SAPROF 2nd Edition manual updated Research chapter

May 2012

M. de Vries Robbé & V. de Vogel

Research with the SAPROF

Retrospective file studies

Research with the SAPROF is being conducted in various settings around the world. Findings from validation studies in The Netherlands concern different samples of forensic psychiatric patients. Two retrospective file studies were carried out with discharged male violent (N=105) and sexual (N=83) offenders, who had completed mandatory treatment in a forensic psychiatric hospital of an average length of 5.5 years (De Vries Robbé & De Vogel, in press; De Vries Robbé, De Vogel, & De Spa, 2011; De Vries Robbé, De Vogel, Koster, & Bogaerts, in preparation). Patients in both samples were primarily diagnosed with personality disorders and to a much lesser extent with major mental disorders. In the sexual offender sample all patients had been convicted for hands-on sexual offenses, about a fifth of these included child victims. Based on file information, the SAPROF and the HCR-20 were coded at the end of treatment. For 70 cases the files were coded by two raters. Results showed excellent interrater reliability for the SAPROF total score as well as the Final Protection Judgment for both the violent (ICC = .88; ICC = .85, $p < .01$, single measure)[1] and sexual offender (ICC = .85; ICC = .73, $p < .01$, single measure) samples.

Outcome data for all offenders were retrieved from official criminal records with a follow-up period in the community of at least 3 years after discharge. Predictive validity was examined by using Receiver Operating Characteristics analyses (see Area Under the Curve, p. 31). Table 3 shows the predictive validity for violent recidivism of SAPROF post-treatment ratings for both groups of offenders as well as the combined sample. As Table 3 indicates, the post-treatment SAPROF total score ratings for both the violent and the sexual offender samples showed good predictive validity for violent reconvictions at short-term (1 year) as well as medium-term (3 year) and long-term (average 11 year) follow-up. Although these results were equally good for violent and sexual offenders, analyses revealed different SAPROF factors that were most predictive of no future (sexual) violence for the two offender groups: *Self-control, Work* and *Finances* for the violent offenders; and *Coping, Self-control, Motivation* and *Attitudes* for the sexual offenders.

---

[1] Interrater reliability was studied using the Intraclass Correlation Coefficient (ICC; McGraw & Wong, 1996). The critical values used for ICCs (single measure) are: ICC ≥ .75 = excellent; .60 ≤ ICC < .75 = good; .40 ≤ ICC < .60 = moderate; ICC < .40 = poor (Fleiss, 1986).

By subtracting the SAPROF total scores from the HCR-20 total scores a new measure was created in which violence risk is counterbalanced by the available protection. Table 3 shows that overall the predictive accuracy of this combined measure was higher than that of the HCR-20 alone, indicating that the addition of the SAPROF adds to the predictive power of risk-only tools. At long-term follow-up, the difference in predictive accuracy between the SAPROF and the HCR-20 and between the combined measure and the HCR-20 was significant (X^2 (1, 188) = 9.1, p < .01; X^2 (1, 188) = 13.4, p < .01). The Final Protection Judgment and the Integrative Final Risk Judgment predicted violent recidivism almost equally well. In general, final judgments made on a five-point rating scale predicted violence more accurately than those made on a three-point rating scale (see 3. *Coding the Final Protection Judgment*, p. 27). Nevertheless, AUC-values for the total scores on both tools were slightly better than those for the final judgments.

Area Under the Curve

A widely used statistical method to assess predictive validity of tools is receiver operating characteristics (ROC) analysis (Douglas et al., 2007; Rice & Harris, 1995). The major advantage of this statistical method is its insensitivity to base rates. The ROC analyses result in a plot of the true positive rate (sensitivity) against the false positive rate (1 minus specificity) for every possible cut-off score of the tool. The area under the curve (AUC) can be interpreted as the probability that a randomly selected recidivist would score higher on the risk tool or lower on the protection tool than a randomly selected non-recidivist. An AUC of .50 represents chance prediction, and an AUC of 1.0 perfect prediction. In general, AUC values of .70 and above are considered moderate to large, and above .75 large (Douglas et al., 2007). Since protective factors are meant to predict non-recidivism, the AUC values for the SAPROF represent the accuracy of predictions for non-recidivism.

Table 3. Predictive validity (AUC-values) for reconvictions for a violent offense of the SAPROF post-treatment ratings for discharged offenders (N=188 ♂)

Follow-up	1 year			3 year			8/15 year (M)		
	Violent N=105	Sexual N=83	Total N=188	Violent N=105	Sexual N=83	Total N=188	Violent N=105	Sexual N=83	Total N=188
Total score SAPROF	.85**	.83**	**.85**	.74**	.77**	**.75**	.71**	.74**	**.73**
Total score HCR-20	.81**	.91**	**.84**	.68*	.80**	**.73**	.67**	.67**	**.64**
Total HCR-20 – Total SAPROF	.85**	.89**	**.87**	.72**	.80**	**.76**	.71**	.72**	**.70**
Final Protection Judgment 3-pt	.82**	.79*	**.80**	.71**	.73**	**.72**	.67**	.66*	**.66**
Final Protection Judgment 5-pt	.83**	.81*	**.83**	.72**	.71*	**.71**	.69**	.65*	**.67**
Integrative Final Risk Judgment 3-pt	.80**	.79*	**.79**	.65*	.70*	**.67**	.66*	.67**	**.66**
Integrative Final Risk Judgment 5-pt	.85**	.83**	**.84**	.71**	.73**	**.72**	.68**	.68**	**.68**

Note. $* = p < .05$, $** = p < .01$ (two-tailed).

Measuring changes during treatment with the SAPROF

For 120 of the 188 cases, the SAPROF was also coded at the beginning of treatment, based on the case history and file information from the first treatment phase. Figure 2 shows the changes in average SAPROF scores between the ratings at the beginning of treatment and the ratings at the end of treatment. Items are divided into Static (item 1-2), Dynamic improving (items 3-14) and Dynamic decreasing (items 15-17), according to their expected direction of change during treatment (see also Using the SAPROF in practice, p. 28). Comparison of the pre- and post-treatment scores showed a significant increase in the total scores on the Dynamic improving SAPROF items during treatment. As expected, the external Dynamic decreasing items showed a significant decrease towards the end of treatment. This clear shift in protection from External to Motivational and Internal is what is aimed for in treatment. In fact, the changes in SAPROF Dynamic improving item scores during treatment proved predictive of no new convictions for violent offenses long after treatment had ended (AUC = .75 at 10-year follow-up, $p < .01$). Thus, patients who showed more treatment progress as measured by the SAPROF, recidivated less after treatment had ended.

Figure 2. Pre- and post treatment ratings SAPROF (N=120)
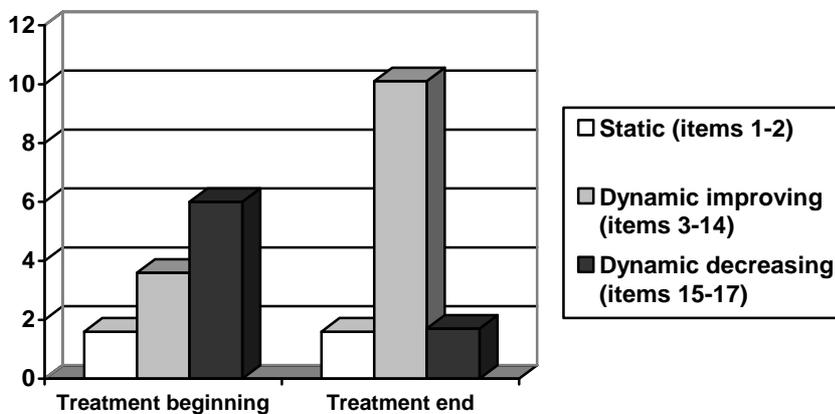
Note. Scores on items 1-2: 1.6 for both ratings; scores on items 3-14 changed from 3.6 to 10.1, $p < .001$; scores on items 15-17 changed from 6.0 to 1.7, $p < .001$.

Prospective clinical studies

Preliminary results from a prospective study including the assessments of 315 (male and female) inpatient forensic psychiatric patients was carried out to examine the clinical interrater reliability and predictive validity of the SAPROF for violent incidents during treatment (De Vries Robbé & De Vogel, 2011; De Vries Robbé, De Vogel, Wever, & Douglas, in preparation). Assessments were carried out by three independent raters from different disciplines (nursing staff member, treatment supervisor and researcher) and final scores on each item were agreed upon in a consensus meeting. Good interrater reliability was found for the SAPROF total score between the three raters of 250 multidisciplinary assessments (ICC = .70, $p < .01$, single measure). Consensus scores were used to calculate the predictive validity. Table 4 shows the predictive validity of SAPROF and HCR-20 scores for violent incidents towards others during the year following the assessment. Results from this clinical prospective study for the male population were very similar to those of the retrospective file studies. Again, the SAPROF predicted as well for male violent offenders as for male sexual offenders and the total scores predicted slightly better than the final judgments. Predictive values for female offenders were fairly good, although not significant in this small sample and lower than for the male patients. Again different factors were best predictors for the different groups: *Self-control, Attitudes, Work, Motivation* and *Medication* for the male violent offenders; *Coping, Leisure activities, Attitudes* and *Network* for the male sexual offenders; and *Intelligence, Coping, Work* and *Finances* for the female offenders. Overall, SAPROF scores had good predictive validity and combined HCR-SAPROF scores outperformed predictions by either instrument alone.

Table 4. Predictive validity for violent incidents during treatment for the SAPROF clinical ratings (N=315, 1 year follow-up)

Follow-up	Violent N=148 ♂	Sexual N=97 ♂	Total ♂ N=245 ♂	Total ♀ N=70 ♀	Total N=315
Total score SAPROF	.77**	.81**	.78**	.70	**.77****
Total score HCR-20	.74**	.85**	.79**	.78*	**.79****
Total HCR-20 – SAPROF	.81**	.84**	.82**	.76*	**.81****
Final Protection Judgment	.69*	.73**	.70**	.69	**.70****
Integrative Final Risk Judgment	.75**	.81**	.77**	.72*	**.76****

Note. * = $p < .05$, ** = $p < .01$ (two-tailed). Final judgments are made on a five-point scale.

As in the retrospective file studies, the SAPROF proved dynamic and patients' protective factors showed to be increasing as treatment progressed (see Figure 3). At the same time the ratings on the dynamic Clinical and Risk management factors of the HCR-20 showed to decrease over time. Together the improved protective factors and the diminished risk factors resulted in an overall reduction in violence risk. It was found that there were far fewer incidents of violence at the later stages in treatment and that predictions of no violent incidents by the SAPROF factors were particularly good at these later stages in treatment, when patients had had more chance to build up their protective factors during treatment (De Vries Robbé & De Vogel, 2011; De Vries Robbé et al., in preparation). These results show the applicability of the SAPROF to (forensic) clinical practice and the usefulness for measuring treatment progress. Further prospective studies will have to corroborate these findings.

Figure 3. Risk- and protective factor total scores at different stages of treatment (N=315)
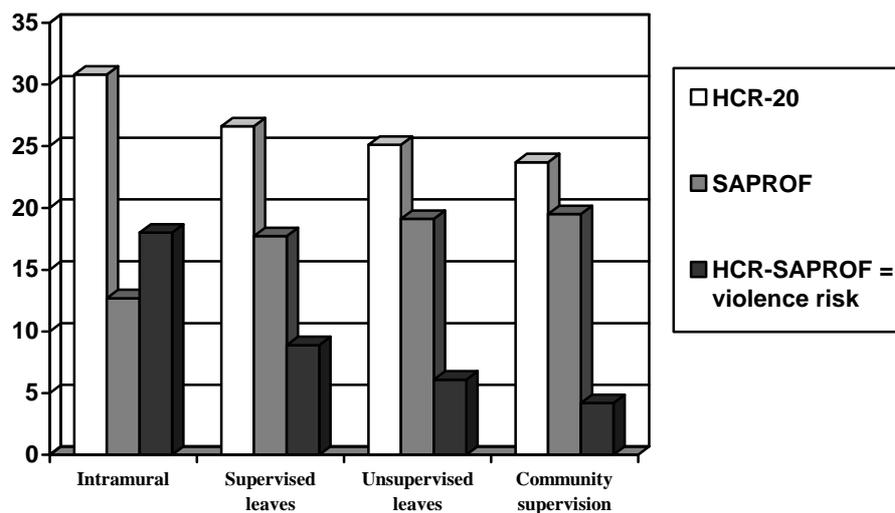
Note. Violent incident rates at the different stages of treatment: Intramural 29%; Supervised leaves 15%; Unsupervised leaves 7%; Community supervision 3%.

Summarizing, the SAPROF can be reliably coded and has demonstrated to be a strong predictor of future violence both during and after clinical treatment. These results are consistent for different types of forensic psychiatric offenders at the Van der Hoeven Kliniek. Further research in different countries and settings will have to consolidate these findings and in addition will have to focus on the usability of the SAPROF for different groups of offenders such as general offenders and offenders in outpatient treatment settings.