Eur Arch Psychiatry Clin Neurosci (2007) 257:129–134 DOI 10.1007/s00406-006-0678-y

ORIGINAL PAPER

Frank Urbaniok · Jérôme Endrass · Astrid Rossegger · Thomas Noll · William T. Gallo · Jules Angst

The prediction of criminal recidivism

The implication of sampling in prognostic models

Received: 2 February 2006 / Accepted: 13 June 2006 / Published online: 14 September 2006

Abstract *Background* Instruments based on actuarial forensic risk assessment models are sensitive to the calibration sample, and the inclusion criteria for the subjects of a study population will influence the features of the resulting model. If the same instrument is used in populations that are not part of the calibration sample, the discriminative validity of the instrument is jeopardized; thus the definition of the calibration sample is very important. The aim of this study was to examine how sensitive prognostic models are to the calibration sample. *Method* Two samples ($N = 773$) of offenders sentenced to at least 10 months in prison for a violent or sexual offense were used in this study. The "sanction sample" (recruited during August 2000, $N = 515$) consisted of all violent and sexual offenders actively administrated by the Criminal Justice System of Zurich, Switzerland. The "verdict sample" (recruited over two years, $N = 258$) included all offenders convicted in the Canton of Zurich during a two-year period. Both samples were unbiased, since all subjects that met the study criteria were included. In the first analysis, differences between the two samples were shown with respect to socio-demographic, criminological, and psychiatric variables using bivariate logistic regressions. In the second analysis, recidivism was estimated separately for both samples, using a logistic regression model as a function of a set of psychiatric, socio-demographic and criminological variables. *Results* Bivariate logistic regression showed that different risk factors for recidivism existed for both samples. *Conclusion* Forensic risk assessment models are very sensitive to the calibration sample. There is strong evidence that, even when index-offenses and the socio-cultural background are the same, risk factors for recidivism differ depending on the stage of the judicial process in which the subjects are (e.g. whether a subject is indicted, on conditional release, on parole, or no longer under the supervision of a parole board). Unfortunately, none of the currently available actuarial risk assessment instruments that have been validated in European countries consider the different stages of the judiciary process.

Key words risk assessment · actuarial instruments · validity

F. Urbaniok · J. Endrass · A. Rossegger · T. Noll
Psychiatric-Psychological Service
Zurich Department of Justice
Zurich, Switzerland

W.T. Gallo
Yale University School of Medicine
New Haven, CT, USA

J. Angst
Department of Psychiatry
University of Zurich
Zurich, Switzerland

A. Rossegger (✉)
Psychiatrisch-Psychologischer Dienst
Justizvollzug Kanton Zürich
Feldstrasse 42
8090 Zürich, Switzerland

Introduction

Forensic risk assessment plays an important role in law enforcement and the criminal justice system and can be performed during different stages of the judicial process: at the very beginning of this process, when a person is reported, a short-term prognostic assessment may be necessary; it is common, in many countries, for forensic expert opinions to be sought after an individual has been indicted; and after the verdict, prognostic assessment can influence the decisions of the board of parole, e.g. whether or not a subject can be released from prison [1].

The principal methods of assessing the risk for reoffending are the intuitive (non-standardized, and using clinical experience), the clinical (mainly using

structured clinical guides), and the statistical (actuarial risk assessments). Actuarial methods are based on the observation of offenders over a period of time and comparing recidivists to non-recidivists; this makes all actuarial instruments evidence-based [2].

In Anglo-Saxon countries, the actuarial method predominates, supported by the results of numerous studies indicating the greater accuracy of statistically developed methods of assessment over non-statistically generated methods (e.g. [3, 5, 7, 9, 12, 15, 16]). Examples of state-of-the-art actuarial risk assessment instruments are: the Violence Risk Appraisal Guide (VRAG) [10], and the Static-99 [6]. However, recent publications suggest that the better results of the actuarial methods can be explained with poor standardization of most clinical approaches [4].

Regardless of their accuracy, actuarial models produce sample-specific estimates of risk. Since coefficients estimated with a statistical model depend on the composition of the study sample, statistical models of recidivism cannot be generalized to other populations, and are, strictly speaking, only valid for specific populations. For instance, it would be inappropriate to use a sample of non-violent, white-collar criminals to validate a risk assessment tool developed on a sample of violent offenders. Another example of inappropriate generalization would be to assess the risk of recidivism in a sample of prison inmates using a risk assessment instrument that was developed in a forensic psychiatric institution and would thus contain many more subjects with mental disorders. This means that a statistically generated risk assessment instrument that has been developed in Anglo-Saxon countries must be validated for another language area (e.g. in a German speaking area) [17, 18, 20], before it can be used outside Anglo-Saxon countries. If there is no careful validation study, the instrument is not suitable to assess the risk for recidivism.

During recent years, there has been a growing interest in European countries for unbiased and scientifically validated prediction of sexual and violent reoffending; however, most of the statistical risk assessment instruments have been developed in Canada. Before applying an actuarial risk assessment instrument in Europe, the validity of this instrument for the population of interest must be demonstrated. In the German language area, for example, there are few validation studies of risk assessment instruments. Exceptions to this are the studies by Stadtland and Nedopil [14], Dahle [4], and Urbaniok et al. [19]. Further validation studies are needed.

The following question therefore arises: What are characteristics of a population that have to be taken into account? For example, what are the implications for the application of actuarial instruments knowing that Stadtland and Nedopil [14] as well as Urbaniok et al. [19] conducted their validation studies among offenders for whom a forensic expert opinion had been sought (before being convicted), while Dahle [4], by contrast, examined offenders released from prison, and Soyka et al. [13] patients discharged from a psychiatric ward?

In clinical epidemiology, important characteristics of the study sample are specific causative agents in a particular region affecting a specific population (for an example of an epidemiological study among offenders see Repo-Tiihonen et al. [11]). Among forensic calibration samples, one other important characteristic is the current phase of the judicial process the offenders are in when a prognosis is asked for.

In most constitutional states, there are numerous stages of judiciary processing (report and preliminary investigation, indictment, verdict, and sanction by the judicial authorities). For forensic psychiatric purposes, these phases can be roughly classified into two more general phases where risk assessment plays an important role. The first phase lasts until sentencing (in the form of psychiatric expert opinions), and the second consists of the time during the process of administration by the legal authorities (prognosis for release or graduated enforcement schemes). A sample for the statistical validation of a prognosis instrument will have different characteristics, depending on during which of the two phases it is drawn. If a study population consists of subjects recruited at the time of their verdict, every convicted offender has the same probability of being included in the sample; if for example a verdict sample is drawn annually, each offender would appear only once (per verdict) in a particular year. This is different when drawing the sample after conviction (e.g. offenders administrated by the correctional service). Offenders with longer sentences (mostly the more dangerous offenders) have a distinctly higher probability of being included in a sanction sample; if such a sample was drawn annually, an offender with a sentence of 20 years would appear twenty times, whereas an offender with a sentence of two years would do so only twice. Both offenders, however, would appear only once in a verdict sample.

Assuming that each of these phases characterizes a specific population, and prognostic models are sensitive to the calibration sample, the two phases should be interpreted as limitations for the risk model. This means that the validity of a risk assessment instrument has to be tested for each.

It is common among researchers in the field of forensic psychiatry to pose the question of whether a model is accurate for sexual and violent offenders, but not whether it is equally valid for a verdict sample and a sanction sample.

The aim of this study was to estimate statistical models for recidivism in two different samples that differed with respect to judicial status (a verdict sample and a sanction sample) but were comparable concerning significant socio-economic variables.

We hypothesized that the estimation of a statistical model for recidivism of an offender-population,

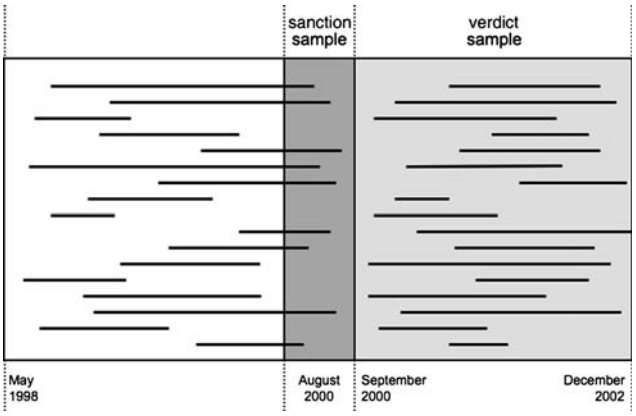

Fig. 1 Proportion of short term sentences and long term sentences in the two samples

which is currently actively administrated by the judicial authorities (sanction sample), would differ substantially from the model of an offender population that included all sentenced individuals over a specific time period (verdict sample).

Method

Subjects

All subjects ($N = 773$) were male and convicted of committing a sexual or violent offense in the State (Canton) of Zurich, Switzerland, an urban and pre-alpine area with a population of approximately 1,200,000. All subjects were sentenced to at least 10 months in jail or prison.

Two samples were investigated. The "verdict sample" comprised all individuals (258 subjects) sentenced between September 2000 and December 2002 for violent or sexual offenses in the Canton of Zurich. An offender was immediately included in this sample upon conviction. The "sanction sample" comprised all 515 convicts administered by the Justice Department of the Canton of Zurich in August 2000. In the sanction sample, a violent or sex offender was included if actively administered for being in jail, on parole, on probation, or in forensic therapy (inpatient or outpatient).

The chances of being included in the verdict sample were equal for all convicted offenders. In contrast, not all offenders with a sanction had the same probability of being part of the sanction sample. For instance as mentioned before, the likelihood of being registered in the month of the census was higher for offenders with long sentences or forensic therapies than for delinquents with short sanctions and no therapy (Fig. 1).

Measures

A series of psychiatric, psychological, criminological, and socio-economic variables were assessed on the basis of the files of the convicts. The main source of recidivism–data was the nationally registered penal records, where convictions–but not charges–appear. The record contains date of conviction, type of offense, length of sentence, reprieve, expulsion from the country, date of conditional release, and duration of probation. Not listed are the exact date of the beginning of incarceration, date of non-conditional release and name of prison.

In a pilot study, the interrater agreement was computed for the $N = 60$ variables yielding alpha > 0.65. The interrater reliability was assessed with Krippendorff's alpha [8]. The advantage of this measure is that it can be used to analyze the agreement of multiple raters, even if there are unequal sample sizes or missing data, and it can be computed when the variables are nominal, ordinal, or continuous. The variables are listed in the results.

Statistical analysis

To assess the differences between the compositions of the two samples, bivariate logistic regression analyses were carried out using the sample-identifier as the dependent variable.

In a second analysis, stratified for each sample, bivariate logistic regression analyses were computed to estimate the association between the selected criteria and recidivism.

All beta-weights of the logistic regression models were exponentiated to e in order to obtain odds ratios. An odds ratio (OR) of 1 indicates that an event under study is equally likely among recidivists and non-recidivists. An OR > 1 indicates that the event is more likely among recidivists. An OR < 1 means that the event is less likely among recidivists and can thus be interpreted as a protective factor (for recidivism). All logistic models were computed using STATA 8.2 SE.

Results

Differences between the verdict sample and the sanction sample

As hypothesized, the logistic regression shows that the two samples differ substantially, and thus the assumption that there is a need to test the sensitivity of prognostic models to the judicial process was confirmed. Subjects who lived in a foster home before the age of 15 (OR = 1.8) were more likely to be in the sanction sample. Other important socio-demographic variables such as age, education, having a child, being married, and/or being a foreigner living in Switzerland did not differ between the two samples (Table 1).

Subjects convicted for committing arson at home (OR = 0.6) were less likely to be in the sanction sample. Offenders who committed severe crimes such as homicide (OR = 3.9) and robbery (OR = 1.8), as well as subjects suffering from delusional symptoms at the time of the offense (OR = 4.1), were more likely to be among the sanction sample. Other offense-related variables, such as whether the offenses were "endangering human life", "physical assault", "sexual abuse" or "rape", and whether the victim was a family member or unknown, did not differ between the two samples.

Subjects from the sanction sample were twice as likely to be recidivists (OR = 2.0) and to have been in forensic psychiatric treatment (OR = 2.0). Furthermore, those subjects were more likely to have a history of alcohol and drug abuse (OR = 1.9), were more likely to have been hospitalized in a psychiatric clinic (OR = 1.8), and were three times more likely to undergo antipsychotic treatment (OR = 3.1). Additionally, there were more subjects with a history of alcohol abuse or psychiatric disorder in the nuclear family (OR = 2.1) within the sanction sample (Table 2).

Table 1 Differences between the two populations: Bivariate logistic regression for socio-demographic and biographic variables

Variable	Verdict	Sanction	OR	SE	*P*	95% CI	
Age (1 unit = year)	X = 32.72 SD = 10.98	X = 33.49 SD = 10.55	1.007	0.007	0.343	0.993	1.021
Swiss national	135 *N* = 265 52.73%	267 *N* = 514) 51.95%	0.969	0.148	0.836	0.718	1.308
Lived in a foster home before the age of 15	26 *N* = 216 12.03%	86 *N* = 433) 19.86%	1.811	0.437	0.014*	1.129	2.906
Completed school	159 *N* = 226 70.35%	282 *N* = 445) 63.37%	0.729	0.128	0.072	0.517	1.029
Vocational education	129 *N* = 239 53.97%	222 *N* = 466) 47.64%	0.776	0.124	0.112	0.568	1.061
Married	67 *N* = 255 26.27%	127 *N* = 506) 25.10%	0.940	0.165	0.725	0.667	1.326

Note:
Verdict = Verdict sample
Sanction = Sanction sample
OR = Odds Ratio, SE = Standard-Error, 95% CI = 95% confidence interval
*$P \leq 0.05$

Stratified estimation of the risk for recidivism

The results of the bivariate logistic regression show that subjects in both samples were more likely to re-offend if their victim was a stranger (verdict sample: OR = 2.9, $P = 0.002$, sanction sample: OR = 2.131, $P = 0.000$).

Other features were either predictive for one group or the other, resulting in different risk-assessment models.

For the verdict sample (but not for the sanction sample), a history of alcohol or drug abuse (OR = 4.6, $P = 0.000$) was a strong predictor for reoffending. In addition, actual or past intimacy between victim and offender (OR = 0.3, $P = 0.010$) was a predictive protective factor.

For the sanction sample (but not for the verdict sample), delusional symptoms at the time of offense (OR = 0.4, $P = 0.023$), and being married (OR = 0.9, $P = 0.002$), were strong protective factors against re-offending. Swiss citizenship (OR = 2.5, $P = 0.000$), and living in a foster home before the age of 15 (OR = 3.6, $P = 0.000$), were strong risk factors for reoffending (Table 3).

Discussion

The aim of this study was to empirically demonstrate and emphasize the importance of selecting the appropriate sample when validating or developing risk assessment tools.

We based our study on two different samples of sexual and violent offenders. The inclusion criteria for both samples were the same, but were carried out at different phases of the judiciary process. The verdict sample consisted of all convicted offenders in a specific time period, and the sanction sample of all offenders administrated by the criminal justice system at a specific time.

The assumption that the samples would differ was confirmed. The sanction and verdict samples drawn from the same population differed substantially in mental health and criminological variables, whereas the composition of the samples was comparable with respect to socio-demographic variables. Therefore the assumption that the effect of sampling needs to be tested in forensic settings due to the stages of the judicial process was justified.

A series of bivariate logistic regression analyses demonstrated that there were some criteria that were predictively valid for both samples, but many results were significant only for one sample and meaningless for the other. Being in a foster home before the age of 15 increased the risk for recidivism in the sanction sample by about 260%, but was not a significant risk factor in the verdict sample. Then again, a history of alcohol or drug abuse increased the risk for recidivism in the verdict sample by about 360%, but was not a significant risk factor in the sanction sample. By contrast, having an unknown victim seems to be a robust/stable risk factor; having an unknown victim was predictive in both groups, increasing the risk of recidivism by approximately 190% in the verdict sample and 110% in the sanction sample.

Conclusion

What are the implications of these findings? The findings lead us to the conclusion that characteris-

Table 2 Differences between the two populations: Bivariate logistic regression for offense related and forensic psychiatric variables

Variable	Verdict	Sanction	OR	SE	*P*	95% CI	
Homicide	22 *N* = 257 08.56%	138 *N* = 515 26.80%	3.910	0.955	0.000*	2.423	6.310
Exposure to loss of life	13 *N* = 257 05.06%	18 *N* = 515 03.50%	0.680	0.253	0.300	0.328	1.410
Physical assault	24 *N* = 257 09.34%	48 *N* = 515 09.32%	0.998	0.262	0.993	0.596	1.669
Robbery	32 *N* = 269 11.90%	48 *N* = 515 09.32%	1.778	0.399	0.010*	1.146	2.760
Arsonist	24 *N* = 257 09.34%	28 *N* = 515 05.44%	0.558	0.162	0.044*	0.317	0.984
Sexual abuse	69 *N* = 257 25.85%	72 *N* = 515 13.98%	1.836	0.745	0.134	0.829	4.065
Rape	41 *N* = 257 15.95%	69 *N* = 515 13.40%	0.815	0.174	0.339	0.536	1.240
Domestic crime scene	97 *N* = 248 39.11%	131 *N* = 496 26.41%	0.559	0.092	0.000*	0.404	0.772
Delusional at the time of offense	06 *N* = 234 02.56%	46 *N* = 473 09.73%	4.094	1.808	0.001*	1.722	9.730
Victim was family member or partner	45 *N* = 255 17.65%	413 *N* = 498 17.07%	0.960	0.195	0.842	0.645	1.430
Unknown victim	89 *N* = 255 34.90%	180 *N* = 498 36.14%	1.056	0.170	0.736	0.770	1.448
Criminal record	140 *N* = 255 54.90%	355 *N* = 511 69.47%	1.869	0.296	0.000*	1.371	2.549
Previous forensic therapy	36 *N* = 255 14.12%	125 *N* = 506 24.70%	1.996	0.414	0.001*	1.329	2.996
Previous hospitalized in a psychiatric clinic	49 *N* = 222 22.07%	155 *N* = 454 33.48%	1.777	0.339	0.003*	1.224	2.581
Previous treated for schizophrenia	06 *N* = 223 02.69%	35 *N* = 444 07.88%	3.095	1.392	0.012*	1.282	7.473
History of alcohol or drug abuse	86 *N* = 225 38.22%	241 *N* = 445 54.16%	1.909	0.319	0.000*	1.377	2.649
Family history of alcohol abuse or psychiatric disorder	38 *N* = 165 23.03%	147 *N* = 384 38.28%	2.073	0.441	0.001*	1.366	3.145

Note:
Verdict = Verdict sample
Sanction = Sanction sample
OR = Odds Ratio, SE = Standard-Error, 95%CI = 95% confidence interval
*$P \leq 0.05$

tics of the sample defined through the judicial process set close limits on the possibility of generalizing the prognostic results of a statistically generated risk assessment tool. The validation of an instrument for a sanction sample does not automatically suggest predictive quality for a verdict sample.

There is a wide consensus that separate prognostic models must be developed and validated for different categories of offenses (e.g. violent offenses, white-collar crime, etc.). Evidence from validation studies of risk assessment tools indicates that actuarial risk assessment instruments must be specifically designed for a target population (e.g., convicts in the Canton of

Table 3 Predicting pertinent criminal record: Bivariate logistic regression for each population

Variable	Verdict Sample					Sanction Sample				
	OR	SE	*P*	95% CI		OR	SE	*P*	95% CI	
Lived in a foster home before the age of 15	1.376	0.692	0.525	0.514	3.685	3.632	0.913	0.000*	2.219	5.944
Completed school	0.962	0.367	0.919	0.456	2.030	0.811	0.166	0.305	0.543	1.210
Vocational Education	0.726	0.247	0.347	0.372	1.415	0.940	0.182	0.749	0.643	1.374
Married	0.736	0.287	0.431	0.343	1.580	0.487	0.115	0.002*	0.306	0.775
Swiss national	1.380	0.456	0.329	0.723	2.637	2.531	0.491	0.000*	1.730	3.701
Delusional at the time of offense	0.886	0.982	0.913	0.101	7.780	0.417	0.160	0.023*	0.196	0.886
History of alcohol or drug abuse	4.624	1.678	0.000*	2.270	9.418	1.358	0.272	0.127	0.917	2.010
Victim was partner	0.279	0.139	0.010*	0.105	0.739	0.681	0.221	0.237	0.361	1.287
Unknown victim	2.897	0.973	0.002	1.500	5.594	2.131	0.415	0.000*	1.454	3.122

Note:
OR = Odds Ratio, SE = Standard-Error, 95% CI = 95% confidence interval
* $P \leq 0.05$

Zurich) [20] in order to give accurate estimates of recidivism risks. However, there is limited research that aims to identify specific sub-populations defined by the criminal justice process. Since the samples we used were comparable with respect to socio-demographic variables, there is evidence that forensically relevant differences can be attributed to different stages of the criminal justice process.

If actuarial risk-assessment instruments are developed without taking the population's stage in the criminal justice process into consideration (e.g. a verdict versus a sanction population), the discriminative validity could be jeopardized at the onset of the investigation. Furthermore, the application of an actuarial risk assessment instrument that has not been carefully validated could lead to severe misjudgments–at the expense of the civil rights of the individual or the safety of society. Unfortunately, there is currently no actuarial risk assessment instrument available that has been validated in the German language area and takes the criminal justice process into consideration.

Showing the importance of the stage of the judicial process highlights the population problem that each actuarial risk assessment tool faces. To minimize misleading conclusions, future research and experts using risk assessment tools should pay special attention to the weaknesses of actuarial risk assessment tools.

References

1. Albrecht PA (1993) Multidimensional personality assessment in penal measures. An evaluation questionnaire as the deciding factor of dangerousness prognosis. Psychiatr Prax 20:9–14
2. Andrews DA, Bonta J (2003) The psychology of criminal conduct. Anderson, Cincinnati
3. Bonta J (1998) Mentally disordered offenders. In: Solicitor General Canada
4. Dahle K-P (2005) Psychologische Kriminalprognose. Centaurus, Herbolzheim
5. Grove W, Zald D, Lebow B, Snitz B, Nelson C (2000) Clinical versus mechanical prediction: a meta-analysis. Psychol Assess 12:19–30
6. Hanson RK, Thornton D (2000) Improving risk assessments for sex offenders: A comparison of three actuarial scales. Law Hum Behav 24:119–136
7. Hilton NZ, Simmons JL (2001) The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. Law Hum Behav 25:393–408
8. Krippendorff K (2003) Content analysis: an introduction to its methodology. Sage Publications, Thousand Oaks, CA
9. Mossman D (1994) Assessing predictions of violence: being accurate about accuracy. J Consult Clin Psychol 62:783–792
10. Quinsey VL, Harris G, Rice M, Cormier CA (2003) Violent offenders: appraising and managing risk. APA, Washington DC
11. Repo-Tiihonen E, Halonen P, Tiihonen J, Virkkunen M (2002) Total serum cholesterol level, violent criminal offences, suicidal behavior, mortality and the appearance of conduct disorder in Finnish male criminal offenders with antisocial personality disorder. Euro Arch Psychiat Clin Neurosci 252:8–11
12. Rice ME, Harris GT (1997) Cross-validation and extension of the violence risk appraisal guide for child molesters and rapists. Law Hum Behav 21:231–241
13. Soyka M, Morhart-Klute V, Schoech H (2004) Delinquency and criminal offenses in former schizophrenic inpatients 7–12 years following discharge. Euro Arch Psychiat Clin Neurosci 254:289–294
14. Stadtland C, Kleindienst N, Kroner C, Eidt M, Nedopil N (2005) Psychopathic Traits and Risk of Criminal Recidivism in Offenders with and without Mental Disorders. Int J Forensic Mental Health 4:89–97
15. Swets J, Dawas R, Monahan J (2000) Psychological science can improve diagnostic decisions. Psychol Sci Publ Interest 1:1–26
16. Urbaniok F (2004) Validität von Risikokalkulationen bei Straftätern - Kritik an einer methodischen Grundannahme und zukünftige Perspektiven. Fortschritte für Neurologie und Psychiatrie 72:260–269
17. Urbaniok F, Endrass J, Rossegger A, Noll T (eingereicht) Violent and sexual offences: A validation of the predicitve quality of the PCL:SV in Switzerland
18. Urbaniok F, Noll T, Grunewald S, Endrass J (angenommen) Prediction of sexual and violent offences: A replication study of the VRAG in Switzerland. J Forensic Psychiat Psychol
19. Urbaniok F, Noll T, Grunewald S, Steinbach J, Endrass J (2006) Prediction of sexual and violent offences: A replication study of the VRAG in Switzerland. J Forensic Psychiat Psychol 17:23–31
20. Urbaniok F, Noll T, Rossegger A, Endrass J (2005) Die prädiktive Qualität der Psychopathy Checklist-Revised (PCL-R) bei Gewalt- und sexualstraftätern in der Schweiz. Eine Validierungsstudie. Zur Publikation eingereicht