The Structured Assessment of Protective Factors for Violence Risk (SAPROF): A Meta-Analysis of Its Predictive and Incremental Validity

Psychological Assessment

The Structured Assessment of Protective Factors for Violence Risk (SAPROF): A Meta-Analysis of Its Predictive and Incremental Validity

Matthias Burghart, Corine de Ruiter, Sophia E. Hynes, Nishant Krishnan, Yara Levtova, and Abdo Uyar

Online First Publication, October 13, 2022. http://dx.doi.org/10.1037/pas0001184

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

https://doi.org/10.1037/pas0001184

The Structured Assessment of Protective Factors for Violence Risk (SAPROF): A Meta-Analysis of Its Predictive and Incremental Validity

Matthias Burghart[1], Corine de Ruiter[2], Sophia E. Hynes[2], Nishant Krishnan[2], Yara Levtova[2], and Abdo Uyar[2]

[1] Department of Psychology, University of Konstanz

[2] Department of Forensic Psychology, Faculty of Psychology and Neuroscience, Maastricht University

Although the inclusion of protective factors in risk assessment is believed to improve prediction, most risk assessment tools emphasize risk factors. In response, the Structured Assessment of Protective Factors for violence risk (SAPROF) has been developed, which focuses exclusively on protective factors and is used in conjunction with a structured risk assessment tool. It has received increasing attention from both researchers and forensic mental health practitioners, and additional versions have been developed for use with adolescents (SAPROF-YV) and sex offenders (SAPROF-SO). To assess their psychometric performance, we conducted a meta-analysis of validation studies. Our final sample included 39 articles with 5,434 subjects from 16 countries. Overall, the SAPROF(-YV/-SO) showed good interrater reliability and moderate-to-good predictive performance for the absence of recidivism and institutional misconduct. All three instruments exhibited incremental validity when used in conjunction with a risk-focused assessment tool. Our meta-analysis additionally showed that changes on the SAPROF are associated with decreased violent and general recidivism after controlling for baseline risk. We also uncovered several shortcomings in current research with the SAPROF(-YV/-SO). Studies did not report calibration indices and most studies were retrospective and limited to male offenders. The present findings provide support for the relevance of protective factors in risk assessment, but future research should focus on their hypothesized role in treatment and risk management.

Public Significance Statement

This meta-analysis shows that the Structured Assessment of Protective Factors for violence risk (SAPROF), a risk assessment tool that focuses exclusively on protective factors, can significantly predict the absence of violent behavior in an institution and after discharge into the community. It highlights the importance of adding dynamic protective factors to structured risk assessment to lower the risk of future violence.

Keywords: meta-analysis, predictive validity, protective factors, risk assessment, risk management

Supplemental materials: https://doi.org/10.1037/pas0001184.supp

Since the 1990s, over 150 specialized risk assessment tools have been developed for the prediction, prevention, and management of violent behavior of individuals within institutional and community settings (Heilbrun et al., 2010; Singh et al., 2011; Yang et al., 2010). Despite the proliferation of violence risk assessment instruments, the prediction and prevention of future violence remain a complex task (Abbiati et al., 2020). Given that risk estimates derived from risk assessment tools may result in decisions on the deprivation of

Matthias Burghart https://orcid.org/0000-0001-7300-1846

The authors thank Abbiati, de Vogel, Haines, Kleeven, Oziel, Tozdan, Turner, Willis, and Yoon for sharing additional data from their studies with us. The authors would also like to express our deepest gratitude to the late Jodi Viljoen who also shared additional data with us, and whose work on strength-based factors in risk assessment and risk management has been an inspiration in our research.

Corine de Ruiter is one of the coauthors of the Structured Assessment of Protective Factors for violence risk. She receives royalties from the publisher of the instrument. The remaining authors of this article do not declare any conflicts of interest.

Matthias Burghart played lead role in data curation, formal analysis and methodology and equal role in investigation, validation, writing of original draft and writing of review and editing. Corine de Ruiter played lead role in conceptualization, project administration and supervision and equal role in investigation, writing of original draft and writing of review and editing. Sophia E. Hynes played equal role in investigation, validation, writing of original draft and writing of review and editing. Nishant Krishnan played equal role in investigation, validation, writing of original draft and writing of review and editing. Yara Levtova played equal role in investigation, validation, writing of original draft and writing of review and editing. Abdo Uyar played equal role in investigation, validation, writing of original draft and writing of review and editing.

Our raw data are publicly available on open science framework (OSF; https://osf.io/jpqys/). In addition, the study protocol was preregistered on PROSPERO (ID = CRD42020172571) on April 28, 2020.

Correspondence concerning this article should be addressed to Matthias Burghart, Department of Psychology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany. Email: Matthias.2.Burghart@uni-konstanz.de

individual liberty, or permission for leave or release into the community (Monahan & Skeem, 2016; Singh et al., 2011), importance must be placed on the use of evidence-based, structured, and transparent risk assessment tools (Fazel et al., 2022).

Until the early years of the 21st century, risk assessment tools focused almost exclusively on risk factors. Risk-only evaluations, however, have been criticized for their potential to have significant negative consequences for forensic populations by leading to implicitly biased and inherently inaccurate assessments and contributing to professional negativism (Rogers, 2000). In direct response to this criticism and in an effort to create a more balanced risk assessment practice, new tools evolved that incorporate protective factors (de Ruiter & Nicholls, 2011). Protective factors refer to strengths that counterbalance or mitigate risk factors and thereby assist in preventing individuals from recommitting violence (de Vries Robbé et al., 2013). The addition of a protective, strengths-based approach to risk assessment could result in fairer risk evaluations, the provision of more motivating perspectives for both offenders and treatment providers, and more personalized treatment and risk management plans (de Vries Robbé & Stam, 2012; Jones et al., 2015; Klepfisz et al., 2017).

Developed as a companion instrument for use alongside risk-focused assessment tools, such as the Historical Clinical Risk Management–20 (HCR-20; Douglas et al., 2013), the Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel et al., 2012) is the first assessment instrument that focuses exclusively on identifying protective factors that guard against future violence in adult populations (de Vogel et al., 2012). The SAPROF comprises 17 protective factors, organized into three subscales: internal factors, motivational factors, and external factors. The internal factors are historical and static, while the other 15 are dynamic and, in principle, changeable through intervention. The dynamic items are rated for the upcoming 6–12 months, which implies rating of these factors depends on the context (e.g., while residing in an institution or in the community). SAPROF factors are rated on a 3-point scale (0, 1, 2), with higher scores indicating more protection. The SAPROF assessment concludes with a "Final Protection Judgment" (FPJ; low–moderate–high). Subsequently, the protective factors assessment is taken together with the risk factors assessment using the preferred risk assessment instrument, creating a risk formulation, risk scenarios, a risk management and treatment plan, and an integrated final risk judgment (de Vries Robbé et al., 2020).

Since its development, studies have shown that the SAPROF provides a more comprehensive approach to predicting and managing violence than risk-only evaluations, and has subsequently become a popular tool for evaluating protective factors for (sexual) violence risk (O'Shea & Dickens, 2016). Strengths-based assessment tools, such as the SAPROF, are representative of

> the latest evolution in risk instruments (i.e., fourth generation), […] specifically designed to be integrated into (a) the process of risk management, (b) the selection of intervention modes and targets for treatment, and (c) the assessment of rehabilitation progress (Andrews & Bonta, 2006; Andrews et al., 2006). (Campbell et al., 2009, p. 569)

Thus, an important test of the utility of these tools is the predictive validity of protective factor change scores beyond the posttreatment risk level. Several recent studies provide supportive evidence that treatment-related improvements on the SAPROF indeed result in decreased violent recidivism after controlling for baseline risk scores (Coupland & Olver, 2020; Olver & Riemer, 2021).

Newer SAPROF Versions: Sexual Offending and Juvenile Offending

In recent years, two new adaptations of the SAPROF have been developed. First, the Structured Assessment of Protective Factors for violence risk–Sexual Offence version (SAPROF-SO; Willis et al., 2017) is an assessment tool for protective factors against sexual reoffending. On the basis of reviews of the empirical literature, 24 SAPROF-SO items were developed. Item descriptions were informed by research specific to sexual offending as well as general recidivism research, including the strengths-based good lives model (GLM; Ward & Fortune, 2013) of offender rehabilitation. Many original SAPROF items were revised including attempts to incorporate an understanding of protective mechanisms into scoring instructions, and some were renamed to fit the new item descriptions (Willis et al., 2020). New items in the SAPROF-SO include protective factors hypothesized to specifically prevent sexual recidivism: sexual self-regulation, prosocial sexual interests, and prosocial sexual identity.

Second, the Structured Assessment of Protective Factors for violence risk–Youth Version (SAPROF-YV; de Vries Robbé, Geers, et al., 2015) is the adolescent counterpart of the adult version. The SAPROF-YV was designed for the assessment of protective factors for violence in juveniles. It consists of 16 dynamic protective factors: four resilience, six motivational, three relational, and three external factors. Preliminary findings support the association between the SAPROF-YV protective factors and absence from reoffending in a retrospective file study in a mixed sample of juvenile and young adult offenders (Kleeven et al., 2022).

The Current Meta-Analysis

A review of the existing literature shows that no published meta-analysis has focused on the effectiveness of the SAPROF, and its more recently developed sister instruments, as a predictor of the absence of future reoffending and institutional misconduct. Given (a) the recent development of a protective, strengths-based approach to risk assessment, (b) the SAPROF's standing as one of the important tools for the assessment of protective factors, and (c) the recent proliferation of studies on the instrument, we considered it timely to undertake a meta-analysis of the current research on the predictive validity of the SAPROF(-YV/-SO) for the absence of recidivism and institutional misconduct.

Method

Protocol, Open Data, and Conflict of Interest

The recommended Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed (Moher et al., 2009). In addition, the study protocol was preregistered on PROSPERO on April 28, 2020. The raw data are available on the Open Science Framework under the following link: https://osf.io/jpqys/. Since openly available data were used in this article, ethical approval was not required.

For full transparency, it should also be noted that the second author of the present article is one of the developers of the SAPROF

and receives royalties from its publication. The other authors do not have any conflicts of interest to declare.

Search Strategy

Three electronic databases were used to search for eligible articles: PsycINFO, PubMed, and Web of Science. To keep the search as general as possible, the search terms were limited to *SAPROF* OR *SAPROF-YV* OR *SAPROF-SO* OR *structured assessment of protective factors for violence risk*. Articles were additionally acquired using ResearchGate and by manually searching Google Scholar and relevant reference lists. Last, the manager of the official SAPROF website as well as other researchers in the field were contacted for unreported or ongoing studies.

Study Eligibility

An article was deemed relevant for the current meta-analysis if it (a) assessed the interrater reliability and/or the predictive validity of the SAPROF(-YV/-SO); (b) originated from peer-reviewed publications, gray literature, book chapters/books, or unpublished dissertations; (c) used community, clinical, and/or offender samples; (d) was written in either English or German. Single-case studies, as well as articles that did not contain primary study data (e.g., conference abstracts, book reviews), were excluded from the analysis.

Study Selection and Data Extraction

Titles and abstracts of all retrieved articles were examined for eligibility by two independent raters. Upon agreement, the articles were considered for full-text screening. Authors agreed on 81% of studies in the title/abstract screening phase and 96% of studies in the full-text screening phase. Disagreements between raters were resolved by the second author. Subsequently, two independent raters extracted relevant data from the articles based on a predetermined list of study characteristics and possible moderators. The authors agreed on 87% of the extracted data. Disagreements were resolved by the first and second authors. If information was lacking or needed clarification, the corresponding authors were contacted.

Data were extracted for two types of outcomes: recidivism in the community and institutional misconduct. The former deals with reoffending after an individual is released from a secure institution, whereas the latter describes incidents that occur while the person is still institutionalized. For each outcome, three types of incidents were considered as follows: (a) violent incidents, (b) sexually violent incidents, and (c) any type of incident.

Performance Measures

The interrater reliability of the SAPROF(-YV/-SO) was assessed using intraclass correlation coefficients (ICCs) between two or more raters. Other measures of interrater reliability were also extracted, but are reported separately due to their small number.

Predictive performance was assessed by measures of discrimination. The most widely used discrimination index is the area under the receiver operating characteristic curve (AUC; Singh et al., 2013). It quantifies the probability that the score of a random recidivist will be higher (or in the case of the SAPROF: lower) than the score of a random nonrecidivist. AUCs can range from 0 to 1, with .5 indicating prediction at the chance level. In case articles did not report corresponding 95% confidence intervals ($n = 5$), these were estimated from the number of recidivists and nonrecidivists in the sample (Hajian-Tilaki & Hanley, 2002).

It is relevant to note that the predictive validity of a risk assessment tool is not only determined by its ability to discriminate between recidivists and nonrecidivists, but also requires information about how well a tool's prediction matches the true risk that an individual will recidivate (Cook, 2007). This is referred to as calibration and is a prerequisite to capture the full picture of the predictive utility of a risk assessment tool (Singh, 2013). There are many different calibration indices (e.g., Brier score, E/O index, calibration slopes), all of which have their own strengths and limitations (for an overview, see Huang et al., 2020). For the purposes of this meta-analysis, the intercept (i.e., b0) and coefficient (i.e., b1) of logistic regression analyses with the SAPROF(-YV/-SO) were included given that other calibration indices are underreported (Fazel et al., 2022). These indices were then used to create calibration plots (Lee et al., 2020).

Finally, the exponent of the coefficient (i.e., eB) from the final step of a hierarchical regression analysis (Cox or logistic regression) served as an effect size measure of the incremental validity of the SAPROF(-YV/-SO). It quantifies the extent to which the SAPROF(-YV/-SO) uniquely predicts the outcome of interest when included in the model simultaneously with another risk assessment tool (Walters, 2012). The same approach was used to examine whether a change in SAPROF(-YV/-SO) scores over the course of treatment can incrementally predict recidivism/institutional misconduct beyond baseline risk scores alone.

Synthesis of Results

ICCs were summarized descriptively using the arithmetic mean, median, and interquartile range (IQR). This approach was preferred over a meta-analytical model because most studies did not report corresponding measures of uncertainty (e.g., 95% CI).

AUCs were first converted to Cohen's *d* effect sizes, following the recommendations of Rice and Harris (2005), and subsequently pooled for each type of outcome. Depending on the independence of the effect sizes, three different meta-analytical models were applied to the data. That is, when effect sizes were not independent from each other (e.g., due to studies reporting results for multiple follow-up periods or using overlapping samples) a three-level meta-analysis was used. This multilevel approach accounts for nonindependence by breaking down the heterogeneity of the true effects (τ^2) into the variance within samples ($\sigma_2{}^2$; effect size-level) and the variance between samples ($\sigma_1{}^2$; study level; Cheung, 2014). Articles that used the same sample were treated as coming from the same study level (see Supplemental Table 1, for more details about sample overlap). Whenever nonindependence between effect sizes was not an issue, random-effects and fixed-effect models were conducted. In addition, the length of follow-up, base rate of incidents, country of publication (grouped by continent: "Asia" vs. "Europe" vs. "North America"), and setting ("correctional" vs. "forensic psychiatric hospital/mental health facility") were examined for their moderating influence on the predictive validity of the SAPROF(-YV/-SO).

The same meta-analytical approach was applied to pool the effect size measures of calibration (i.e., b0 and b1 of logistic regression

analyses) and incremental validity (i.e., eB of the final step of hierarchical logistic/Cox regressions). However, given the small number of effect sizes compared with those of predictive validity, no moderation analyses could be performed.

Whenever possible, results were also reported for the subscales and the FPJ of the SAPROF(-YV/-SO). All analyses were performed in R (Version: 4.0.2; R Core Team, 2020) with the metafor package (Viechtbauer, 2010).

Risk of Bias

The risk of bias was assessed for all included studies with the Prediction Model Risk of Bias Assessment Tool (PROBAST; Wolff et al., 2019). The PROBAST rates the risk of bias as either low, high, or unclear in four domains: (a) participants, (b) predictors, (c) outcome, and (d) analysis. The first domain is concerned with how a sample was selected. High risk of bias is to be expected if, for instance, a certain group of participants is systematically not included in the sample (e.g., individuals from minority groups). The predictor domain addresses potential sources of bias associated with the rating of the predictor (i.e., SAPROF). Sources of bias may include items that are systematically missing from the entire sample (e.g., because no data were available to code that particular item) or raters of the instrument who are not blind to the outcome in a retrospective study design. The outcome domain covers biases resulting from an inappropriate operationalization of the outcome variable. For example, when self-reports are used instead of official records of recidivism. Last, the analysis domain refers to issues concerning the analytical approaches used to determine the predictive validity of a tool. This includes a low prevalence of the outcome variable in a sample (i.e., fewer than 100 participants recidivated). PROBAST ratings were conducted in pairs of two authors, with an average agreement of 85%. Remaining disagreements were subsequently discussed among all authors.

Results

Description of Included Studies

Our systematic literature search yielded 729 articles. After removing duplicates and noneligible articles, 39 articles remained for inclusion in the quantitative analysis, of which 28 assessed the SAPROF, nine the SAPROF-YV, and two the SAPROF-SO (Supplemental Figure 1). When taking into account the overlap between samples, our meta-analysis included 5,434 (SAPROF: n = 3,449; SAPROF-YV: n = 1,735; SAPROF-SO: n = 250) independent participants from 16 different countries. A detailed description of all included articles can be found in Supplemental Table 1.

Twenty-one articles examined recidivism after discharge, while 12 examined institutional misconduct, and three examined both.[1] Although results were often reported for more than one outcome (violence, sexual violence, or any incident), the prediction of violence predominated. Furthermore, studies reporting recidivism had longer follow-up periods (*Mdn* = 36 months; range = 6–180 months) and lower base rates (*Mdn* = 20%; range = 6–79%) than studies reporting institutional misconduct (follow-up period: *Mdn* = 9 months; range = 1–30 months; base rate: *Mdn* = 28%; range = 9–66%).

Interrater Reliability

The interrater reliability of the SAPROF(-YV/-SO) was assessed in 27 articles (SAPROF: n = 18; SAPROF-YV: n = 7; SAPROF-SO: n = 2), most of which (n = 25) reported ICCs (including both absolute agreement and consistency). Across these articles, the mean ICCs for the total score of the SAPROF, SAPROF-YV, and SAPROF-SO were .80 (IQR = .74–.86), .80 (IQR = .77–.90), and .94 (IQR = .92–.96), respectively. According to Koo and Li's (2016) conventions for interpreting ICCs, these values can be considered indicative of good to excellent interrater reliability. The remaining two articles reported Spearman rho (ρ = .83; Abidin et al., 2013) and Pearson r (r = .86; Oziel et al., 2020) correlation coefficients for the SAPROF total score, which also indicate good to excellent reliability. Excluding studies conducted by authors involved in the tool's development did not significantly alter the mean ICCs for the SAPROF (.79) and SAPROF-YV (.81). All reliability coefficients reported for the SAPROF-SO were obtained from studies conducted by the original developers of the instrument, which is why no sensitivity analysis could be performed.

Interrater reliability was also assessed for the FPJ as well as for all subscales of the SAPROF and SAPROF-YV (these data were unavailable for the SAPROF-SO). The average ICCs for these scales ranged from moderate to good and were, for the most part, slightly lower than the average ICCs for the total scores. The largest variation in reported ICCs was found for the External subscale of the SAPROF, ranging from .31 to .92 (mean ICC = .69). The ICC for the FPJ of the SAPROF-YV was comparatively low (.61), but based on only one study (Kleeven et al., 2022).

All results concerning the interrater reliability of the SAPROF(-YV/-SO) are reported in Table 1. Additional box plots showing the range of all individual coefficients can be found in the Supplemental Material (Supplemental Figure 2).

Predictive Validity

Recidivism After Discharge

Of the 24 articles that examined recidivism after discharge, 14 reported findings on the predictive validity of the SAPROF, six on the SAPROF-YV, and just one on the SAPROF-SO. The meta-analytical results for all three tools are shown in Table 2.

Analysis of the total scores yielded statistically significant pooled effect sizes for the prediction of violent recidivism, sexual recidivism, and any type of recidivism. The magnitude of these pooled effects ranged from moderate to large (Cohen, 1988), but differed across the instruments depending on the outcome variable. More specifically, the SAPROF produced the largest pooled effect size for the prediction of violent recidivism, while the SAPROF-YV predicted any type of recidivism best. The SAPROF-SO performed best in predicting sexual recidivism. Almost all meta-analytical models indicated a statistically nonsignificant degree of heterogeneity (based on the Q statistic); however, I^2 ranged from low to moderate (Higgins et al., 2003). Heterogeneity was the largest for models that

[1] The sum of these articles (n = 36) does not match the total number of included articles (n = 39) since three articles did not examine the predictive performance of the SAPROF(-YV/-SO) but were still included because they reported interrater reliability coefficients.

Table 1
Intraclass Correlation Coefficients (ICCs) for the SAPROF, SAPROF-YV, and SAPROF-SO

Scale	*k*	*M*	*Mdn*	IQR
SAPROF				
Total score	18	.80	.78	.74, .86
Internal	9	.70	.73	.61, .78
Motivational	9	.79	.79	.71, .84
External	9	.69	.74	.58, .76
FPJ	6	.72	.74	.66, .77
SAPROF-YV				
Total score	7	.80	.84	.77, .90
Motivational	2	.80	.80	.73, .86
Relational	2	.82	.82	.79, .84
Resilience	2	.72	.72	.67, .78
External	2	.75	.75	.74, .76
FPJ	1	.61	—	—
SAPROF-SO				
Total score	3	.94	.94	.92, .96

Note. *k* = number of effect sizes (i.e., ICCs); IQR = interquartile range; FPJ = Final Protection Judgement; SAPROF = Structured Assessment of Protective Factors for violence risk; SAPROF-YV = Structured Assessment of Protective Factors for violence risk–Youth Version; SAPROF-SO = Structured Assessment of Protective Factors for violence risk–Sexual Offence.

assessed the prediction of any recidivism. Furthermore, no significant moderators were found.

Fewer articles reported findings on the FPJ and the subscales of the SAPROF and SAPROF-YV, and none reported such findings for the SAPROF-SO (Table 2). While the FPJ of the SAPROF produced larger pooled effect sizes for the prediction of all three types of recidivism compared to the total score, the opposite was observed for the SAPROF-YV. As for the subscales, the External subscale of both the SAPROF and the SAPROF-YV performed poorly when predicting violent, sexual, and any recidivism. In contrast, the Motivational subscale of both tools showed good predictive performance, with the exception for sexual recidivism.

Institutional Misconduct

The predictive validity of the SAPROF (n = 12) and the SAPROF-YV (n = 3) for institutional misconduct was reported in 15 articles. The pooled effect sizes are presented in Table 3.

The SAPROF total score significantly predicted the absence of violent, sexual, and any institutional misconduct, with moderate-to-large effects (Cohen, 1988). While the SAPROF total score performed particularly well for violent misconduct, the total score of the SAPROF-YV failed to reach statistical significance for predicting institutional violence. Data on the SAPROF-YV for sexual and any misconduct were not available. The estimated heterogeneity was low to moderate, but again statistically nonsignificant. However, we did find a moderating effect of base rate and country in which the study was conducted on the predictive validity of the SAPROF total score for violent misconduct. Specifically, the base rate had a negative impact on the predictive performance ($\beta = -1.73$, 95% CI [−2.84, −.63], $p < .01$) and studies from Asia reported significantly larger effect sizes than studies from North America, but not than studies from Europe ($Q_{between} = 12.42$, $p < .01$; $d_{Asia} = 1.52$, 95% CI [0.91, 2.13], $p < .001$; $d_{Europe} = .96$, 95% CI [0.73, 1.19], $p < .001$; $d_{North\ America} = .44$, 95% CI [0.14, 0.75], $p < .01$).

Meta-analytical models for the FPJ of the SAPROF yielded smaller but still significant summary effect sizes compared to the total score for violent and any misconduct. For the prediction of sexual misconduct, the one available effect size was larger than that for the total score. The FPJ of the SAPROF-YV was the only significant predictor of violent institutional misconduct, but this finding is based on just one effect size. Again, the External subscale showed mostly poor predictive performance. None of the subscales of the SAPROF-YV significantly predicted the outcome variable (violence), and all summary effect sizes were small in magnitude.

Incremental Validity

Data on incremental validity were extracted from 19 articles (SAPROF: n = 13, SAPROF-YV: n = 5, and SAPROF-SO: n = 1), which used different risk assessment tools in conjunction with the SAPROF(-YV/-SO). These risk assessment instruments were: HCR-20, Psychopathy Checklist–Revised (PCL-R; Hare, 2003), Static-99 (Hanson & Thornton, 2000), Violence Risk Scale–Youth Version (VRS-YV; Wong et al., 2004–2011), youth level of service/case management inventory (YLS/CMI; Hoge & Andrews, 2004), and the risk scale of the Structured Assessment of violence risk in Youth (SAVRY; Borum et al., 2006). Results are summarized in Table 4.

The summary effect sizes showed that the SAPROF significantly predicted the absence of recidivism (violence and any type) and institutional misconduct (violence and any type) after controlling for the score on a risk-focused instrument. However, the SAPROF did not add incremental power to the prediction of absence of sexual recidivism. The statistically significant odds ratios (*ORs*) ranged from .80 to .96 and can be interpreted as the factor by which the SAPROF total score reduces the odds of an outcome occurring (e.g., in the case of violent recidivism, by a factor of .91), after controlling for the predictive value of the risk-focused assessment tool. The sister tools of the SAPROF performed similarly well, with significant results in the same range for the prediction of absence of violent and any recidivism (SAPROF-YV) and sexual recidivism (SAPROF-SO).

Predictive Validity of SAPROF Change Scores

Whether changes in protective factors (e.g., through interventions) add incremental predictive power beyond the level of baseline risk was examined in three articles (Coupland & Olver, 2020; de Vries Robbé, de Vogel, Douglas, et al., 2015; Olver & Riemer, 2021). The multilevel meta-analyses produced statistically significant summary effect sizes for the prediction of absence of violent recidivism (OR = .94, 95% CI [.92, .96], $p < .001$) and absence of any recidivism (OR = .97, 95% CI [.95, 1.00], p = .02). Only one effect size was available for the prediction of sexual nonrecidivism, which was statistically nonsignificant (OR = .96, 95% CI [.90, 1.02], p = 1.68; Olver & Riemer, 2021). These results suggest that, on average, with each one-point increase of the SAPROF total score, the odds of violent recidivism decrease by 6% and the odds of any type of recidivism by 3%, after controlling for baseline risk level.

As for the SAPROF's sister tools, only one article investigated the predictive performance of change on the SAPROF-YV (Koh et al., 2021). In this study, however, the SAPROF-YV change score failed to significantly predict the absence of reoffending (violent and any type).

Table 2

Results of Meta-Analytical Models (Cohen's d) Assessing the Predictive Validity of Decreased Recidivism for the SAPROF, SAPROF-YV, and SAPROF-SO

Scale	*n*	*k*	Model	*d* [95% CI]	I^2 in %	*Q*-test
Violent recidivism						
SAPROF						
Total score	13	21	ML	.63 [.44, .82]***	38.86	25.07
Internal	9	11	ML	.46 [.30, .61]***	12.25	11.10
Motivational	9	11	ML	.62 [.49, .76]***	0	6.08
External	9	11	ML	.22 [.05, .40]*	27.86	11.66
FPJ	5	9	ML	.75 [.51, .99]***	0	3.37
SAPROF-YV						
Total score	5	10	ML	.51 [.37, .65]***	0	1.32
Motivational	2	5	ML	.52 [.35, .69]***	0	.40
Relational	2	5	ML	.45 [.27, .64]***	6.98	1.74
Resilience	2	5	ML	.37 [.21, .54]***	0	.75
External	2	5	ML	.22 [.06, .39]**	0	1.10
FPJ	2	4	ML	.39 [.10, .68]**	0	.20
SAPROF-SO						
Total score	1	1	PE	.58 [.14, 1.15]**	—	—
Sexual recidivism						
SAPROF						
Total score	6	7	ML	.41 [.11, .71]**	28.17	5.99
Internal	4	4	RE/FE	.39 [.10, .69]**/.41 [.15, .67]**	14.22/12.55	3.43
Motivational	4	4	FE	.32 [.06, .58]*	0	.84
External	4	4	FE	.19 [–.07, .45]	0	1.11
FPJ	2	3	ML	.51 [.04, .99]*	0	.03
SAPROF-SO						
Total score	1	1	PE	1.24 [.82, 1.84]***	—	—
Any recidivism						
SAPROF						
Total score	11	13	ML	.59 [.39, .80]***	43.88	16.49
Internal	10	12	ML	.47 [.30, .64]***	24.15	12.37
Motivational	10	12	ML	.57 [.40, .73]***	20.27	10.41
External	10	12	ML	.25 [.01, .48]*	55.35	23.15*
FPJ	3	3	FE	.65 [.27, 1.04]***	0	.46
SAPROF-YV						
Total score	5	10	ML	.77 [.49, 1.05]***	69.11	19.69*
Motivational	2	5	ML	.64 [.49, .79]***	0	.84
Relational	2	5	ML	.51 [.36, .66]***	0	.59
Resilience	2	5	ML	.41 [.26, .56]***	0	1.09
External	2	5	ML	.30 [.15, .44]***	0	1.03
FPJ	2	4	ML	.60 [.32, .87]***	45.58	5.08
SAPROF-SO						
Total score	1	1	PE	.47 [.21, .78]**	—	—

Note. *n* = number of articles; *k* = number of effect sizes included in the model; ML = multilevel model; RE = random-effects model; FE = fixed-effect model; PE = point estimate (i.e., no meta-analysis was performed); *d* = Cohen's *d*; CI = confidence interval; I^2 = between study heterogeneity; *Q*-test = test of heterogeneity between studies; FPJ = Final Protection Judgement; SAPROF = Structured Assessment of Protective Factors for violence risk; SAPROF-YV = Structured Assessment of Protective Factors for violence risk–Youth Version; SAPROF-SO = Structured Assessment of Protective Factors for violence risk–Sexual Offence. When $I^2 > 0$ and no dependence between effect sizes, the results of both random-effects and a fixed-effect model are reported (FE = RE when $I^2 = 0$).
* $p < .05$. ** $p < .01$. *** $p < .001$.

Calibration

None of the articles included in this meta-analysis directly investigated calibration indices of the SAPROF(-YV/-SO) as recommended by Huang et al. (2020). Fortunately, after contacting the authors, we were able to obtain coefficients for the intercept and slope of a logistic regression analysis for seven articles (SAPROF: $n = 4$, SAPROF-YV: $n = 2$, SAPROF-SO: $n = 1$), which we then used to create calibration plots (Supplemental Figure 3). Overall, they revealed that individuals with higher protective scores were predicted to be less likely to recidivate or engage in institutional misconduct. In this sense, the SAPROF(-YV/-SO) works as intended. However, these results should be viewed as preliminary and interpreted with caution for two reasons. First, the calibration plots for each outcome were mostly based on a single effect size. Second, we included effect sizes from de Vogel et al. (2019), who examined the SAPROF in an all-female sample in which the SAPROF performed poorly.

Risk of Bias

PROBAST

With the exception of Olver and Riemer (2021), Lovatt et al. (2022), and Chu et al. (2020), all studies had a high risk of bias in the

Table 3

Results of Meta-Analytical Models (Cohen's d) Assessing the Predictive Validity of Decreased Institutional Misconduct for the SAPROF and SAPROF-YV

Scale	*n*	*k*	Model	*d* [95% CI]	I^2 in %	*Q*-test
			Violent misconduct			
SAPROF						
Total score	10	14	ML	.88 [.63, 1.14]***	39.50	16.81
Internal	6	7	ML	.79 [.49, 1.10]***	17.83	5.03
Motivational	6	7	ML	.84 [.57, 1.10]***	0	4.26
External	6	7	ML	.30 [.01, .60]*	14.74	5.75
FPJ	6	8	ML	.71 [.34, 1.08]***	53.56	12.64
SAPROF-YV						
Total score	3	3	ML	.42 [–.01, .85]	0	.68
Motivational	2	2	ML	.26 [–.21, .73]	0	.12
Relational	2	2	ML	.33 [–.14, .80]	0	.12
Resilience	2	2	ML	.31 [–.17, .78]	0	.11
External	2	2	ML	.30 [–.17, .78]	0	.78
FPJ	1	1	PE	.67 [.07, 1.46]*	—	—
			Sexual misconduct			
SAPROF						
Total score	1	1	PE	.75 [.11, 1.61]*	—	—
Internal	1	1	PE	.55 [–.21, 1.46]	—	—
Motivational	1	1	PE	.44 [–.11, 1.09]	—	—
External	1	1	PE	.51 [–.28, 1.53]	—	—
FPJ	1	1	PE	.86 [.26, 1.74]*	—	—
			Any misconduct			
SAPROF						
Total score	3	3	RE/FE	.69 [.02, 1.37]*/.49 [.13, .85]**	57.37/56.76	4.63
Internal	3	3	RE/FE	.67 [–.05, 1.40]/.46 [.10, .82]*	63.32/60.02	5.00
Motivational	3	3	RE/FE	.71 [.05, 1.37]*/.50 [.14, .86]**	55.74/56.34	4.58
External	3	3	FE	.17 [–.18, .52]	0/0	.03
FPJ	3	3	RE/FE	.53 [.02, 1.03]*/.44 [.08, .80]*	33.02/39.47	3.30

Note. *n* = number of articles; *k* = number of effect sizes included in the model; ML = multilevel model; RE = random-effects model; FE = fixed-effect model; PE = point estimate (i.e., no meta-analysis was performed); *d* = Cohen's *d*; CI = confidence interval; I^2 = between study heterogeneity; *Q*-test = test of heterogeneity between studies; FPJ = Final Protection Judgement; SAPROF = Structured Assessment of Protective Factors for violence risk; SAPROF-YV = Structured Assessment of Protective Factors for violence risk–Youth Version. When $I^2 > 0$ and no dependence between effect sizes, the results of both random-effects and a fixed-effect model are reported (FE = RE when $I^2 = 0$).
* $p < .05$. ** $p < .01$. *** $p < .001$.

analysis domain (Supplemental Figure 4). In most cases this was due to too few outcome events, the benchmark being set at 100 individuals showing the outcome (Wolff et al., 2019). Problems within the other three domains (i.e., Participants, Predictors, and Outcome) were less common and, if present, mostly related to selection bias (e.g., exclusion of a particular group of participants) or inappropriate use of the SAPROF(-YV/-SO; e.g., failure to rate all items or using a modified version).

Authorship Bias

In 10 of the 39 included articles, one or more of the developers of the SAPROF(-YV/-SO) were involved as the author(s). The risk of authorship bias was particularly pronounced for the SAPROF-SO as this tool has not yet been validated by an independent research group. In addition, nearly half of the effect sizes quantifying the predictive performance of the SAPROF for violent recidivism after discharge are from studies by de Vries Robbé and colleagues. Still, including the risk of authorship bias (i.e., "yes" vs. "no") as a moderator to the meta-analytical models, did not yield significant differences in any of the summary effect sizes.

Discussion

The current meta-analysis is the first to examine the existing research on the interrater reliability, predictive and incremental validity of the SAPROF and its sister instruments, the SAPROF-YV, and the SAPROF-SO. Thirty-nine articles were included in this review, comprising 5,434 independent participants from 16 countries. Overall, all three tools demonstrated good to excellent interrater reliability as well as statistically significant predictive validity for nonrecidivism after discharge. The SAPROF further showed good predictive power for the absence of institutional misconduct.

While the SAPROF generally provided better predictions for the absence of violent recidivism than for the absence of sexual and any type of recidivism, the SAPROF-SO performed best in predicting the absence of sexual recidivism. This was expected as the SAPROF was specifically developed to assess protective factors against violence risk (de Vogel et al., 2012), while the SAPROF-SO was designed to assess protective factors against sexual violence risk (Willis et al., 2017). Pooling the effect sizes of the three articles that used the original SAPROF in a juvenile sample (Awrey, 2021; Klein et al., 2015; Zeng et al., 2015) and comparing them to the summary effect sizes for the SAPROF-YV, makes it clear that the youth

Table 4

Results of Meta-Analytical Models (OR) Assessing the Incremental Validity of the SAPROF, SAPROF-YV, and SAPROF-SO for Different Outcomes

Scale	*n*	*k*	Model	*OR*	95% CI
			SAPROF		
Recidivism					
Violence	5	11	ML	.91**	[.87, .96]
Sexual	4	5	ML	.97	[.91, 1.04]
Any	6	7	ML	.96*	[.92, .99]
Misconduct					
Violence	3	5	ML	.92**	[.87, .97]
Any	2	2	FE	.80*	[.68, .96]
			SAPROF-YV		
Recidivism					
Violence	3	6	ML	.96**	[.94, .99]
Any	5	7	ML	.94*	[.88, 1.00]
			SAPROF-SO		
Recidivism					
Sexual	1	1	PE	.89***	[.83, .94]

Note. n = number of articles; k = number of effect sizes included in the model; ML = multilevel model; FE = fixed-effect model; PE = point estimate (i.e., no meta-analysis was performed); *OR* = odds ratio; CI = confidence interval; SAPROF = Structured Assessment of Protective Factors for violence risk; SAPROF-YV = Structured Assessment of Protective Factors for violence risk–Youth Version; SAPROF-SO = Structured Assessment of Protective Factors for violence risk–Sexual Offence. An $OR < 1$ indicates a risk mitigating effect of the SAPROF(-YV/-SO). * $p < .05$. ** $p < .01$. *** $p < .001$.

version is better suited for the assessment of young offenders than the adult SAPROF (violent recidivism: $d_{\text{SAPROF}} = .40$, 95% CI [.06, .74], $p = .02$ vs. $d_{\text{SAPROF-YV}} = .51$; any recidivism: $d_{\text{SAPROF}} = .43$, 95% CI [.13, .73], $p < .05$ vs. $d_{\text{SAPROF-YV}} = .77$).

An unexpected finding was the SAPROF's lower predictive accuracy for the absence of violence after discharge into the community ($d_{\text{SAPROF}} = .63$) compared to the absence of violent institutional misconduct ($d_{\text{SAPROF}} = .88$). One possible explanation for this finding could be that measures of institutional misconduct are more accurate than measures of recidivism in the community. Official criminal justice records tend to underestimate true recidivism rates (Hardin & Scurich, 2022) and therefore may reduce predictive validity. An alternative explanation may be that coding the SAPROF while the individual is in the controlled environment of an institution is less prone to error because there are relatively fewer "unknowns." On the contrary, coding of the SAPROF at the time of discharge from the institution may not accurately represent the "true protection score" for the individual while they are residing in the community at a later date. Thus, future research should attempt to conduct repeated SAPROF assessments during the process of reintegration into society. This way we can examine the tool's predictive potential across different reintegration phases. In addition, future research should also focus on the differential performance of the SAPROF subscales and possibly specific items thereof, as several studies have shown that the Motivational and Internal subscales in particular predict decreased violent recidivism (e.g., Olver & Riemer, 2021).

As a structured professional judgement (SPJ) tool, the results of a SAPROF(-YV/-SO) assessment can be formulated not only in terms of a total score but also in the form of a categorical FPJ. Although fewer studies have examined the predictive performance of the FPJ, our meta-analysis provides some evidence for its predictive validity. In the context of recidivism after discharge, the SAPROF's FPJ yielded larger summary effect sizes than the total score, for all three outcome variables. However, this was not the case for the SAPROF-YV, and due to a lack of studies, could not be examined for the SAPROF-SO. In predicting institutional misconduct, the FPJ generally performed worse than the total score, with the exception for violent misconduct (SAPROF-YV) and sexual misconduct (SAPROF). It is important to note, however, that the latter estimates are each based on a single effect size. The practical utility of categorical risk and protection estimates is currently still debated in the literature (Scurich, 2018). Opponents argue that risk categories (e.g., low, moderate, high) are often not consistently interpreted in terms of their level of risk and therefore should not be used to communicate results. We do nevertheless think a categorical risk/protective estimate, such as the SAPROF's FPJ, may also have advantages over the total score in that it facilitates triaging offenders into groups that require high- versus low-intensity treatment/risk management. However, for the development of a tailored risk formulation as well as a risk management and treatment plan, the focus should be on the individual SAPROF items since "for one person, the 'key' factors may be self-control, work, and social network, while for another they may be medication, treatment motivation, and living circumstances" (de Vries Robbé et al., 2020, p. 411).

Particularly, striking was the comparatively poor performance of the External subscale of the SAPROF and the SAPROF-YV for predicting the absence of recidivism and institutional misconduct (data were not available for the subscales of the SAPROF-SO). The External subscale of the SAPROF comprises five items, including social network, intimate relationship, professional care, living circumstances, and external control, whereas the External subscale of the SAPROF-YV comprises three items, namely pedagogical climate, professional care, and court order. Considering that most of these items are expected to decrease (rather than increase) when reintegration into the community is successful, this may explain its poor performance in predicting decreased recidivism. That is, because well-functioning individuals in the community do not require professional care, supervised living, or court-mandated supervision, an increase in protection is not reflected in a higher score on the External subscale, but (ideally) in a "transfer of protection" away from external factors toward the internal and motivational factors (de Vries Robbé et al., 2020, p. 425). In addition, institutionalized patients are expected to have similar levels of external control and living circumstances, which consequently reduces variance in scores on the External subscale in institutionalized patients. Due to this restriction of variance, the External subscale is likely to lose predictive significance for decreased institutional misconduct. Finally, the large variation in interrater reliability reported for the External subscale (Supplemental Figure 2) may indicate some issues in the utility of its item descriptions and coding instructions, which may attenuate its predictive performance.

Only examining absolute total and subscale scores diminishes the true dynamic nature of the SAPROF(-YV/-SO). Rather than achieving a maximum total score, the goal should be to map changes in protective factors over the resocialization period (de Vries Robbé et al., 2020). The results of our meta-analysis support this view, showing that changes on the SAPROF are associated with decreased

violent and general recidivism after controlling for baseline risk levels. This is an important finding because it underscores the utility of the SAPROF in guiding clinicians during risk management and rehabilitation by monitoring changes in protection that are directly associated with decreased recidivism. That being said, the one study that investigated change scores on the SAPROF-YV (Koh et al., 2021), did not find these to be a significant predictor of decreased recidivism. Further research on the risk-mitigating effects of changes in protective factors, with multiple time points in a prospective design, is needed to draw firm conclusions.

The SAPROF(-YV/-SO) is intended to complement risk-focused assessment tools. This naturally raises the question of whether protective factors can incrementally predict decreases in recidivism or institutional misconduct beyond risk factors alone. Our meta-analysis of studies that examined the incremental validity of the SAPROF(-YV/-SO) suggests that they can. With the exception of sexual recidivism (SAPROF), all recidivism and misconduct outcomes were incrementally predicted by the three tools. That is, increasing levels of protection significantly offset an individual's risk for recidivism or institutional misconduct. The fact that the SAPROF did not incrementally predict sexual recidivism, while the SAPROF-SO did, corroborates the argument that these tools should be used as intended by their developers. Additional findings supporting the combined use of protective and risk factors come from Olver and Riemer (2021) and Lovatt et al. (2022), who used survival analyses to display trajectories of recidivism as a function of PCL-R risk levels and SAPROF protection levels as well as VRS-YV risk levels and SAPROF-YV protection levels, respectively. Their plots show that survival curves become progressively flatter (i.e., survival rate diminishes less over time) with higher levels of protection in both high- and low-risk offenders. These results are not surprising considering that structural equation models suggest that protective factors and risk factors are not simply opposite ends of the same continuum, but measure independent latent constructs (Klepfisz et al., 2020). Importantly, adding a focus on protective factors in offender rehabilitation has been shown to enhance the working alliance between patients and staff, by fostering a more holistic approach to risk management (de Vries Robbé & Stam, 2012; Wanamaker et al., 2018).

Our moderation analyses revealed two significant moderators, both influencing the SAPROF's prediction of decreased violent misconduct. First, larger base rates were associated with lower predictive performance. This finding is difficult to interpret and requires further investigation. Although it is possible that the SAPROF performs particularly well in situations where the prevalence of incidents is low, this argument is not supported by our results showing better performance in institutional settings, where the average base rate was higher than in the community ($Mdn_{\text{Institution}} = 28\%$ vs. $Mdn_{\text{Community}} = 20\%$). Second, articles from Asia reported significantly higher effect sizes compared to studies from North America, but not compared to studies from Europe. However, this finding should be interpreted with caution because (a) only two effect sizes from Kashiwagi et al. (2018) "represented" Asia in this particular meta-analytical model and (b) all other models (also including samples from Singapore) did not yield a significant moderation by country. All remaining moderators were either not significant (time at risk and setting) or could not be investigated due to limited data (lack of women in included studies).

Limitations

In this section, we discuss several limitations of our meta-analysis deemed important to take into account in planning future SAPROF(-YV/-SO) validation studies. First, there was a lack of women in the validation samples. Only one article (de Vogel et al., 2019) validated the SAPROF in an all-female sample. This is concerning, because not only do the manuals of the three tools state that they are also suitable for women but also the results of de Vogel and colleagues challenge the assertion in the manuals. It follows that our findings cannot simply be generalized to female offenders.

Second, our analyses were limited by the statistical parameters reported in the available research. No article explicitly reported common measures of calibration. Although we were able to obtain some calibration data after contacting the authors, these findings are preliminary. Thus, we can only draw firm conclusions regarding the SAPROF(-YV/-SO)'s ability to discriminate between recidivists and nonrecidivists, but not about how well their predictions match the true likelihood that an individual will refrain from violence or misconduct (Cook, 2007). In other words, it remains unclear whether the predictions are accurate or whether the estimates are systematically too high or too low.

Third, some of our results were based on single effect sizes. For instance, all data concerning the SAPROF-SO were extracted from Nolan et al. (2022). This should not be surprising given the recent development of the SAPROF-SO. Nevertheless, an effect size estimate based on a single study is evidently not a "meta-analysis."

Finally, the PROBAST indicated a high risk of bias in most articles, which was mainly due to too few outcome events in a sample. Despite this result, our findings should not be dismissed because the risk of bias was less pronounced in the other three domains (i.e., participants, predictors, and outcome). Furthermore, the PROBAST has been criticized for being too harsh in the context of risk assessment validation studies, primarily because it was developed to assess prediction models in medicine, which tend to have simpler outcome definitions (Fazel et al., 2022).

Implications for Future Research

Because our meta-analysis uncovered differences in the predictive performance of the SAPROF and SAPROF-YV between institutional and community settings, as well as across their subscales, it is recommended that future research also examines predictive validity at the item level. Specific items may differ in their predictive utility depending on the context or phase of the individual's reintegration into the community. A more fine-grained examination of the predictive performance of individual SAPROF(-YV/-SO) items could also inform the effective management and prevention of violent behavior.

At present, most validation studies of the SAPROF(-YV/-SO) have been *retrospective* file studies ($n = 34$), that is, the tool was rated on the basis of file information that was not collected for the purpose of coding the SAPROF(-YV/-SO). This is far from ideal for testing the performance of a tool intended to serve as a *prospective* guide to treatment and risk management at the individual level. Thus, prospective studies, in which the instrument is used as intended, are sorely needed.

The SAPROF is intended as a dynamic measure of protection against violent behavior and should be sensitive to an individual's

responsiveness to intervention. To date, however, only three studies have investigated SAPROF change scores (Coupland & Olver, 2020; de Vries Robbé, de Vogel, Douglas, et al., 2015; Olver & Riemer, 2021), and only one article has assessed SAPROF-YV change scores (Koh et al., 2021). Further research is urgently needed as it will ultimately provide deeper insight and useful guidance for successful violence risk management. According to Lovatt et al. (2022), such research should also include the assessment of concurrent changes in risk factors. Moreover, investigating the predictive validity of the SAPROF(-YV/-SO) in relation to positive community outcomes (e.g., quality of life, social relationships, employment), as proposed by de Vries Robbé et al. (2020), would be a valuable new approach to assessing treatment effectiveness. So far, this has only been addressed in one study (Coupland & Olver, 2020).

Conclusion

The SAPROF exhibits good interrater reliability as well as moderate to good predictive validity for violent incidents after discharge into the community and within institutional settings. Its sister tools, the SAPROF-YV and the SAPROF-SO, perform similarly well and are best used in their respective target groups. When combined with a risk-focused assessment instrument, protective factors add incremental validity and contribute to a more holistic risk assessment and risk management plan. Furthermore, changes in protection from pre-to-posttreatment predict decreased violent recidivism beyond baseline risk. This meta-analysis highlights several limitations of the current research base on the SAPROF(-YV/-SO), notably a lack of information on calibration parameters, predictive performance at the item level, prospective designs, and research in female samples. Future research that addresses these limitations has the potential to provide more fine-grained insights into the predictive performance of the SAPROF(-YV/-SO). This will likely contribute to our understanding of the role of protective factors in effective risk management and prevention of future violence.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Abbiati, M., Azzola, A., Palix, J., Gasser, J., & Moulin, V. (2017). Validity and predictive accuracy of the structured assessment of protective factors for violence risk in criminal forensic evaluations: A Swiss cross-validation retrospective study. *Criminal Justice and Behavior*, *44*(4), 493–510. https://doi.org/10.1177/0093854816677565

Abbiati, M., Golay, P., Gasser, J., & Moulin, V. (2020). Protective factor assessments: What are we measuring?—An investigation of the internal validity of the structured assessment of protective factors for violence risk. *Criminal Justice and Behavior*, *47*(4), 383–398. https://doi.org/10.1177/0093854819901157

*Abbiati, M., Palix, J., Gasser, J., & Moulin, V. (2019). Predicting physically violent misconduct in prison: A comparison of four risk assessment instruments. *Behavioral Sciences & the Law*, *37*(1), 61–77. https://doi.org/10.1002/bsl.2364

*Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. G. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: Prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BMC Psychiatry*, *13*(1), Article 197. https://doi.org/10.1186/1471-244X-13-197

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). LexisNexis.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, *52*(1), 7–27. https://doi.org/10.1177/0011128705281756

*Awrey, M. (2021). *Predicting recidivism versus desistance in adolescents with sexual offenses using tools developed for adults: The Static-99 and SAPROF* [Doctoral dissertation]. University of Windsor.

*Bhanwer, A. K. (2016). *The structured assessment of protective factors for violence risk—Youth version (SAPROF-YV): The association between protective factors and aggression in adolescents* [Unpublished master thesis]. Simon Fraser University.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the structured assessment for violence risk in youth (SAVRY)*. Psychological Assessment Resources.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*(6), 567–590. https://doi.org/10.1177/0093854809333610

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

*Chu, C. M., Xu, X., Li, D., Ruby, K., & Chng, G. S. (2020). The utility of SAPROF-YV ratings for predicting recidivism in male youth under community supervision in Singapore. *Criminal Justice and Behavior*, *47*(11), 1409–1427. https://doi.org/10.1177/0093854820949595

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

*Coid, J. W., Ullrich, S., Kallis, C., Freestone, M., Gonzalez, R., Bui, L., Igoumenou, A., Constantinou, A., Fenton, N., Marsh, W., Yang, M., DeStavola, B., Hu, J., Shaw, J., Doyle, M., Archer-Power, L., Davoren, M., Osumili, B., McCrone, P., … Bebbington, P. (2016). Improving risk management for violence in mental health services: A multimethods approach. *Programme Grants for Applied Research*, *4*(16), 185–202. https://doi.org/10.3310/pgfar04160

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, *115*(7), 928–935. https://doi.org/10.1161/CIRCULATIONAHA.106.672402

*Coupland, R. B. A., & Olver, M. E. (2020). Assessing protective factors in treated violent offenders: Associations with recidivism reduction and positive community outcomes. *Psychological Assessment*, *32*(5), 493–508. https://doi.org/10.1037/pas0000807

*de Ruigh, E. L., Kleeven, A. T. H., Jansen, L. M. C., de Vries Robbé, M., Vermeiren, R. R. J. M., Mulder, E. A., de Paauw, B., van de Ven, P. M., & Popma, A. (2021). Predicting youth reoffending after incarceration: Added value of protective factors and heart rate variability. *Journal of Forensic Psychiatry & Psychology*, *32*(4), 449–480. https://doi.org/10.1080/14789949.2020.1860250

de Ruiter, C., & Nicholls, T. L. (2011). Protective factors in forensic mental health: A new frontier. *International Journal of Forensic Mental Health*, *10*(3), 160–170. https://doi.org/10.1080/14999013.2011.600602

*de Vogel, V., Bruggeman, M., & Lancel, M. (2019). Gender-sensitive violence risk assessment: Predictive validity of six tools in female forensic psychiatric patients. *Criminal Justice and Behavior*, *46*(4), 528–549. https://doi.org/10.1177/0093854818824135

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2012). *SAPROF: Guidelines for the assessment of protective factors for violence risk* (2nd ed.). Van Der Hoeven Stichting.

*de Vries Robbé, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study of the SAPROF. *International Journal of Forensic Mental Health*, *10*(3), 178–186. https://doi.org/10.1080/14999013.2011.600232

*de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry & Psychology*, *24*(4), 440–457. https://doi.org/10.1080/14789949.2013.818162

*de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior*, *39*(1), 53–61. https://doi.org/10.1037/lhb0000089

*de Vries Robbé, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015). Assessing protective factors for sexually violent offending with the SAPROF. *Sexual Abuse*, *27*(1), 51–70. https://doi.org/10.1177/1079063214550168

de Vries Robbé, M., de Vogel, V., & Veldhuizen, A. (2020). The structured assessment of protective factors for violence risk (SAPROF). In K. S. Douglas & R. K. Otto (Eds.), *Handbook of violence risk assessment* (2nd ed., pp. 410–437). Routledge. https://doi.org/10.4324/9781315518374-21

*de Vries Robbé, M., de Vogel, V., Wever, E. C., Douglas, K. S., & Nijman, H. L. I. (2016). Risk and protective factors for inpatient aggression. *Criminal Justice and Behavior*, *43*(10), 1364–1385. https://doi.org/10.1177/0093854816637889

de Vries Robbé, M., Geers, M. C. K., Stapel, M., Hilterman, E. L. B., & de Vogel, V. (2015). *SAPROF–Youth version: Guidelines for the assessment of protective factors for violence risk in juveniles*. Van der Hoeven Kliniek.

de Vries Robbé, M., & Stam, J. (2012). Protective factors for violence risk: The value for clinical practice. *Psychology*, *3*(12), 1259–1263. https://doi.org/10.4236/psych.2012.312A187

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence—User guide*. Simon Fraser University.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice*, *81*, Article 101902. https://doi.org/10.1016/j.jcrimjus.2022.101902

*Haines, A., Brown, A., Javaid, S. F., Khan, F., Noblett, S., Omodunbi, O., Sadiq, K., Zaman, W., & Whittington, R. (2018). Assessing protective factors for violence risk in U.K. General mental health services using the structured assessment of protective factors. *International Journal of Offender Therapy and Comparative Criminology*, *62*(12), 3965–3983. https://doi.org/10.1177/0306624X17749449

Hajian-Tilaki, K. O., & Hanley, J. A. (2002). Comparison of three methods for estimating the standard error of the area under the curve in ROC analysis of quantitative data. *Academic Radiology*, *9*(11), 1278–1285. https://doi.org/10.1016/S1076-6332(03)80561-5

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*(1), 119–136. https://doi.org/10.1023/A:1005482921333

Hardin, K., & Scurich, N. (2022). The dark figure of violence committed by discharged psychiatric inpatients. *Journal of Forensic Practice*, *24*(3), 229–240. https://doi.org/10.1108/JFP-11-2021-0058

Hare, R. D. (2003). *Hare psychopathy checklist-revised: PCL-R* (2nd ed.). Multi-Health Systems.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 1–17). Routledge.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hoge, R. D., & Andrews, D. A. (2004). *Youth level of service/case management inventory (YLS/CMI) user's manual*. Multi-Health Systems.

Huang, Y., Li, W., Macheret, F., Gabriel, R. A., & Ohno-Machado, L. (2020). A tutorial on calibration measurements and calibration models for clinical prediction models. *Journal of the American Medical Informatics Association: JAMIA*, *27*(4), 621–633. https://doi.org/10.1093/jamia/ocz228

Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders: The Service Planning Instrument. *Criminal Justice and Behavior*, *42*(3), 321–338. https://doi.org/10.1177/0093854814547041

*Judges, R. C. (2016). *An exploration into the value of protective factors in violence risk assessment of psychiatric inpatients* [Doctoral dissertation]. University of Nottingham. http://eprints.nottingham.ac.uk/32959/1/JUDGES%20R%20RESEARCH%20THESIS%20AMENDED.pdf

*Kashiwagi, H., Kikuchi, A., Koyama, M., Saito, D., & Hirabayashi, N. (2018). Strength-based assessment for future violence risk: A retrospective validation study of the structured assessment of PROtective factors for violence risk (SAPROF) Japanese version in forensic psychiatric inpatients. *Annals of General Psychiatry*, *17*(1), Article 5. https://doi.org/10.1186/s12991-018-0175-5

*Kleeven, A. T. H., de Vries Robbé, M., Mulder, E. A., & Popma, A. (2022). Risk assessment in juvenile and young adult offenders: Predictive validity of the SAVRY and SAPROF-YV. *Assessment*, *29*(2), 181–197. https://doi.org/10.1177/1073191120959740

*Klein, V., Rettenberger, M., Yoon, D., Köhler, N., & Briken, P. (2015). Protective factors and recidivism in accused juveniles who sexually offended. *Sexual Abuse*, *27*(1), 71–90. https://doi.org/10.1177/1079063214554958

Klepfisz, G., Daffern, M., & Day, A. (2017). Understanding protective factors for violent reoffending in adults. *Aggression and Violent Behavior*, *32*, 80–87. https://doi.org/10.1016/j.avb.2016.12.001

*Klepfisz, G., Daffern, M., Day, A., Lloyd, C. D., & Woldgabreal, Y. (2020). Latent constructs in the measurement of risk and protective factors for violent reoffending using the HCR-20^{v3} and SAPROF: Implications for conceptualizing offender assessment and treatment planning. *Psychology, Crime & Law*, *26*(1), 93–108. https://doi.org/10.1080/1068316X.2019.1634197

*Koh, L. L., Day, A., Klettke, B., Daffern, M., & Chu, C. M. (2021). Youth violence assessment instruments: Are they sensitive to change and are changes related to recidivism? *Psychology, Crime & Law*, *28*(4), 416–433. https://doi.org/10.1080/1068316X.2021.1909019

*Koh, L. L., Day, A., Klettke, B., Daffern, M., & Chu, C. M. (2022). The predictive validity of three youth violence assessment instruments: The SAVRY, VRS-YV, and SAPROF-YV. *International Journal of Offender Therapy and Comparative Criminology*, *66*(2–3), 168–185. https://doi.org/10.1177/0306624X20970887

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lee, S. C., Hanson, R. K., & Blais, J. (2020). Predictive accuracy of the Static-99R and Static-2002R risk tools for identifying Indigenous and White individuals at high risk for sexual recidivism in Canada. *Canadian Psychology*, *61*(1), 42–57. https://doi.org/10.1037/cap0000182

*Lovatt, K. M., Stockdale, K. C., & Olver, M. E. (2022). Dynamic violence risk, protective factors, and therapeutic change in a gender and ethnoculturally diverse sample of court-adjudicated youth. *Psychological Assessment*, *34*(2), 105–124. https://doi.org/10.1037/pas0001071

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*(1), 489–513. https://doi.org/10.1146/annurev-clinpsy-021815-092945

*Neil, C., O'Rourke, S., Ferreira, N., & Flynn, L. (2020). Protective factors in violence risk assessment: Predictive validity of the SAPROF and HCR-20. *International Journal of Forensic Mental Health*, *19*(1), 84–102. https://doi.org/10.1080/14999013.2019.1643811

*Nolan, T., & Willis, G. M., Thornton, D., Kelley, S. M., & Christofferson, S. B. (2022). Attending to the positive: A retrospective validation of the structured assessment of protective factors-sexual offence version. *Sexual Abuse*. Advance online publication. https://doi.org/10.1177/10790632221098354

*Olver, M. E., & Riemer, E. K. (2021). High-psychopathy men with a history of sexual offending have protective factors too: But are these risk relevant and can they change in treatment? *Journal of Consulting and Clinical Psychology*, *89*(5), 406–420. https://doi.org/10.1037/ccp0000638

O'Shea, L. E., & Dickens, G. L. (2016). Performance of protective factors assessment in risk prediction for adults: Systematic review and meta-analysis. *Clinical Psychology: Science and Practice*, *23*(2), 126–138. https://doi.org/10.1037/h0101736

*Oziel, S., Marshall, L. A., & Day, D. M. (2020). Validating the SAPROF with forensic mental health patients. *Journal of Forensic Psychiatry & Psychology*, *31*(5), 667–698. https://doi.org/10.1080/14789949.2020.1785525

*Páv, M., Sebalo, I., Vňuková, M., Pabiánová, Š., Málová, V., Hollý, M., & Perkins, D. (2022). Predicting discharge from long-term forensic treatment: Patients characteristics, protective factors, needs and treatment-related factors study in the Czechia. *Journal of Forensic Psychiatry & Psychology*, *33*(1), 89–111. https://doi.org/10.1080/14789949.2022.2027995

*Persson, M., Belfrage, H., Fredriksson, B., & Kristiansson, M. (2017). Violence during imprisonment, forensic psychiatric care, and probation: Correlations and predictive validity of the risk assessment instruments COVR, LSI-R, HCR-20, and SAPROF. *International Journal of Forensic Mental Health*, *16*(2), 117–129. https://doi.org/10.1080/14999013.2016.1266420

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, *29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, *24*(5), 595–605. https://doi.org/10.1023/A:1005575113507

*Scholten, D. S. (2017). *Risk and protective factors for inpatient youth violence: The predictive value of the SAVRY and the SAPROF-YV in youth with and without a psychiatric disorder* [Unpublished master thesis]. University of Leiden.

Scurich, N. (2018). The case against categorical risk estimates. *Behavioral Sciences & the Law*, *36*(5), 554–564. https://doi.org/10.1002/bsl.2382

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, *31*(1), 8–22. https://doi.org/10.1002/bsl.2052

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law*, *31*(1), 55–73. https://doi.org/10.1002/bsl.2053

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513. https://doi.org/10.1016/j.cpr.2010.11.009

*Soeterbroek, D. (2019). *Predictive validity of the SAVRY and the SAPROF-YV for aggressive incidents in Dutch juvenile justice institutions* [Unpublished master thesis]. Tilburg University. http://arno.uvt.nl/show.cgi?fid=150170

*Tong, K., Basrak, N., Quane, T., Murray, M., Kennedy, H. G., & Davoren, M. (2022). *Protective factor for violence in a secure hospital: Global function, treatment response and recovery* [Poster]. Trinity College Dublin, Ireland.

*Tozdan, S., Briken, P., Yoon, D., & von Franqué, F. (2016). Risiko- und Schutzfaktoren bei sexualdelinquent gewordenen Menschen: Vorhersage erneuter Straftaten und Veränderungen im Behandlungsverlauf. *Psychiatrische Praxis*, *43*(3), 154–159. https://doi.org/10.1055/s-0034-1387404

*Turner, D., Rettenberger, M., Yoon, D., Klein, V., Eher, R., & Briken, P. (2016). Risk assessment in child sexual abusers working with children. *Sexual Abuse*, *28*(6), 572–596. https://doi.org/10.1177/1079063214564390

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring gender differences in the utility of strength-based risk assessment measures. *International Journal of Forensic Mental Health*, *15*(2), 149–163. https://doi.org/10.1080/14999013.2016.1170739

Walters, G. D. (2012). Criminal thinking and recidivism: Meta-analytic evidence on the predictive and incremental validity of the Psychological Inventory of Criminal Thinking Styles (PICTS). *Aggression and Violent Behavior*, *17*(3), 272–278. https://doi.org/10.1016/j.avb.2012.02.010

Wanamaker, K. A., Jones, N. J., & Brown, S. L. (2018). Strengths-based assessments for use with forensic populations: A critical review. *International Journal of Forensic Mental Health*, *17*(2), 202–221. https://doi.org/10.1080/14999013.2018.1451414

Ward, T., & Fortune, C.-A. (2013). The good lives model: Aligning risk reduction with promoting offenders' personal goals. *European Journal of Probation*, *5*(2), 29–46. https://doi.org/10.1177/206622031300500203

*Willis, G. M., Kelley, S. M., & Thornton, D. (2020). Are protective factors valid constructs? Interrater reliability and construct validity of proposed protective factors against sexual reoffending. *Criminal Justice and Behavior*, *47*(11), 1448–1467. https://doi.org/10.1177/0093854820941039

Willis, G. M., Thornton, D., Kelley, S. M., & de Vries Robbé, M. (2017). *The structured assessment of protective factors for violence risk—Sexual offending. SAPROF-SO* [Unpublished manual].

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & the PROBAST Group†. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51–58. https://doi.org/10.7326/M18-1376

Wong, S. C. P., Lewis, K., Stockdale, K. C., & Gordon, A. (2004–2011). *The violence risk scale-youth version* [Unpublished manuscript]. University of Saskatchewan.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*(5), 740–767. https://doi.org/10.1037/a0020473

*Yoon, D., Turner, D., Klein, V., Rettenberger, M., Eher, R., & Briken, P. (2018). Factors predicting desistance from reoffending: A validation study of the SAPROF in sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, *62*(3), 697–716. https://doi.org/10.1177/0306624X16664379

*Zeng, G., Chu, C. M., & Lee, Y. (2015). Assessing protective factors of youth who sexually offended in singapore: Preliminary evidence on the utility of the DASH-13 and the SAPROF. *Sexual Abuse*, *27*(1), 91–108. https://doi.org/10.1177/1079063214561684

Received January 7, 2022
Revision received August 26, 2022
Accepted August 29, 2022 ■