

Performance of protective factors assessment in risk prediction for adults: systematic review and meta-analysis

O'Shea, Laura E.
Dickens, Geoffrey L.

This is the peer reviewed version of the following article:

O'Shea, L. E. and 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): pp.126-138,

which has been published in final form at <https://dx.doi.org/10.1111/cpsp.12146>.

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Performance of Protective Factors Assessment in Risk Prediction for Adults: Systematic
Review and Meta-Analysis

Laura E O'Shea^{ab}, Geoffrey L Dickens^{*ac}

^aSt Andrew's Academic Department, Billing Road, Northampton, NN1 5DG, UK

^bInstitute of Psychiatry, King's College London, 16 De Crespigny Park, London, SE5 8AF,
UK

^cSchool of Social and Health Sciences, Abertay University, Bell Street, Dundee, DD1 1HG,
UK

***Correspondence:** Geoffrey L. Dickens, School of Social and Health Sciences, Abertay
University, Bell Street, Dundee, Scotland, DD1 1HG, UK. Email: g.dickens@abertay.ac.uk.
Tel: +441382 308257

Abstract

Protective factors are neglected in risk assessment in adult psychiatric and criminal justice populations. This review investigated the predictive efficacy of selected tools that assess protective factors. Five databases were searched using comprehensive terms for records up to June 2014, resulting in 17 studies ($N = 2,198$). Results were combined in a multilevel meta-analysis using the R (R Core Team, 2015) metafor package (Viechtbauer, 2010). Prediction of outcomes was poor relative to a reference category of violent offending, with the exception of prediction of discharge from secure units. There were no significant differences between the predictive efficacy of risk scales, protective scales and summary judgments. Protective factor assessment may be clinically useful but more development is required; claims that it is therapeutically beneficial require testing.

Keywords: risk assessment, protective factors, violence, self-harm

Performance of Protective Factors Assessment in Risk Prediction for Adults: Systematic
Review and Meta-Analysis

Structured professional judgment schemes are now the gold standard technique for violence risk assessment in mental health and criminal justice settings (National Institute for Mental Health in England, 2004). However, they have been widely criticised (e.g., Hart, 2001; [Rogers, 2000](#)) because their almost exclusive focus on negative factors that increase risk may contribute to a negative bias, overestimation of risk, and ultimately to client stigmatization ([Rogers, 2000](#)) and unnecessary restrictions of personal freedom (de Ruiter & Nicholls, 2011; [Rogers, 2000](#)). In contrast, protective factors, the “conditions or attributes of individuals, families, communities, or the larger society that reduce or eliminate risk” (Child Welfare Information Gateway, Children's Bureau, FRIENDS National Resource Center For Community-Based Child Abuse Prevention, & Families, 2014; [p. 4](#)), may mitigate the risk factors held by an individual. Their consideration, in conjunction with risk factors should facilitate a more balanced evaluation process (Laub & Lauritsen, 1994). Further, consideration of protective factors may assist in identification of treatment targets (Nonstad et al., 2010) and facilitate the development of therapeutic relationships (de Ruiter & Nicholls, 2011).

A number of systematic reviews and meta-analyses have aimed to determine the efficacy of violence risk assessments (e.g., O'Shea, Mitchell, Picchioni, & Dickens, 2013; [Singh, Grann, & Fazel, 2011](#); [Singh, Serper, Reinharth, & Fazel, 2011](#); [Yang, Wong, & Coid, 2010](#)). While some of these reviews have included tools that assess protective factors (e.g., [Singh, Grann, et al., 2011](#)), they have not been the primary focus. One possible reason for the underdevelopment of assessment tools that are based on protective factors, in comparison to those based on risk factors, may be due to difficulties in conceptualisation of the former. The concept of risk is widely understood; risk factors are those attributes that are associated with

increased likelihood of negative outcomes, such as morbidity or mortality. In the medical context, for example, tobacco smoking is a risk factor for lung cancer; in the forensic psychiatric/psychological context we are concerned with risk factors for behaviours such as violence, self-harm and self-neglect (Jessor, Van Den Bos, Vanderryn, Costa, & Turbin, 1995). A large empirical literature has developed around risk factors, at least for violence. For example, it is widely agreed that factors including psychopathy, previous violence, and major mental disorder are associated with increased risk for violence (e.g., [Hart, 1998](#); [Monahan, 1992](#)). However, there is less consensus about the concept and operationalization of protective factors (Jessor et al., 1995). Protective factors have been variously defined as A) simply the absence of a risk factor (e.g., no previous violence), B) a factor that lies at the opposite end of a continuum to a risk factor (e.g., history of kindness as opposed to history of violence measured on a single continuum), or C) conceptually distinct, with no corresponding risk factor (e.g., history of kindness irrespective of history of violence) (O'Shea & Dickens, 2014). In both research and clinical practice, the consideration of protective factors in adults is lagging behind what has been achieved in adolescents; Viljoen, McLachlan, and [Vincent \(2010\)](#) found that risk assessment reports with adolescent offenders were significantly more likely to include protective factors than those completed with adults. Rogers (2000) described research studies in adult populations as one-sided, such that they emphasise risk factors to the detriment of protective factors. This may in part be because adolescent tools were developed at a later time point to adult tools, allowing them to learn from the existing research and criticisms in the adult literature, or may be due to a greater focus on rehabilitation or therapeutic optimism in juvenile practice ([Viljoen et al., 2010](#)).

Contribution of the Current Study

This gap in the literature has been addressed in recent years with the development of a number of tools that are designed to support the assessment of protective factors in adults for

a range of outcomes, but no studies have aimed to systematically identify and evaluate all the existing empirical literature for a range of protective factor assessment tools. Therefore, the aim of the current paper was to 1) identify protective factor (PF) assessment tools in the area of forensic psychiatry/psychology, 2) systematically identify studies assessing the predictive efficacy of PF assessment tools in adults, and 3) conduct a meta-analysis to examine their ability to predict their intended outcomes.

Method

Review Protocol

This review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement (Moher, Liberati, Tetzlaff, Altman, & The Prisma Group, 2009) which is designed to facilitate transparent reporting. The studies selected for inclusion were identified and retrieved as part of a wider literature search strategy about the accuracy of assessments utilising protective factors in both adult and adolescent populations; a meta-analysis of the role of protective factors in adolescents will be conducted in due course.

Tool Selection

Based on reviews of the literature about protective factors in forensic risk evaluation (e.g., de Ruiter & Nicholls, 2011; de Vries Robbé, de Vogel, & Stam, 2012), and from a search of computerised databases for studies involving assessment of protective factors for multiple adverse outcomes, we identified 17 instruments that aim to assist practitioners with the assessment of protective factors for the prediction of a range of adverse outcomes occurring in both institutional/correctional settings, and in the community following discharge or release: the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009), the START: Adolescent Version (START: AV; Viljoen, Cruise, Nicholls, Desmarais, & Webster, 2012), the Structured Assessment of

PROtective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2012), the Structured Assessment of Violence Risk in Youth (SAVRY; [Borum, 2006](#)), the Dangerousness UNDERstanding, Recovery and Urgency Manual 3 and 4 (DUNDRUM-3/4; Kennedy, O'Neill, Flynn, & Gill, 2010), the Inventory of Offender Risk, Needs, and Strengths (IORNS; Miller, 2006), the San Diego Regional Resiliency Check-up (SDRRC; [Turner & Fain, 2006](#)), the Multiplex Empirically Guided Inventory of Ecological Aggregates for Assessing Sexually Abusive Children and Adolescents (MEGA; Miccio-Fonseca, 2013), the Columbia Suicide Severity Rating Scale (CSSRS; Posner et al., 2011), and the Reasons for Living Inventory (RFL; Linehan, Goodstein, Nielsen, & Chiles, 1983), and its variations: BRFL (Ivanoff, Jang, Smyth, & Linehan, 1994), RFL-A ([Gutierrez, Osman, Kopper, & Barrios, 2000](#)), BRFL-A (Osman et al., 1996), RFL-YA ([Gutierrez et al., 2002](#)), RFL-OA (Edelstein et al., 2009), and RFL-CS ([Westefeld, Cardin, & Deaton, 1992](#)).

Search Strategy

The aim of the literature search was to identify all empirical studies conducted to evaluate the predictive validity of one or more of the above identified tools for any of a range of adverse outcomes. Adverse outcomes of interest included all those covered by the START Outcome Scale (Nicholls et al., 2007), i.e., aggression, self-harm, suicide, unauthorised leave, self-neglect, victimisation, substance abuse, sexual offending, and stalking; in addition we were interested in studies reporting on risk outcomes including fire-setting/arson, recidivism, recovery, and programme completion. Multiple computerised databases (PsycINFO, Scopus, Web of Knowledge, Cochrane Library, CINAHL, and NCJRS) were searched for articles published prior to June 25, 2014. Search terms pertaining to risk outcomes were paired with terms relating to the identified risk assessment tools (see example in Appendix A, available on-line). All studies, including grey literature such as unpublished manuscripts, theses, and conference presentations were eligible for inclusion. Wild card terms (i.e. those ending with

*) were used to include all permutations of the stem of each search term. Additional studies were located through personal correspondence and hand searching of reference lists of papers identified at the previous step.

Study Selection

The title and abstract of all articles returned by the search were reviewed by the first author to identify those that documented an empirical investigation of one or more of the included PF assessment tools. Eligibility of full-text articles for inclusion was then assessed by the first author; the second author independently reviewed 25% of the articles to establish inter-rater reliability; kappa = .91. Discrepancies ($n=1$) were resolved through discussion.

Inclusion and exclusion criteria. In order to be included in the current study, an article must have described an empirical, prospective or pseudo prospective account of the predictive efficacy of one or more of the assessment tools, for any of the outcomes intended by the tools' authors, in an adult population. An Area under the Curve (AUC) value, or sufficient statistical information to calculate such a value, must have been included. Studies were excluded if they did not contain original primary research (i.e., reviews), were written in non-English language, or if the assessment instruments used were scored in a manner that differed from that recommended in the tools' manual. If multiple articles contained overlapping samples, the article with the largest sample size was retained; however, if different tools or outcomes were examined, both articles were retained.

Data extraction. For each included study, we extracted the following information: number of participants, country of data collection, study setting, length of follow-up period, assessment tool(s) used, type of adverse outcomes measured, demographic and clinical characteristics of the sample, and the AUC value for each PF assessment scale-outcome combination, and for any summary judgement based, at least in part, on the assessment of PFs. Additionally, we extracted AUC values for predictions made using the risk scale of any

of the included PF assessment tools (e.g. the START vulnerability scale) in order to provide contextual information about the magnitude of effect sizes obtained for protective factors.

Risk of Bias

The quality of all primary studies was assessed independently by both authors using guidelines suggested by Haney and colleagues (2012); weighted kappa =.95, discrepancies ($n=9$) resolved through discussion. This procedure examined sources of bias including clear definition and appropriate sampling of the study populations, valid/reliable outcome measurements and risk-assessment tools, independence of outcome assessment from risk assessment, and whether confounders were adequately controlled. Each domain is rated as “yes”, “unclear”, or “no” and overall risk of bias (low, unclear, high) is based on the authors’ opinion regarding the likelihood that any bias has lowered the confidence that can be placed on the results; all studies were included in the subsequent meta-analysis irrespective of bias.

Data Synthesis

The vast majority of studies inverted scores on protective factor scales, such that higher scores represented fewer protective factors, or used these scales to predict outcome absence; two studies did not adopt either of these procedures and we therefore inverted AUC values in these cases in order to facilitate combination of results. We extracted effect sizes for each predictor-outcome combination; therefore, the data for the current review has a four-level structure, with participants nested within scales/outcomes, scales/outcomes nested within studies, and studies nested within author groups (as four author groups contributed multiple papers to the current review). A number of studies (Chu, Thomas, Ogloff, & Daffern, 2013; de Vries Robbe, de Vogel, & Douglas, 2013; de Vries Robbe, de Vogel, Koster, & Bogaerts, 2014; Inett, Wright, Roberts, & Sheeran, 2014) reported multiple follow-up periods; in these cases, data were taken from the time point with the maximum sample size. Risk outcomes were defined in each individual primary study; therefore some variation

would be expected in the definitions used across studies. Effect sizes for inpatient physical aggression against others, and for inpatient physical aggression against objects were pooled with those for any inpatient physical aggression in order to minimise the number of outcomes reported on and because some studies reported only this composite outcome. Where a study reported on both physical aggression towards others, and physical aggression towards objects, a mean effect size was calculated. Similarly, inpatient self-harm and inpatient suicide were combined into inpatient self-harm/suicide. Inpatient verbal abuse and inpatient verbal threats were combined into a verbal aggression effect size.

The meta-analysis was conducted in R (R Core Team, 2015). AUC values were converted to Cohen's d values according to the tables provided by Rice and Harris (2005); Cohen's d values are a commonly used measure of effect size well suited to random effects models (Yang, et al., 2010). However, Cohen's d values are biased and provide an overestimation of the true effect, especially when sample sizes are small (Lakens, 2013); therefore, d values were converted to Hedge's g values, which corrects for bias, using the formulas from the compute.es package (Del Re, 2013).

Hedge's g values were weighted by the inverse variance weight and pooled using the rma.mv function in the metafor package (Viechtbauer, 2010); this function conducts a multilevel meta-analysis which can account for the non-independence of effect sizes caused by the nested data structure. Random effects were added for both outcome and scale, nested within study, to account for the fact that estimates derived from the same scale, or predicting the same outcome, are likely correlated with each other. Similarly, random effects were added for study nested within author, and for author, as effect sizes within studies and authors are likely to be more similar than those between studies/authors. Effect sizes were estimated for each outcome of interest, relative to violent offending, as the ability of structured professional judgment schemes for predicting this outcome is well established (e.g., Singh,

Grann, et al., 2011); effect sizes for protective scales and summary judgments were estimated with risk scales as the reference category. Gender, coded as the percentage of the sample that is female, was also included as a moderator, as previous research (O'Shea & Dickens, 2015a; O'Shea, et al., 2013; O'Shea, Picchioni, Mason, Sugarman, & Dickens, 2014) has found that risk assessment schemes such as the HCR-20 and the START perform more accurately in women. Finally, the risk of bias as determined by the quality assessment was included to investigate the effect of bias on estimated effect size. The magnitude of effect size estimates were classified as small (0.2), moderate (0.5) and large (0.8) according to Cohen's criteria (Cohen, 1992); equivalent AUC values were presented for estimated effect sizes to facilitate comparison with previous research.

Results

Study Characteristics

Our search strategy resulted in the identification of 1016 articles, of which 671 remained after removal of duplicates (see Figure 1). Application of exclusion criteria to abstracts resulted in the exclusion of 556 records; we were unable to obtain seven theses/dissertations and one conference presentation despite attempted contact with the authors. The full texts of 107 articles were reviewed; 90 records were excluded (see Figure 1 for exclusion reasons), resulting in 17 studies for inclusion in the meta-analysis. Three articles that presented data from samples overlapping with other included studies were retained as they examined the predictive validity for different outcomes (de Vries Robbe, de Vogel, Wever, Douglas, & Nijman, 2014; de Vries Robbe, et al., 2015; O'Shea, Picchioni, & Dickens, 2015).

The total sample size was 2,198 (mean $N=122$); 15 were journal articles published between 2010 and 2014, one was a manuscript submitted for publication, and one was a master's thesis. The START ($k=12$) was the most commonly investigated assessment tool,

followed by the SAPROF ($k=4$), the DUNDRUM-3 ($k=2$), and the DUNDRUM-4 ($k=2$). None of the included studies investigated the IORNS, CSSRS, SDRRC, or any of the RFL variations; the remaining identified tools are designed to examine protective factors in adolescents and/or children and hence studies investigating their efficacy were excluded as they were not conducted in adult populations. Studies were conducted in the United Kingdom ($n=5$), Canada ($n=3$), Netherlands ($n=3$), Ireland ($n=3$), Australia ($n=2$), and Norway ($n=1$) (see Appendix B for sample characteristics, available on-line).

There were a total of 126 AUC values contributed from the 17 studies. There was wide variation in both the number (range 2-19) and magnitude of effect sizes (range -0.39-1.83). Inpatient physical aggression was the most researched outcome with nine studies examining the predictive ability of the START for this outcome. Discharge, moves to higher level of security, and sexual offending were the least researched outcomes, only being examined by one study each; predictive validity for the first two outcomes were examined by the DUNDRUM family measures while sexual offending was examined by the SAPROF. Violent offending was also only examined by SAPROF ($k=2$). Any inpatient aggression was the only outcome to be examined by both the START and the SAPROF ($k=8$); the START was the only tool used to examine the predictive ability of the remaining outcomes ($k=3 - k=6$).

Characteristics of Included Tools

The START. The START comprises 20 items, each rated in terms of protective factors (Strengths) and risk factors (Vulnerabilities) on a 3-point unipolar scale (0 - no strengths/ vulnerabilities in this area, 1 - some strengths or vulnerabilities in this area, 2 - many strengths or vulnerabilities). The START appears to be most consistent with definition C, such that it is possible for individuals to be rated differently in terms of both Strengths and Vulnerabilities for any given item simultaneously. After item-rating, an overall risk estimate

is made for seven specific outcomes (violence, self-harm, suicide, victimisation, self-neglect, unauthorised leave, and substance abuse) based on scoring of Strengths and Vulnerabilities items, rating of the HCR-20 historical scale (Webster, Douglas, Eaves, & Hart, 1997), any other identified risk or protective factors, history of the risk outcome in question, consideration of key Strengths or critical Vulnerabilities, any signature risk signs, and, for violent and self-harm outcomes, the presence of imminence and severity. The model was developed because the authors believed there was a need to focus inter-disciplinary discussions; as a result an evidence-based scheme was devised (Webster et al., 2009). The manual details the authors' belief that there is a need to take note of positive, promotive characteristics or strengths in addition to risk evaluations. The START is intended to be used by mental health, correctional and forensic professionals with mentally and personality disordered adult clients.

The SAPROF. The SAPROF comprises 17 items deemed to be protective factors over three subscales (internal, external and motivational); 15 items are considered to be dynamic by the tools' authors. Each item is rated on a 3-point unipolar scale (0 – item is absent, 1 - item is possibly present or to a limited extent, 2 - item is definitely present). The aim is to use the tool in conjunction with other risk assessment tools, and specifically the HCR-20, to assess risk for future violence (including sexual violence) in adult offenders. An overall final protection summary judgement for future violence is made through consideration of the item ratings and consideration of any items identified as 'key'; a final risk judgment is made in conjunction with the corresponding HCR-20 risk assessment. When used as recommended in conjunction with a violence risk assessment, the conceptual model underlying SAPROF seems most consistent with definition C; i.e., that individuals hold both risk and protective factors simultaneously. The authors claim that consideration of protective factors allows for a more balanced assessment, and that the inclusion of mostly dynamic

items will provide concrete guidelines for effective and achievable treatment and interventions, positive treatment planning, risk management and clinical evaluation.

The DUNDRUM-3. The DUNDRUM-3 is a 7-item tool aimed at assessing level of programme engagement among inpatients in secure/forensic settings in order to inform decisions about their readiness for transfer to lower level of security. Each item is rated on a 5-point criterion referenced, ordinal scale (0 represents greatest engagement, and 4 represents least engagement). Scoring criteria suggest an underlying model similar to that of the SAPROF such that a higher score represents an absence of engagement while a lower score represents active engagement. The underlying model draws on elements from Maslow's (1943) hierarchy of needs, the recovery model of mental health ([Andresen, Oades, & Caputi, 2003](#)), and the trans-theoretical stages of change model ([Prochaska & DiClemente, 1983](#)). The aim is to measure the true level of engagement with services rather than simply to gauge compliance or symptom remission. It is recommended that the DUNDRUM-3 be used in conjunction with the HCR-20 or other risk assessment instruments since it does not purport to measure risk but 'something complimentary to risk' (O'Dwyer et al., 2011). Again, as with the SAPROF, the underlying conceptual model most closely resembles definition C.

The DUNDRUM-4. The DUNDRUM-4 is a 6-item tool for measuring recovery that is intended to be used in conjunction with the DUNDRUM-3 for assessment of a patient's readiness to move between levels of security. Items are intended to be dynamic and are each rated on a 5-point criterion referenced scale. Lower scores (i.e. 0) represent greater strengths or protective factors on each item. The scoring criteria for DUNDRUM-4 suggests active recovery at one pole but continued active risk rather than absence of recovery at the other such that it appears more bipolar in nature than DUNDRUM-3 (e.g., for Therapeutic Rapport 0 = *maintains professional contact regularly and spontaneously*; 4 = *may seek to secrete, deceive or subvert*) and thus the underlying model most closely resembles definition B.

Broadly, the five point scale is intended to equate to the need for an appropriate level of security (4 = *high security*, 3= *medium security*, 2= *low security*, 1 = *community or supervised setting*, 0 = *may be suitable for absolute discharge*).

Risk of Bias

The quality assessment of primary studies revealed that six studies were rated as low risk of bias, ten studies were rated as unclear risk, and one study was rated as high risk of bias (see Appendix C, available on-line). The most common causes of potential bias were having the same person collect outcomes data as completing the risk assessment, not providing an adequate description of assessor independence or blinding, or failing to provide evidence of consecutive or random sampling of participants.

Mean Weighted Effect Sizes

The results of the meta-analysis are presented in Table 1. Overall, the moderators accounted for a significant proportion of the heterogeneity of effect sizes ($Q[16]=58.56$, $p<.001$); however, there was still a significant amount of unexplained heterogeneity ($Q[109]=286.92$, $p<.001$). The estimated Hedge's g value when the moderators were equivalent to the reference categories (i.e. Outcome = violent offending, scale type = risk, bias = low) was 0.88 (AUC=.73). There was no significant effect of scale type on estimated effect size ($Q[2]=3.51$, $p=.173$); however, effect sizes were lower for protective scales and higher for summary judgments, compared with risk scales. Estimated effect sizes differed significantly as a function of outcome ($Q[11]=52.39$, $p<.001$). With the exception of discharge, all estimated effect sizes were smaller than for violent offending; this difference was significant for sexual offending (-0.47; $p=.021$), inpatient self-harm (-0.68; $p=.004$), inpatient substance abuse (-0.53; $p=.035$), inpatient self-neglect (-0.50; $p=.037$), and inpatient victimisation (-0.70; $p=.004$). The proportion of females in the sample did not moderate magnitude of effect sizes ($Q[1]=1.02$, $p=.313$) and there was no significant effect of bias

($Q[2]=0.46, p=.796$); however, estimated effect sizes were slightly larger for studies with a high risk of bias ($0.16; p=.695$).

Estimated mean weighted Hedge's g values for each outcome are presented in Figure 2. The largest effect size was for discharge ($1.30; AUC=.82$) and the smallest was for inpatient victimisation ($0.17; AUC=.55$). Half of the effect sizes were in the moderate to large range, while the remainder would be considered small; however, only the estimated effect sizes for discharge and violent offending were significantly greater than 0, based on inspection of 95% confidence intervals.

Discussion

The current study aimed to collate and synthesize the evidence for the predictive efficacy of a range of tools that have been devised to assist mental health and criminal justice professionals in the assessment of protective factors for important outcomes in adults. Our systematic search strategy revealed that there is currently no empirical evidence for the predictive efficacy of the IORNS, SDRRC, CSSRS or RFL variants for their target, or indeed any, outcomes in adults. Further, there was no evidence for the predictive validity of PF tools in a criminal justice service context, which may highlight a lack of research with this population. Of the four PF scales for which evaluative studies of predictive efficacy have been conducted, the START is the most researched. Our results revealed that protective factor scales performed poorly relative to the risk scales for the violent offending reference category, but not significantly so. This is consistent with evidence that strengths based models do not produce significantly different effects than other service models in people diagnosed with severe mental illness in terms of either level of functioning and quality of life ([Ibrahim, Michail & Callaghan, 2014](#)), or in reduction of incidents of violent and criminal behaviour ([Troquete, van den Brink, Beintema, et al., 2013](#)). However, prediction of violent offending from risk scales in turn performed more poorly than summary judgements. This

suggests that some element of the summary judgement is informed by factors not captured by risk assessment scales and is consistent with previous research which has found superior predictive efficacy for summary judgements for a range of outcomes (e.g., O'Shea et al., 2013). Both the START and SAPROF guidance recommends that the individual's HCR-20 assessment should be considered in conjunction with the protective scales when forming overall judgements and could partly explain this finding.

The current design does not allow us to conclude that protective scales contribute uniquely to the formulation of a more accurate summary judgement although we have previously found evidence for this to be true to a small extent in the case for prediction of violence using the START (O'Shea, et al., 2015). A lack of consensus about how protective factors should be conceptualised is one possible reason for their lack of inclusion in assessment tools. However, the tools included in the current study were largely consistent with our definition C, corresponding to conceptually distinct protective factors, that are balanced against risk factors either at the item level, as is the case with the START, or, at least in the formulation of overall risk judgments, as in the SAPROF. This suggests that tools based on this underpinning theory are preferred clinically.

Predictive efficacy for the range of outcomes beyond violent offending considered by PF schedules, most notably the START and DUNDRUM family measures, was generally significantly poorer than for the reference category; this may in part reflect the relative clarity of outcomes related to offending in comparison with other outcomes. With the notable exception of discharge, Hedge's g values were not significantly different from 0 for sexual offending, moves to higher security, unauthorised leave, inpatient aggression, self-harm, substance abuse, self-neglect and victimisation. This is consistent with previous examinations of the START (O'Shea & Dickens, 2014) and suggests that there are still considerable strides to be made in terms of developing protective factors tools for prediction of non-violent

outcomes. Further, it suggests that it may be overambitious for one tool, such as the START, to aim to assist in risk-assessment for a range of disparate outcomes.

We interpret significant results for prediction of discharge by the DUNDRUM-3 and DUNDRUM-4 as, at least partly, resulting from the fact that the outcome of interest is essentially a clinical decision, which is based on the criteria measured by the tools rather than a patient-determined outcome. This likely contributes to the large effect sizes since, even though outcomes were independent of rating, the clinical criteria for making a decision to discharge a patient are precisely the same criteria captured by the tools.

Effect sizes were found not to be moderated by either female gender, nor by study quality. For gender, this may seem surprising since a recent study has found that the START has superior predictive efficacy in females compared with males (O'Shea & Dickens, 2015a). It is possible that protective factors in the remaining tools are inherently less prone to gender bias; we could not examine this using the current study design and recommend that it is subject to further investigation. It is also perhaps surprising that potential study bias did not moderate results, as it has been suggested that lower quality studies may overestimate effects due to less methodological rigour ([Sterne, Egger, & Smith, 2001](#)). The large confidence intervals observed in the current study and the significant residual heterogeneity suggest that there may be additional factors that moderate effect size; future research, when the literature on PFs is more established, should aim to investigate the potential influence of factors such as authorship bias, ethnicity, diagnosis, study setting, and research vs. routine assessment.

Limitations

There were a number of records that we could not obtain, despite attempted communication with the authors; however, we did review a number of articles that would be considered as grey literature and ultimately included one thesis and one manuscript submitted for publication, reducing the likelihood of publication bias. The overall quality of included

studies was disappointing, with 11 of the 17 studies being rated as unclear or high risk of bias. Further, there were a number of scale-outcome combinations where there was only one available study which limited our ability to conduct more detailed analysis examining the interaction between outcome and scale. However, these criticisms are more a criticism of the state of the existing literature, than the current review itself, and highlight the need for more high quality studies investigating the role of protective factors for adverse outcomes in vulnerable adults.

Implications and Future Directions

The current study demonstrates that the incorporation of assessment of protective factors in determining risk among adults under the care of mental health and criminal justice services is in its infancy. Compared with studies of the predictive value of violence risk assessment tools (e.g., [Singh, Grann, et al., 2011](#)) the evidence base is very limited, particularly for criminal justice populations for whom there was no empirical evidence. There is a need for a clearer conceptualisation of protective factors to facilitate future research regarding their efficacy; such research should determine if their consideration adds to risk-assessment beyond what is achieved by risk factors alone and determine relevance for outcomes other than aggression. A number of authors have claimed that the value of PF assessment may lie elsewhere, for example in its promotion of a holistic view and development of therapeutic relationship. These claims however are hitherto unsubstantiated and deserve research attention in their own right as they cannot be answered by designs aimed at ascertaining predictive efficacy.

References

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

- *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), 197. doi: 10.1186/1471-244X-13-197
- Andresen, R., Oades, L., & Caputi, P. (2003). The experience of recovery from schizophrenia: towards an empirically validated stage model. *Australasian Psychiatry, 37*(5), 586-594. doi: 10.1046/j.1440-1614.2003.01234.x
- Borum, R. (2006). *Manual for the structured assessment of violence risk in youth (SAVRY)*. Odessa, Florida: Psychological Assessment Resources.
- *Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). *The International Journal of Forensic Mental Health, 9*(4), 271-281. doi: 10.1080/14999013.2010.534378
- Child Welfare Information Gateway, Children's Bureau, FRIENDS National Resource Center For Community-Based Child Abuse Prevention, & Families (2014). *Prevention Resource Guide: Making Meaningful Connections*. Washington: Department of Health and Human Services.
- *Chu, C. M., Thomas, S. D., Ogloff, J. R., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *The International Journal of Forensic Mental Health, 10*(4), 337-345. doi: 10.1080/14999013.2011.629715

- *Chu, C. M., Thomas, S. D., Ogloff, J. R., & Daffern, M. (2013). The short- to medium-term predictive accuracy of static and dynamic risk assessment measures in a secure forensic hospital. *Assessment, 20*(2), 230-241. doi: 10.1177/1073191111418298
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.
doi.org/10.1037/0033-2909.112.1.155
- *Davoren, M., Abidin, Z., Naughton, L., Gibbons, O., Nulty, A., Wright, B., & Kennedy, H. G. (2013). Prospective study of factors influencing conditional discharge from a forensic hospital: The DUNDRUM-3 programme completion and DUNDRUM-4 recovery structured professional judgement instruments and risk. *BMC Psychiatry, 13*, 185. doi:10.1186/1471-244X-13-185
- *Davoren, M., O'Dwyer, S., Abidin, Z., Naughton, L., Gibbons, O., Doyle, E., . . . Kennedy, H. G. (2012). Prospective in-patient cohort study of moves between levels of therapeutic security: The DUNDRUM-1 triage security, DUNDRUM-3 programme completion and DUNDRUM-4 recovery scales and the HCR-20. *BMC Psychiatry, 12*, 80. doi: 10.1186/1471-244X-12-80
- de Ruiter, C., & Nicholls, T. L. (2011). Protective factors in forensic mental health: A new frontier. *The International Journal of Forensic Mental Health, 10*(3), 160-170. doi: 10.1080/14999013.2011.600602
- de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2012). *SAPROF. Guidelines for the assessment of protective factors for violence risk (English version 2nd Edition)*. Utrecht: Forum Educatief.
- *de Vries Robbe, 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. doi: 10.1080/14789949.2013.818162

- de Vries Robbe, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2014). 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. doi.org/10.1037/lhb0000089
- *de Vries Robbe, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015). Assessing protective factors for sexually violent offending with the SAPROF. *Sexual Abuse: Journal of Research and Treatment, 27*(1), 51-70. doi: 10.1177/1079063214550168
- de Vries Robbé, M., de Vogel, V., & Stam, J. (2012). Protective factors for violence risk: The value for clinical practice. *Psychology, 3*(12A), 1259-1263. doi: 10.4236/psych.2012.312A187
- *de Vries Robbe, M., de Vogel, V., Wever, E. C., Douglas, K. S., & Nijman, H. L. I. (2014). Risk and protective factors for inpatient aggression. *Manuscript submitted for publication.*
- Del Re, A. C. (2013). compute.es: Compute Effect Sizes. *R package version 0.2-2.*
- *Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24*(3), 685-700. doi: 10.1037/a0026668
- Edelstein, B. A., Heisel, M. J., McKee, D. R., Martin, R. R., Koven, L. P., Duberstein, P. R., & Britton, P. C. (2009). Development and psychometric evaluation of the Reasons for Living-Older Adults Scale: A suicide risk assessment inventory. *Gerontologist, 49*(6), 736-745. doi: 10.1093/geront/gnp052
- *Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., . . . Snowden, R. J. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *The International Journal of Forensic Mental Health, 10*(4), 305-313. doi: 10.1080/14999013.2011.631692

Gutierrez, P. M., Osman, A., Barrios, F. X., Kopper, B. A., Baker, M. T., & Haraburda, C. M.

(2002). Development of the reasons for living inventory for young adults. *Journal of Clinical Psychology*, 58(4), 339-357. doi: 10.1002/jclp.1147

Gutierrez, P. M., Osman, A., Kopper, B. A., & Barrios, F. X. (2000). Why young people do

not kill themselves: The Reasons for Living Inventory for Adolescents. *Journal of Clinical Child Psychology*, 29(2), 177-187. doi: 10.1207/S15374424jccp2902_4

Haney, E. M., O'Neil, M. E., Carson, S., Low, A., Peterson, K., Denneson, L. M., . . .

Kansagara, D. (2012). Suicide risk factors and risk assessment tools: A systematic review. *VA-ESP Project #05-225*.

Hart, S. D. (1998). Psychopathy and risk for violence. In D. J. Cooke, A. E. Forth, & Hare, R.

D. (Eds). *Psychopathy: Theory, research and implications for society* (pp. 355-373): Dordrecht, Netherlands: Kluwer.

Hart, S. D. (2001). Assessing and managing violence risk. In K. S. Douglas, C. D. Webster,

S. D. Hart, D. Eaves & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-26): Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University, and Department of Mental Health Law & Policy, University of South Florida.

Ibrahim, N., Michail, M., & Callaghan, P. (2014). The strengths based approach as a service

delivery model for severe mental illness: A meta-analysis of clinical trials. *BMC Psychiatry*, 14(1), 243. doi:10.1186/s12888-014-0243-6

*Inett, A., Wright, G., Roberts, L., & Sheeran, A. (2014). Predictive validity of the START

with intellectually disabled offenders. *Journal of Forensic Practice*, 16(1), 78-88. doi: 10.1108/JFP-12-2012-0029

- Ivanoff, A., Jang, S. J., Smyth, N. J., & Linehan, M. M. (1994). Fewer reasons for staying alive when you are thinking of killing yourself: The brief reasons for living inventory. *Journal of Psychopathology and Behavioral Assessment, 16*(1), 1-13.
- Jessor, R., Van Den Bos, J., Vanderryn, J., Costa, F. M., & Turbin, M. S. (1995). Protective factors in adolescent problem behavior: Moderator effects and developmental change. *Developmental Psychology, 31*(6), 923.
- Kennedy, H., O'Neill, C., Flynn, G., & Gill, P. (2010). *The Dundrum Toolkit: Dangerousness, UNDERstanding, Recovery and Urgency Manual (The Dundrum Quartet) VI. 0.21. Four Structured Professional Judgment Instruments for Admission Triage, Urgency, Treatment Completion and Recovery Assessments*. Dublin, Ireland: Trinity College Dublin.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12. doi: <http://dx.doi.org/10.3389/fpsyg.2013.00863>
- Laub, J. H., & Lauritsen, J. L. (1994). The precursors of criminal offending across the life course. *Federal Probation, 58*(3), 51-57.
- Linehan, M. M., Goodstein, J. L., Nielsen, S. L., & Chiles, J. A. (1983). Reasons for staying alive when you are thinking of killing yourself: The Reasons for Living Inventory. *Journal of Consulting and Clinical Psychology, 51*(2), 276-286. doi: <http://dx.doi.org/10.1037/0022-006X.51.2.276>
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370. doi: <http://dx.doi.org/10.1037/h0054346>
- Miccio-Fonseca, L. C. (2013). MEGA ♪: A New Paradigm in Risk Assessment Tools for Sexually Abusive Youth. *Journal of Family Violence, 28*(6), 623-634. doi: [10.1007/s10896-013-9527-8](http://dx.doi.org/10.1007/s10896-013-9527-8)

- Miller, H. A. (2006). *Manual of the Inventory of Offender Risk, Needs, and Strengths (IORN)*. Odessa, FL: Psychological Assessment Resources.
- 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), e1000097. doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)
- Monahan, J. (1992). Mental disorder and violent behavior: Perceptions and evidence. *American Psychologist*, 47(4), 511. doi: <http://dx.doi.org/10.1037/0003-066X.47.4.511>
- *Morris, D. (2013). The predictive validity of the Short Term Assessment of Risk and Treatability in an inpatient female forensic population. *Unpublished Masters Thesis, School of Psychology, University of Birmingham, Birmingham, UK.*
- National Institute for Mental Health in England. (2004). Mental Health Policy Implementation Guide. Retrieved 1st March 2012 from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4087136.
- Nicholls, T. L., Gagnon, N., Crocker, A. G., Brink, J., Desmarais, S. L., & Webster, C. (2007). *START Outcomes Scale (SOS)*. Vancouver: BC Mental Health & Addiction Services.
- *Nonstad, K., Nasset, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *The International Journal of Forensic Mental Health*, 9(4), 294-299. doi: [10.1080/14999013.2010.534958](https://doi.org/10.1080/14999013.2010.534958)
- O'Dwyer, S., Davoren, M., Abidin, Z., Doyle, E., McDonnell, K., & Kennedy, H. G. (2011). The DUNDRUM Quartet: Validation of structured professional judgement

instruments DUNDRUM-3 assessment of programme completion and DUNDRUM-4 assessment of recovery in forensic mental health services. *BMC Research Notes*, 4, 229-229. doi: 10.1186/1756-0500-4-229

O'Shea, L. E., & Dickens, G. L. (2014). Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment*, 26(3), 990-1002. doi: 10.1037/a0036794

O'Shea, L. E., & Dickens, G. L. (2015a). Predictive validity of the Short-Term Assessment of Risk and Treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health*, 14(2), 132-146. doi:10.1080/14999013.2015.1033112

*O'Shea, L. E., & Dickens, G. L. (2015b). Predictive validity of the START for unauthorized leave and substance abuse in a secure mental health setting. *International Journal of Nursing Studies*, 52(5), 970-979. Doi: <http://dx.doi.org/10.1016/j.ijnurstu.2015.02.007>

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M., & Dickens, G. L. (2013). Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. *Aggression and Violent Behavior*, 18, 255-270. doi: 10.1016/j.avb.2012.11.016

*O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2015). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes. *Assessment*. doi: 10.1177/1073191115573301

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Predictive validity of the HCR-20 for inpatient self-harm. *Comprehensive Psychiatry*, 55(8), 1937-1949. doi:10.1016/j.comppsy.2014.07.010

- Osman, A., Kopper, B. A., Barrios, F. X., Osman, J. R., Besett, T., & Linehan, M. M. (1996). The brief reasons for living inventory for adolescents (BRFL-A). *Journal of Abnormal Child Psychology*, 24(4), 433-443. doi: 10.1007/bf01441566
- Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., . . . Mann, J. (2011). The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry*, 168(12), 1266-1277. doi: <http://dx.doi.org/10.1176/appi.ajp.2011.10111704>
- Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Towards an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51, 390-395. doi: <http://dx.doi.org/10.1037/0022-006X.51.3.390>
- R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- 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; Law and Human Behavior*, 29(5), 615. doi: 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. doi: 10.1023/A:1005575113507
- 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. doi: [10.1016/j.cpr.2010.11.009](http://dx.doi.org/10.1016/j.cpr.2010.11.009)
- Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: A systematic review of the

validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin*, 37(5), 899-912. doi: 10.1093/schbul/sbr093

[Sterne, J. A. C., Egger, M., & Smith, G. D. \(2001\). Systematic reviews in health care:](#)

[Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, 323, 101-105. doi: http://dx.doi.org/10.1136/bmj.323.7304.101](#)

[Troquete, N. A. C., van Den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P.,](#)

[Schoevers, R. A., & Wiersma, D. \(2013\). Risk assessment and shared care planning in out-patient forensic psychiatry: Cluster randomised controlled trial. *The British Journal of Psychiatry*, 202\(5\), 365-371. doi: 10.1192/bjp.bp.112.113043](#)

[Turner, S., & Fain, T. \(2006\). Validation of the Risk and Resiliency Assessment Tool for](#)

[juveniles in the Los Angeles county probation system. *Federal Probation*, 70\(2\), 49-57.](#)

[Viechtbauer, W. \(2010\). Conducting meta-analyses in R with the metafor package. *Journal of*](#)

[Statistical Software](#), 36(3), 1-48. doi:http://dx.doi.org/10.18637/jss.v036.i03

[Viljoen, J. L., Cruise, K. R., Nicholls, T. L., Desmarais, S. L., & Webster, C. D. \(2012\).](#)

[Taking stock and taking steps: The case for an adolescent version of the Short-Term Assessment of Risk and Treatability. *International Journal of Forensic Mental Health*, 11\(3\), 135-149. doi: 10.1080/14999013.2012.737406](#)

[Viljoen, J. L., McLachlan, K., & Vincent, G. M. \(2010\). Assessing violence risk and](#)

[psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17\(3\), 377-395. doi: 10.1177/1073191109359587](#)

[Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. \(1997\). *HCR-20: Assessing risk for*](#)

[violence \(version 2\). Burnaby, British Columbia: Mental Health, Law and Policy Institute, Simon Fraser University.](#)

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*.

Coquitlam, Canada: British Columbia Mental Health & Addiction Services.

Westefeld, J. S., Cardin, D., & Deaton, W. L. (1992). Development of the College Student Reasons for Living Inventory. *Suicide and Life-Threatening Behavior*, 22(4), 442-452.

*Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior*, 37(6), 377-388. doi: 10.1111/j.1943-278X.1992.tb01038.x

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. doi: 10.1037/a0020473.

Table 1. Estimated mean weighted effect sizes relative to reference category

	Estimate	SE	95% CI	<i>p</i>
Intercept^a	0.88	0.45	[0.00, 1.76]	.050
Sexual Offending	-0.47	0.20	[-0.87, -0.07]	.021
Moves to Higher Security	-0.53	0.43	[-1.38, 0.31]	.216
Discharge	0.42	0.45	[-0.47, 1.31]	.355
Unauthorised Leave	-0.32	0.25	[-0.80, 0.17]	.206
Inpatient Any Aggression	-0.29	0.22	[-0.73, 0.14]	.188
Inpatient Physical Aggression	-0.28	0.23	[-0.74, 0.17]	.225
Inpatient Verbal Aggression	-0.33	0.23	[-0.79, 0.13]	.155
Inpatient Self-harm	-0.68	0.24	[-1.14, -0.22]	.004
Inpatient Substance Abuse	-0.53	0.25	[-1.02, -0.04]	.035
Inpatient Self-neglect	-0.50	0.24	[-0.98, -0.03]	.037
Inpatient Victimization	-0.70	0.24	[-1.18, -0.23]	.004
Protective Scales	-0.03	0.13	[-0.29, 0.24]	.853
Summary Judgments	0.18	0.13	[-0.08, 0.44]	.172
Gender	0.00	0.00	[0.00, 0.01]	.313
Unclear Risk of Bias	-0.07	0.16	[-0.34, 0.24]	.655
High Risk of Bias	0.16	0.40	[-0.63, 0.95]	.695

Note. SE, standard error; 95% CI, 95% confidence interval

^aEstimated Hedge's *g* value for reference category (i.e. Outcome = violent offending, scale type = risk, bias = low risk)

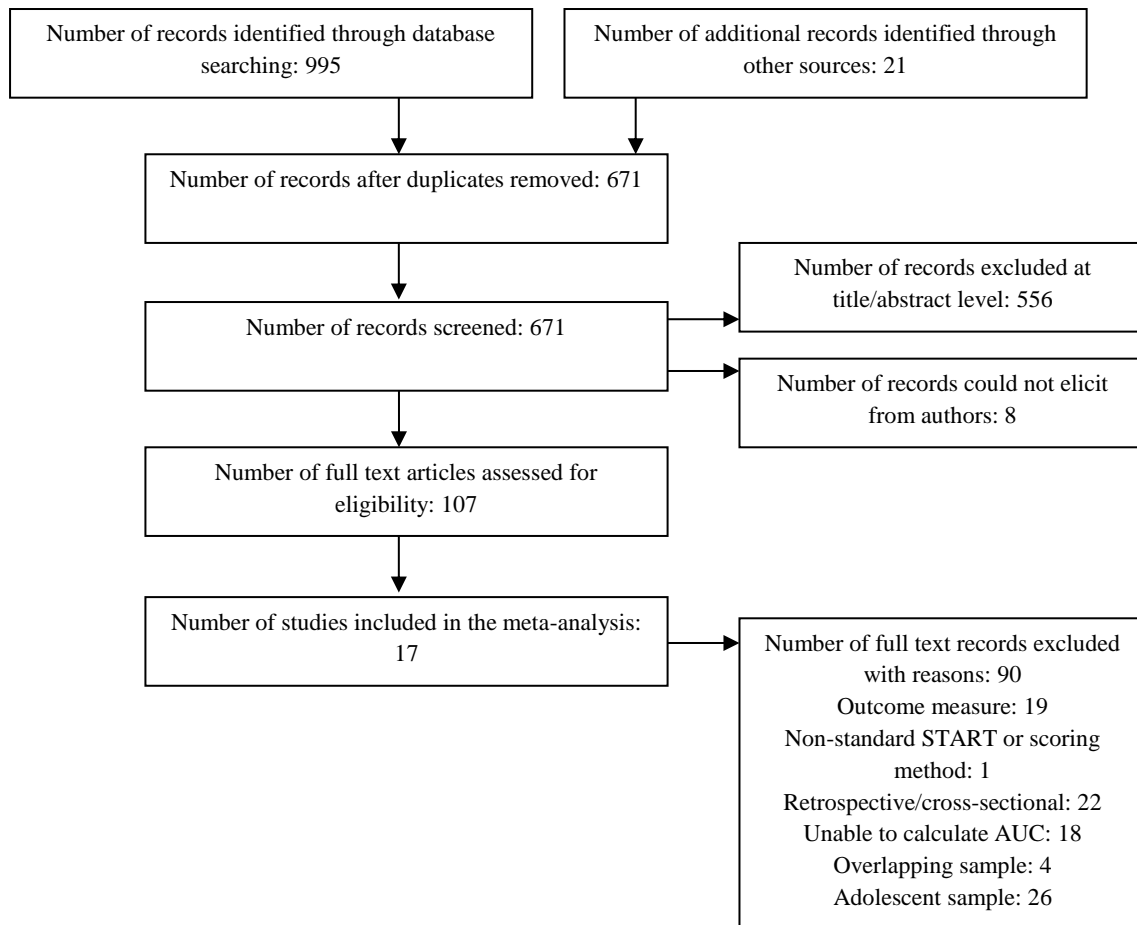


Figure 1. Flow diagram of literature search: Modified from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement flow diagram (Moher et al., 2009). AUC, area under the receiver operating characteristic curve

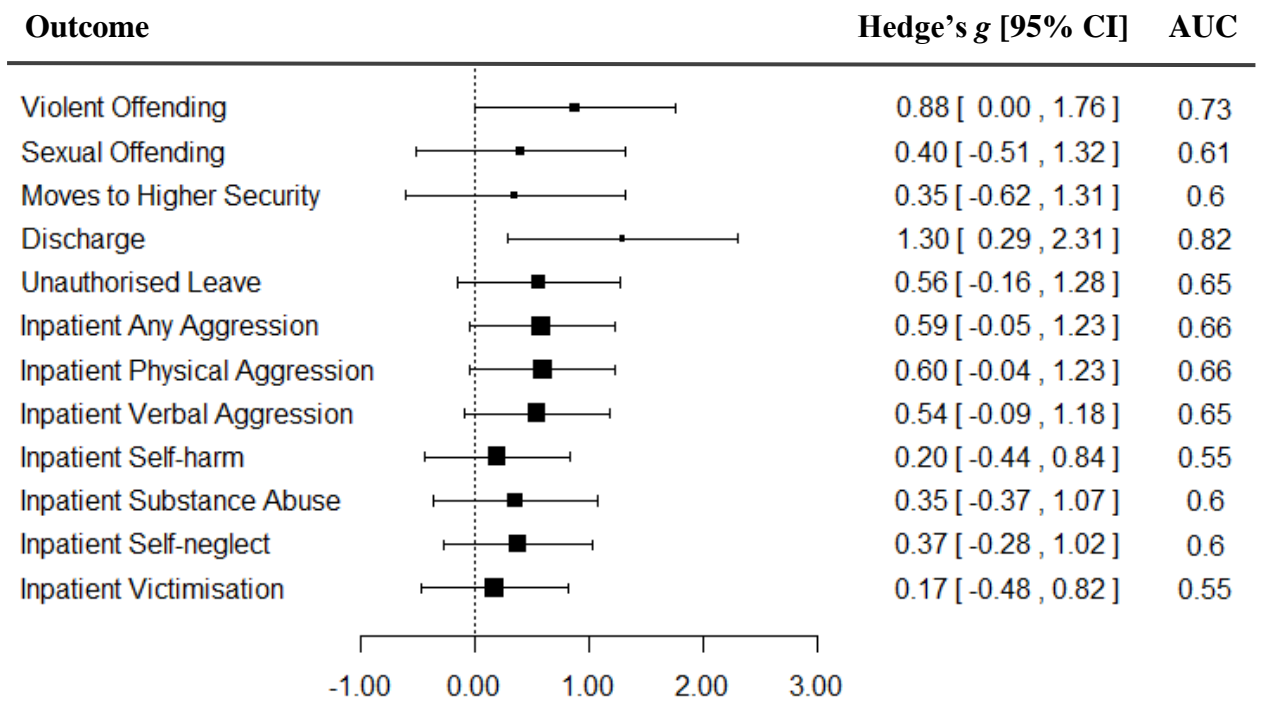


Figure 2. Forest plot of estimated effect sizes by outcome

Supporting Information

Additional supporting information can be found in the online version of this article:

Appendix A: Example search strategy

Appendix B: Characteristics of included studies

Appendix C: Quality of included studies