital-based, prospective research procedure is that it allows insight into an instrument's validity under the circumstances of use for which it is intended. In this context, in the present study only a very small set of predictors survived in the final model, and even several retrospectively well-established predictors were discarded (for instance, age at first conviction; first offender or not). The final prediction model was fairly powerful, though its predictive accuracy would be expected to decline somewhat in a new sample.

Three methodological issues regarding the present study warrant brief consideration. First, a number of checklists ($n = 39$) were only returned to the researchers several months after the patient had actually left the hospital. It is of course possible that these checklists were rated based on imperfect recollections of the patient, and thus affected predictive validity. Yet, it could also be argued that these late returns in fact worked in favor of dynamic predictors: the clinician-rater may have benefited from current information about the patient's functioning in the community context.

Second, like nearly every risk assessment study the present study is hampered by the reoffending "dark number." There can be no doubt that the reconvictions for violent and sexual crimes used as prediction criterion in this study represent only part of the true amount of reoffending. Crimes may never come to the attention of the police, and others remain unsolved. Research setups that solve this problem through intensive follow-up with self-report and collateral interviews are rarely feasible, if only because they require willingness on the patient's behalf to keep in close contact with the forensic system after discharge. This will inevitably result in a select sample. Whether such labor-intensive solutions are at all necessary is not sure. Findings from the Cambridge Study on Criminal Careers (Farrington, 2001) have shown self-reported criminality and officially recorded criminality to result in similar predictor sets. Therefore, though the problem of dark numbers appears in essence to be insoluble, it may not pose as great a problem to file-based follow-up research as is sometimes assumed.

Third, predictors were validated against a relatively low base rate occurrence, a renewed judicial sanction for a violent or sexual offence. This inevitably reduced the power of our study. Yet, we preferred this definition of recidivism to a relapse criterion that would give us a higher base rate, but would be farther removed from the

type and seriousness of violent behavior clinicians would consider when assessing risk and advising the court.

To conclude, it should be noted that in recent years important changes have occurred which are likely to have improved risk assessment practice in TBS already. This is one drawback of prospective research: due to the time needed, developments in the field can easily overtake it. Notably, from 2000 onwards Dutch forensic hospitals have seen the introduction of standardized risk assessment tools as well as the PCL-r, by now mandatory input in any risk assessment. However, it should not be forgotten that long-term prospective, hospital based validation of these instruments as predictors of post-discharge recidivism is still largely absent even internationally, and non-existent in the Netherlands. Whether they have added value as compared to a compact static prediction model like the one presented in this article, remains to be seen.

The findings from our study suggest that observational and clinically interpreted input may not be a good basis for risk assessment, and that other sources of information need to be considered. Future research into dynamic predictors of reoffending may need to focus on new measures that are less susceptible to manipulation by the patient or to clinical observation and evaluation bias. Experimental performance tasks and psycho-physiological measures are alternatives that warrant closer investigation and could well provide ways forward.

ACKNOWLEDGMENTS

The authors wish to thank management and staff of the following institutions for their generous participation in the study: Forensic Psychiatric Clinic of GGz Drenthe, Assen; Veldzicht Forensic Psychiatric Centre, Balkbrug; Hoeve Boschcoörd, Boschcoörd; Forensic Psychiatric Clinic of GGz Eindhoven; Dr. S. van Mesdag Hospital, Groningen; Oldenkotte Forensic Psychiatric Centre, Rekken; and Dr. H. van der Hoeven Hospital, Utrecht. The authors further wish to thank Dr. Marinus Spreen and Dr. Nol van Gemmert for their assistance in retrieving additional data, and anonymous reviewers for their helpful comments on earlier drafts of this paper.

REFERENCES

- American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders*. (3rd ed.). Washington, DC: APA.
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders*. (3rd ed. revised). Washington, DC: APA.
- Andreasen, N. C. (2000). Prediction in clinical psychiatry. What does the future hold? *American Journal of Psychiatry*, *157*(9), 1373–1374.
- Andrews, D. A., & Bonta, J. L. (1995). *The level of service inventory—revised (LSI-R)*. Toronto: Multi-Health Systems.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1995). *Manual for the sexual violence risk-20. Professional guidelines for assessing risk of sexual violence*. Vancouver: Simon Fraser University.
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, *123*, 123–142.
- Buchanan, A. (1999). Risk and dangerousness. *Psychological Medicine*, *29*, 465–473.

- De Vogel, V. (2005). Structured risk assessment of (sexual) violence in forensic clinical practice. Doctoral thesis. Amsterdam: Dutch University Press.
- Dempster, R., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse, 14*, 121–138.
- 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. McMurrin (eds.), *Sex and violence. The psychology of crime and risk assessment* (pp. 83–101). London: Routledge.
- Dolan, M., & Doyle, M. (2000). Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. *The British Journal of Psychiatry, 177*, 303–311.
- Farrington, D. P. (2001). What has been learned from self-reports about criminal careers and the causes of offending? Report for the Home Office. London: Home Office.
- Gagliardi, G. J., Lovell, D., Peterson, P. D., & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior, 28*(2), 133–155.
- Gardner, W., Lidz, C. D., Mulvey, E. P., & Shaw, E. C. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses. *Law and Human Behavior, 20*, 35–48.
- Grann, M. (1998). Personality disorder and violent criminality. A follow-up study with special reference to psychopathy and risk assessment. Doctoral thesis. Stockholm: Karolinska Institutet.
- 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.
- 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*(1), 19–30.
- Grubin, D. (1999). Actuarial and clinical assessment of risk in sex offenders. *Journal of Interpersonal Violence, 14*(3), 331–343.
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*(1), 6–35.
- Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse. A Journal of Research and Treatment, 13*(2), 105–122.
- Hare, R. D. (1991). *The psychopathy checklist—revised*. Toronto: Multi-Health Systems.
- Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior, 23*(1), 25–54.
- Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. *Annals of the New York Academy of Sciences, 989*, 198–210.
- Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*(4), 377–394.
- Hart, S. D. (1999). Assessing violence risk. thoughts and second thoughts. *Contemporary Psychology, 44*(6), 486–487.
- Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*(4), 393–408.
- Kraemer, H., Kazdin, A., Offord, D., Kessler, R., Jensen, P., & Kupfer, D. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337.
- Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471–489.
- Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*(2), 409–443.
- Litwack, T. R. (2002). Some questions for the field of violence risk assessment and forensic mental health or, “Back to basics” revisited. *International Journal of Forensic Mental Health, 1*(2), 171–178.
- Margison, F. R., Barkham, M., Evans, C., McGrath, G., Mellor Clark, J., Audin, K., et al. (2000). Measurement and psychotherapy. Evidence-based practice and practice-based evidence. *The British Journal of Psychiatry, 177*, 123–130.
- Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark, Robbins P., Mulvey, P. E., et al. (2001). *Rethinking risk assessment. The MacArthur study of mental disorder and violence*. Oxford: Oxford University Press.
- Mossman, D. (1994). Assessing predictions of violence. Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783–792.
- Philipse, M. W. G., Koeter, M. W. J., van den Brink, W., & van der Staak, C. P. F. (2004). The structural coherence of clinically derived dynamic indicators of reoffending risk. *Criminal Behaviour and Mental Health, 14*, 263–279.

- Philipse, M. W. G., Koeter, M. W. J., van der Staak, C. P. F. & van den Brink, W. (2005). Reliability and discriminant validity of dynamic reoffending risk indicators in forensic clinical practice. *Criminal Justice and Behavior*, 32(6), 643–664.
- Quinsey, V. L., Harris, G. T., Rice, M. E., Cormier, C. A. (1998). *Violent offenders. Appraising and managing risk*. Washington, DC: American Psychological Association.
- Seto, M., & Barbaree, H. (1999). Psychopathy, treatment behavior, and sex offender recidivism. *Journal of Interpersonal Violence*, 14, 1235–1248.
- Sturidsson, K., Haggård-Grann, U., Lotterber, M., Dernevik, M., & Grann, M. (2004). Clinicians' perceptions of which factors increase or decrease the risk of violence among forensic out-patients. *International Journal of Forensic Mental Health*, 3(1), 23–36.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997a). *HCR-20. Assessing risk for violence (version 2)*. Vancouver: Mental Health, Law & Policy Institute, Simon Fraser University.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997b). Assessing risk of violence to others. In C. D. Webster & M.A. Jackson (Eds.), *Impulsivity. Theory, assessment and treatment* (pp. 251–277). New York: Guilford Press.
- Webster, C. D., Müller-Isberner, R., & Fransson, G. (2002). Violence risk assessment: Using structured professional guidelines professionally. *International Journal of Forensic Mental Health*, 1(2), 43–51.