



Threat/Risk Assessment Literature Review

Task 3: Evaluate Current Threat Assessments and Threat Management –
Test Efficacy

This work is supported by funding by the United States Department of Homeland Security, Science and
Technology Directorate - Public Safety and Violence Prevention Evaluations
70RSAT21G00000002/70RSAT21FR0000084



NCITE NATIONAL COUNTERTERRORISM
INNOVATION, TECHNOLOGY,
AND EDUCATION CENTER

A U.S. DEPARTMENT OF HOMELAND SECURITY CENTER OF EXCELLENCE

**Developed in partnership with NCITE by:
Paul Gill, Zoe Marchment, Amber Seaward**

University of College, London

Threat/Risk Assessment and Management: The Evidence Base

Executive Summary

This report investigates the empirical evidence base for threat/risk assessment and management instruments. We ask: what works, for whom, and in what contexts? To answer these questions, we conducted a systematic sift of the scientific peer-reviewed literature, identified 23 pertinent meta-analyses of threat/risk assessment instruments, and synthesized their key findings. These meta-analyses collectively cover over 2108 empirical studies on over 1.3 million individuals. Our umbrella review leads us to the following 5 conclusions which we elaborate upon in detail in the main body of the report before concluding with some considerations of their implications for the threat assessment and management of terrorism and targeted violence.

Conclusion 1 - Risk assessment instruments can help predictions. The vast majority of meta-analyses place predictive performance in the average-to-good categories. Therefore, there are limitations to what an instrument can do by itself in terms of prediction. One meta-analysis performed well in assessing the onset of violence (whereas the rest focused on recidivism). This points toward the potential of threat assessment instruments in early detection and prevention. The predictive scores also tended to be higher for more common crimes which should lower expectations for how high a threat assessment instrument could predict the likelihood of an act of targeted violence.

Conclusion 2 - Risk assessment performance can be moderated by several design features. A range of meta-analyses investigated whether the predictive performance of risk assessment instruments is impacted (positively or negatively) by a range of potential moderators. First, instruments containing dynamic risk factors perform better. This is important because, by definition, dynamic factors are the most amenable to change, and typically form the targets of interventions and management strategies. Second, assessments will only be as good as the variety of data sources informing them. This has a variety of implications for the need for proportionate ethical data-sharing and multi-agency working. Third, instruments derived from criminological theory and designed for specific harmful outcomes perform better. This implies that instruments designed for one crime type (e.g. sexual violence) are not readily applicable to other crime types, especially lower base rate ones such as targeted violence. Fourth, performance metrics are significantly higher when instrument designers were involved in a validation study. This highlights the need for independent evaluations.

Conclusion 3 – Blending insights from different instruments can improve predictive accuracy. Several meta-analyses demonstrated improved performance when multiple instruments were combined. This is important for practice, but assessors should also have a thorough understanding of the different domains of risk/protective factors incorporated into different instruments and how this may relate to the outcome of interest.

Conclusion 4 – Prediction is not everything. The primary purpose of risk assessment should be prevention. This requires a greater understanding of the individual than a straightforward actuarial score can provide and corroborates the need for structured professional judgement approaches.

Conclusion 5 – Assessments require multivariate explanations of risk. No single factor will explain the emergence of the risk. The key is to think about multivariate explanations of risk, the inter-relationships between risk factors and their cumulative impact upon recidivism likelihood.

Threat/Risk Assessment and Management: The Evidence Base

Risk assessment and management are core components of policing, intelligence, judicial and clinical settings. Risk assessment considers available information and determines an individual's offending risk. Risk management acts to gain or maintain control of the risk posed. Although historically associated with judgements related to the risk of violence recidivism, risk assessment and management principles are increasingly applied to other types of violence (e.g. stalking) as well as newly emergent criminal ventures with high harm potential (e.g. cybercrime).

Its principles (re-badged as *threat* assessment and management) are also increasingly used in the pre-crime space where they address those who are deemed 'at-risk' of crossing the threshold into offending. This is evidenced by PREVENT in the U.K. counterterrorism infrastructure (and its recent various off-shoots in strategies focusing on preventing engagement in cybercrime and serious organized crime). Such 'at risk' individuals may have engaged with extremist materials or settings, and have demonstrated attitudinal affinity with an extremist cause but have yet to mobilize to violence. In the United Kingdom, 4,915 such referrals were made to counterterrorism policing between April 2020 and March 2021. Approximately 14% of these referrals were taken seriously enough for interventions to be implemented to manage their risk.

Plainly speaking, scrutiny of the effectiveness and quality of counterterrorism threat/risk assessment and management come to the forefront when terrorist attacks occur. Unfortunately, this has been too common an occurrence. According to Europol, 2019 witnessed 119 foiled, failed and completed terrorist attacks from a range of ideological influences across 13 EU Member States. Many recent attackers were known to intelligence services, had previously been closed cases or had spent time in prison for terrorist offenses. This naturally raises issues about how these individuals were risk assessed and managed at all points in their trajectory toward offending. Similar concerns have also been evoked about the procedures in place for preventing recent high-profile acts of targeted violence such as mass public murders, school shootings and workplace violence.

As governments face further budget constraints in a world recovering from the COVID pandemic, the onus on threat/risk assessment and management will likely amplify and necessitate better links across practitioner agencies (both policing and non-), academia and industry. This is particularly so for those crime types that are both on the increase and have the potential for high harm, be it physical, psychological, or economic. Consequently, there has been a boom in the availability of threat/risk assessment and management instruments over the past decade focused on forms of targeted violence, particularly terrorism. Some were designed by government entities for in-house work, others are commercially available.

Given both the emergent growing demand and nascent supply of these instruments focused on targeted violence, it is worth taking stock of the more voluminous literature on general violence threat/risk assessment and management. We ask: what works, for whom, and in what contexts? We conduct an umbrella review that synthesizes the results of 23¹ meta-analyses of threat/risk assessment instruments that were identified through a systematic search and sift (see Appendix 1). These meta-analyses collectively cover over 2,108 empirical studies on over 1.3 million

¹ These are denoted with an * in the reference list

individuals. Our umbrella review leads us to the following 5 conclusions which we elaborate upon in turn in the sections that follow:

1. Risk assessment instruments can help predictions
2. Risk assessment performance can be moderated by several design features
3. Blending insights from different instruments can improve predictive accuracy
4. Prediction is not everything
5. Assessments require multivariate explanations of risk

Risk Assessment Instruments Can Help Predictions

A common goal of all risk assessment instruments is to help an assessor grasp what level of risk is currently posed by the individual under assessment. But how accurate are different risk assessments instruments at prediction? Because risk assessment instruments are not designed equally, we should expect a great deal of variability in performance.

A common method of testing the degree to which an instrument can discriminate between those who conduct the outcome of interest (e.g. violent recidivism) from those who do not (e.g. the non-recidivists), is to use Area Under the Curve (AUC). The result determines the probability that a randomly selected, in this case, violent recidivist, was scored highly by the instrument. AUC ROC (Receiver Operating Characteristic) scores of .5 means complete chance prediction (e.g. tossing a coin would be just as good), between .64 and .7 mean average predictive ability, between .7 and .8 means good predictive ability, .8 to .9 means excellent and above .9 outstanding.

Table 1 highlights the aggregated predictive performance (as measured by AUC) of a range of instruments across 18 meta-analyses.

Table 1: Aggregated Predictive Performance of Risk Assessment Instruments

Study	Instrument	Recidivism Type	Sample Specifics	AUC
Babchishin et al. (2012)	Static-99R	Sexual Violence	Sex Offenders	.684
		General Violence		.703
		Any		.718
	RRASOR	Sexual Violence		.661
		General Violence		.614
		Any Recidivism		.582
	STATIC 2002R	Sexual Violence		.686
		General Violence		.708
		Any Recidivism		.732
Campbell et al. (2009)	HCR 20	Institutional Violence	Adult Offenders	.678
		General Violence		.642
	LSI/LSI-R	Institutional Violence		.637
		General Violence		.642
	PCL/PCL-R	Institutional Violence		.585
		General Violence		.637
	PCL: SV	Institutional Violence		.642
	SIR Scale	Institutional Violence		.545
		General Violence		.637
	VRAG	Institutional Violence		.596
		General Violence		.654
Viljoen et al. (2012)	J-SOAP II	Sexual Violence	Sex Offenders	.670
		General Violence		.660
	ERASOR	Sexual Violence		.660
		General Violence		.590

Van Der Put et al. (2019)	All Tools	Domestic Violence	N/A	.647
	CRAT-P			.760
	DVRAG			.715
	IPPA-RAT			.780
	ODARA			.690
	P-TRAIT			.772
	KSID			.555
Van Den Berg (2018)	Dynamic Sex Offender Assessment Instruments	Sexual Violence	Sex Offenders	.692
		General Violence		.619
Pusch & Holtfreder (2018)	LSI	Any	Juvenile Male	.661
		Any	Juvenile Female	.644
		General Violence	Juvenile Male	.669
		General Violence	Juvenile Female	.644
Schwalbe (2007, 2008)	All Tools	Any	Juveniles	.642
	YLS			.641
	PCL			.695
	NCAR			.603
	OCRA			.585
Olver et al. (2009)	YA-LSI	Any	Juveniles	.684
		Non-Violent		.678
		Violent		.666
		Sexual		.608
	YA-PCL	Any		.642
		Non-Violent		.585
		Violent		.642
		Sexual		.540
	SAVRY	Any		.689
		Non-Violent		.719
		Violent		.678
		Sexual		.534
Rossdale et al. (2020)	HCR 20	Violent	Females	.679
		Any		.642
Olver et al. (2014)	LSI	Any	N/A	.672
		Violent		.619
		Non-Violent		.642
		Sexual		.579
O'Shea et al. (2013)	HCR 20	Verbal Aggression	Inpatients	.745
Helmus et al. (2012)	Static 99R	Sexual	N/A	.705
O'Shea & Dickens (2014)	START	Any Aggression	N/A	.714
Fazel et al. (2012)	8 Tools	Violent	N/A	.720
	3 Tools	Sexual		.740
	2 Tools	Criminal Offending		.660
Hanson & Morton-Bourgon (2009)	All Tools	Any	Sex Offenders	.657
		Sexual		.659
		Violent		.641
Singh et al. (2011)	SVR20	Any	N/A	.780
	SORAG			.750
	VRAG			.740
	SAVRY			.710
	HCR			.700
	SARA			.700
	STATIC			.700
	PCL			.660
Tully et al. (2013)	Static 99	Sexual	Sex Offenders	.692
	RRasor			.668
	SORAG			.684
	RM2000/S			.666
	Static 2002			.707
	SVR20			.696
	MnSOST-R			.638
	VRS-SO			.755
	SRA			.737

We can draw a number of conclusions from this table. First, the vast majority of meta-analyses place predictive performance in the average-to-good categories, with none performing excellent or outstanding. Therefore, there are limitations to what an instrument can do by itself in terms of prediction.

Second, looking at the crime types listed, the predictive scores tend to err higher for higher base rate outcomes. For example, general violence occurs more often than sexual violence in the general population and those meta-analyses focused on general violence recidivism prediction fare much better. This has major implications for the adoption of threat and risk assessment instruments for types of violence like mass murder and terrorism which have even lesser base rates. However, alternative measures like attitudinal affinity with extremist measures have higher prevalence rates (Wolfowicz et al., 2020), yet attitudes are typically not considered or tested in the broader risk assessment field where the focus is largely upon overt behavior. It warrants greater consideration therefore of how the assessment of attitudes should be conducted.

Third, the only demonstration of a meta-analysis looking at onset of offending compared to recidivism was that of van der Put et al.'s (2019) systematic review of domestic violence prediction. The results showed that those instruments predicting the *onset* of domestic violence scored significantly higher ($AUC = .744$) than those predicting domestic violence *recurrence* ($AUC = .0643$), which points toward the importance of early detection and prevention and the potential of threat assessment instruments to provide capabilities of spotting this trajectory upstream.

Risk Assessment Performance Can Be Moderated by Several Design Features

A range of meta-analyses investigated whether the predictive performance of risk assessment instruments is impacted (positively or negatively) by a range of potential moderators. The evidence suggests this is the case in terms of (1) the types of risk factor an instrument contains (2) the type of information that informs an assessment (3) the instrument's development and risk specification and (4) authorship bias. We look at each in turn.

Factor 1: Risk Factor Types

There are different ways of categorizing risk factors, and a popular one is to split them into static and dynamic factors. Static risk factors relate to historical factors that are non-changeable aspects of the person being assessed. Dynamic risk factors relate to situational and other types of stressors that are amenable to change and occurred relatively recent to the problem behavior. Campbell et al. (2009) examined the differential performance of instruments that (1) primarily contain static risk factors (2) dynamic risk factors and (3) an equal combination of static and dynamic risk factors. For institutional violence, static instruments had a significantly larger mean effect than primarily dynamic and mixed instruments. The surprising over-performance of static factors here is likely due to the level of control held, and surveillance maintained, within institutional settings on dynamic risk factors (e.g. access to substances, engagement with third parties). Moving to violence in the community, Campbell et al. (2009) found that for violent recidivism, dynamic instruments performed marginally better than instruments that primarily contain static risk factors

and instruments that combine an equal combination of static and dynamic risk factors. Hanson and Morton-Bourgon (2009) also found instruments containing dynamic factors performed better for sexual violence recidivism prediction than those instruments primarily containing static risk factors. Additionally, in a meta-analysis of a single instrument, the HCR-20, the results demonstrated that the dynamic components of the instrument produced larger effect sizes than the static items when predicting inpatient aggression in forensic settings (O'Shea et al., 2013).

The strong performance of instruments with dynamic risk factors is important because, by definition, dynamic factors are the most amenable to change, and typically form the targets of interventions and management strategies.

Factor 2: Data Inputs

Information is crucial to assessments. Depending on the operational context, different types of data sources might inform risk assessments. Some contexts may rely on fully remote assessments with no first-hand contact with the individual being assessed (labelled 'file review only'). Other contexts might rely solely on first-hand interviews with the individual, or solicit self-reports from the individual absent an interview, or conduct a mixture of these practices.

Campbell et al. (2009) tested whether the data sources for assessments impacted predictive performance. File review only and a mixture of file review plus interview produced the largest effect sizes for violent recidivism – almost double that of interview only and self-report only. Similarly, Walters' (2006) meta-analysis demonstrated that combining data sources (in this case file review, interview, and self-report) added incremental validity to predictions compared to relying on a sole source. Schwalbe (2007) demonstrated that 'file review only' and 'direct interview only' performed just as well as each other for juvenile justice risk assessment instruments but did not compare these with a mixture of the two.

The conclusion from this is that assessments will only be as good as the variety of data sources informing them. Some sources may more readily hold insights into particular factors than others and it is important that different sources offset the potential biases generated from alternative sources. It may not always be possible to interview an individual under assessment (e.g. consent was not obtained, or the assessment is part of a covert investigation), but assessors should try to consider offense paralleling behaviors (e.g. behaviors that are functionally similar to the risk outcome), as well as the risk outcome under consideration, to get a more holistic understanding.

Factor 3: Tool Genesis

Campbell et al. (2009) also distinguished risk assessment instruments between those that were derived from criminological theory and/or were created primarily as a risk assessment instrument and those that were neither (e.g. instruments focused on issues such as personality, literacy, and self-esteem) but are used to predict violence outcomes. The former, expectedly, performed significantly better at predicting institutional violence and violent recidivism.

Singh et al.'s (2011) meta-analysis of 9 different risk assessment instruments consistently showed that instruments designed for specific populations produced higher rates of predictive validity than

more generally oriented instruments. More specifically, Hanson and Morton-Bourgon's (2009) meta-analysis similarly showed that risk assessment instruments designed specifically to assess the risk of sexual recidivism performed significantly better than instruments designed for violent recidivism when used to predict sexual recidivism. The inverse was also true, whereby sexual violence risk assessment instruments struggled with predicting violence recidivism compared to those tools designed to assess its likelihood. Van den Berg et al.'s (2018) meta-analysis showed that instruments designed for sexual recidivism risk performed significantly greater for predicting sexual risk compared to violent risk. Olver et al. (2009) conducted a meta-analysis of three risk assessment instruments not designed for sexual recidivism.² For each instrument, performance was significantly greater in predicting general, non-violent and violent recidivism compared to predicting sexual recidivism. In the case of SAVRY, the effect sizes were five times smaller for sexual recidivism, and in the case of PCL it was two to three times smaller. However, van der Put et al.'s (2019) meta-analysis of domestic violence risk assessment demonstrated that those instruments developed specifically for domestic violence performed just as well as other more general violence risk assessment instruments.

The key conclusion therefore is that violence risk assessment instruments perform better than psychometric scales, and that risk assessment instruments typically perform best when built specifically for a certain crime type and deployed against that crime type.

Factor 4: Authorship Bias

Singh et al.'s (2013) meta-analysis of nine risk assessment instruments demonstrated a significant authorship effect on studies validating these instruments. Average effect sizes were significantly higher when instrument designers were involved in a validation study. Singh et al. (2013) propose two alternative explanations. First, the result may be incidental. Designers will be more familiar with the instrument, and those working with the designed may more prodigiously stick to the instrument's guidelines, thus producing a greater effect. Second, designers may be more likely to not submit studies for publication that show poor instrument performance. Either way, the effect size differential is so large that all designers should formally declare their engagement in the study and the potential for authorship bias to have occurred.

Blending Insights from Different Instruments Can Improve Predictive Accuracy

As previously mentioned, there are no shortage of violent extremism threat/risk assessment instruments available (e.g. ERG 22+, IVP, MLG, OVEST, VERA 2R, TRAP 18). The same is true for other forms of violence including general violence (HCR 20 v3, BARR-2002R, VRAG, COVR, VRS, SAPROF, SAVRY), domestic violence (SARA v3, DASH, DVI, ODARA, B-SAFER), sexual violence (SVR 20 v2, RSVP, ERASOR, J-SOAP II, PROFESOR, VASOR 2), and stalking (SRP, SAM). It naturally leads to questions about which instrument is best. If we judge 'best' naively by predictive validity, then there is no clear answer when there are two or more well designed instruments. Because each instrument tends to draw on a similar gene-pool of scientific evidence, often the items they include can be definitionally similar, thus leading to a high-level of correlation between the outcomes generated. For example, O'Shea and Dickens'

² The LSI was developed to assess the risk of general recidivism, the PCL is a diagnostic tool to reliably measure the presence of psychopathic features, and the SAVRY was designed to assess risk of violence.

(2014) systematic review found multiple studies demonstrating positive correlations between the START instrument with SAPROF, PCR:SV, and individual domains of the HCR-20. Viljoen et al.'s (2012) meta-analysis also found correlation between four sex offending risk assessment instruments (J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99). Olver et al.'s (2009) meta-analysis of the performance of LSI, PCL and SAVRY in youth samples similarly shows negligible differences in the performance between these instruments when applied to the same samples.

By synthesizing the results of 20 studies that used three actuarial sex offender risk assessment instruments, Babchishin et al. (2012) investigated whether combining risk scales improved performance. The meta-analysis showed that each scale consistently positively contributed to each other's performance when predicting recidivism. Van den Berg et al. (2018) replicated this finding in their meta-analysis of dynamic sex offending risk assessment instruments. Babchishin et al. (2012) also considered what is the most efficient process to consider multiple scales (e.g. take the highest, take the lowest, or average). The analysis demonstrated it was better to average the risk ratings than to consistently choose the instrument that generated the highest score for an individual case. The combination of multiple instruments therefore managed to negate the risk of underestimating and overestimating the risk of recidivism from case to case.

There are potentially grounds therefore to make use of multiple instruments in a single assessment, but assessors should have a thorough understanding of the different domains of risk/protective factors incorporated into different instruments and how this may relate to the outcome of interest.

Prediction Is Not Everything

There are three approaches to risk assessment: 'unstructured', 'actuarial', and 'structured professional judgement'. Whereas the latter two require risk assessment instruments, the former sees practitioners relying solely on judgement. The identification of characteristics would be based on previous experience, the practitioner's memory of previous empirical findings, and perhaps intuition (Hilton et al., 2006). This method has proven to be less accurate than assessment based on empirical evidence (Monahan, 2008). Meta-analyses have shown poor predictive performance for such unstructured approaches (Hanson & Morton-Bourgon, 2009), with effect sizes between two to five times smaller than alternative approaches (documented below).

If the practitioner uses the actuarial method, the characteristic chosen will be based on empirical evidence from follow-up studies that identify the relationship of the characteristics to the outcome (Hilton et al., 2006). In this manner, the actuarial method differs from the clinical, as it uses statistics to identify the most powerful predictors of offending. However, this can only be done with outcomes that have an extensive and robust scientific basis underpinning them.

Despite the reported success of using actuarial approaches to risk assessment, some note the potential pitfalls of strictly following statistical results (Douglas et al., 1999; Douglas et al., 2005; Dvoskin et al., 2001). Among these is the difficulty of generalizability beyond the samples used in development of the tool, the challenge of applying statistical knowledge to a clinical setting, the propensity of actuarial methods to exclude potentially important risk factors, rigidity of actuarial models and their lack of space for change. Most importantly, actuarial methods fail to address prevention and risk management.

The combination of statistical methods and clinical judgement methods resulted in a new type of risk assessment, namely the structured professional judgement (SPJ) approach (Douglas & Kropp, 2002; Douglas et al., 2003). SPJ allows practitioners to review all available clinical data to identify any potential risk factors, which are found in a structured manual based on empirical evidence (Douglas et al., 2003). Based on these factors, a final structured risk judgement is made, which indicates the risk of offending (Douglas et al., 2003). Unlike actuarial methods, the SPJ does not include fixed guidelines on how to calculate level of risk. Instead, SPJ tools are structured to guide the decision-making process of the practitioners. Tools in this category include a list of risk factors, all of which have been empirically supported, with guidelines on how these risk factors are coded and how to reach a final judgement of different gradations of risk (Douglas et al., 2003).

The major difference between actuarial and SPJ tools is the final risk judgement: actuarial tools only allow for a final decision based on fixed rules and statistical calculations, while SPJ instruments allow the professional to make a final decision considering all risk variables that have been identified (Pedersen et al., 2010). Thus, the expertise of the practitioner is still valued, alongside empirical evidence (Douglas et al., 2003).

As Logan and Lloyd (2019:146-147) outline in regard to SPJ approaches to violence risk assessment and management:

“The substantial part of SPJ-informed risk assessment and management guidance is the identification of risk factors evidenced in the research as being the most relevant to the specific outcome to be prevented...and the review of evidence justifying their importance...Depending on the guidance [document], support is offered on understanding the relevance of risk factors to the individual case, the meaning of violence to the person at risk, the scenarios in which that violence may become manifest, and risk management planning involving a range of strategies”

SPJ can be completed in an abbreviated form whereby an appraiser identifies risk and protective factors and assigns a rating of high, medium, or low risk “based on the pattern of risk factors present in the case.” This is known as ‘SPJ-lite’. ‘Full’ SPJ, on the other hand, additionally requires scenario planning and formulation (Logan and Lloyd 2019:147).

In terms of predictive performance, Singh et al. (2011) found no difference in predictive performance between actuarial and SPJ instruments across many different outcomes. Fazel (2012) found similar. Van der Put et al.’s (2019) meta-analysis of domestic violence prediction showed actuarial tools fared significantly better. However, the actuarial method has the potential to disregard the different dynamics of risk, including the nature, severity, imminence, duration and frequency of future offending (Hart and Logan 2011). So, whilst actuarial methods may aid assessment, they do next to nothing to aid management.

Although prediction is important, the primary purpose of risk assessment should be prevention. This requires a greater understanding of the individual than a straightforward actuarial score can provide. Engaging in violence is a choice. It is carried out with a specific goal, reward, or benefit in mind. The decision might be made quickly, without much forethought, without knowledge of it

being an offense, through carelessness or bad (mis)information – but it is still a decision (Hart & Logan, 2011). SPJ focuses on understanding the factors that led to that decision. Why did they cross the threshold into offending? What was happening for the individual at the time (internally and externally) that facilitated or justified the offending? What were the perceived benefits? What goal did they hope to achieve? Were the negative consequences of offending considered? What were the barriers that could have prevented the offending? With these answers, an assessor can draw inferences about the factors that influenced the onset and continued the problem behavior, as well as consider the factors that might inhibit further problem behavior.

Assessments Require Multivariate Explanations of Risk

Gill et al. (2021) conducted an umbrella review that synthesized the results of 27 meta-analyses and systematic reviews on predictors of violent recidivism. These 27 evidence syntheses examined at least 526,510 participants and 1,118 primary studies. They found 32 different factors that had been subject to meta-analysis. Table 2 summarizes their effect sizes and rank orders them. Effect size ranges were consistently low ($r < .3$), with only one factor experiencing a range that went into the moderate range ($r = .3-.5$). No individual factor had a high effect size ($r > .5$).

At first, the consistently low effect sizes may appear underwhelming. But the opposite is true. The results demonstrate no single factor will explain the emergence of the risk. The key is to think about multivariate explanations of risk, the inter-relationships between risk factors and their cumulative impact upon recidivism likelihood.

Putting gender aside, the strongest effect sizes consistently refer to aspects of personality disorders, prior anti-social/criminal/violent behavior, the interaction of several poor life situations combined (e.g. living conditions, employment, education) and other factors that impact self-regulation (e.g. substance abuse). On the other hand, socio-demographic characteristics like age, race, and others were amongst the smallest effect sizes.

Table 2: Risk Factor Effect Size for Violence Perpetration

Factor	Effect Size from Highest Quality Study	Factor	Effect Size from Highest Quality Study
Anti-Social Personality Disorder	.44	Substance Abuse (any)	.13
Personality Disorder (Any)	.29	Personal/Psychological Problems	.13
Anti-Social Behaviors	.27	Anti-Social Personality Traits	.13
Gender (male)	.24	Psychopathic Personality Disorder	.13
Psychopathic Personality Disorder	.24	Criminal Behavior (previous)	.11
Living Situation (poor)	.23	Childhood Sexual Abuse Victim	.11
Alcohol Use	.22	Family Problems	.10
Violent Behavior (previous)	.21	Relationship Problems	.08
Employment (no) / Poor Education combined	.20	Married (no)	.06
Self-Regulation	.18	Race (minority)	.03
Drug Use	.18	Age (younger)	.01
Leisure (poor)	.17	Socio-Economic Status (low)	Non-Significant
Anti-Social Attitudes	.16	Childhood Physical Abuse Victim	Non-Significant
Age (older)	.16	Mental Disorder (yes)	Non-Significant
Criminal Peers	.14	Employment (no)	Non-Significant
Weapon Use/Threats/ Physical Injuries Caused	.14	Education	Non-Significant

The umbrella review also highlights a number of areas where evidence is lacking. For example, no meta-analyses examined family criminality, self-esteem, social skills, or social ecological factors that consider the impact of community dynamics. Only one low quality design examined the impact of sentence length and did not consider the moderating impact of age. Few studies examined the differential impact of either follow-up periods (e.g. do some risk factors matter more in different follow-up periods) or whether the index offense involved co-offenders or not. Few meta-analyses considered mediating or moderating effects across studies, and this demonstrates potentially major opportunities for future research.

The definition of risk and protective factors was pretty uniform throughout the evidence syntheses. Simply, studies noted the presence of a factor that has a theoretical or pre-existing empirical correlation with offending and/or recidivism. Often syntheses focused on a single factor like psychopathic traits (Asscher et al., 2011). Studies also simply analyzed whether the presence of a ‘risk factor’ was greater in the violent recidivism group compared to the non-violent recidivism group. There was next to nothing which sought to understand the *relevance* of a risk factor to the outcome of violent recidivism. This is for two main reasons. First, *presence* suits idiographic research designs informed by the simple coding that actuarial risk assessment tools produce (e.g. everything coded as yes or no). *Relevance* is potentially much more suited to qualitative nomothetic research designs which are unsuitable for meta-analyses. Second, SPJ manuals do not have an overt focus on prediction in the same way that actuarial tools do. SPJ assesses and manages, whereas actuarial tools predict. Therefore actuarial tools and the data they use lend themselves better to meta-analyses. It was surprising that although the HCR-20 is so widely used, we could not find a single meta-analysis which looks at its more nuanced coding scheme for the presence, relevance and absence of risk factors and their link to violent recidivism.

Conclusion and Implications for Threat/Risk Assessment for Targeted Violence

The empirical evidence base for validated risk and protective factors for terrorism and targeted violence lags behind the evidence base for other forms of violence. This should be unsurprising given these are relatively new literatures, but both have been hampered because of lengthy definitional debates. The contours of what demarcates (a) terrorist engagement from disengagement, (b) radicalization from deradicalization, and (c) re-engagement from recidivism, remain heavily contested. However, this is not a new problem. Gill et al.’s (2021) umbrella review of predictors of violent recidivism demonstrates this is also true for the general violent recidivism literature which means it is surmountable (Table 3).

Table 3: Recidivism Definitions in Different Meta-Analyses

Study	Parole Violations	Rearrest/ Re-offense	New Charges	Reconviction	Reincarceration	Recommitment to Psych Facility	Treatment Failure
Asscher et al. (2011)		X		X			
Bonta & Blais (2014)		X		X		X	
Doherty & Bersani (2018)		X					
Dowden & Brown (2002)	X	X	X	X	X		
Eher et al. (2016)		X					
Eisenberg et al. (2019)	X	X		X			

Geerlings et al. (2020)				X			
Gerhold et al. (2007)		X					
Golenkov et al. (2014)		X					
Hanson & Bussiere (1998)		X					X
Hanson & Morton-Bourgon (2005)	X	X		X	X		
Kreis et al. (2014)		X		X	X		
Maille et al. (2011)		X					
McCann & Lussier (2008)	X	X	X	X			
Mohr, Jensen & Steinhausen (2016)		X					
Olver et al. (2014)		X	X	X			
Olver et al. (2011)			X	X			
Viljoen et al. (2012)		X					
Wibbelink et al. (2017)		X		X	X		
Yu et al. (2012)		X					

Transparent and flexible research practices are the key to pushing science forward in this space. For example, consider the grounds for a rigorous violent extremism re-engagement/recidivism study, and set aside the research difficulties associated with its comparatively lower base rate of index offending, and potentially difficult-to-access official data. A major difficulty in such a study would be overcoming the boundary confusion in terms of where re-engagement starts and ends, and when it transitions into violent extremist recidivism. One route to transparency would be to develop a standardized and widely acceptable definition of violent extremism re-engagement and recidivism prior to any data being collected. Such a definition would articulate inclusion criteria in terms of offending type, and timespan under consideration. From experience, this is likely to be an exhaustive process that is unlikely to achieve the desired consensus that manages to keep all academics and practitioners satisfied. Instead of this exhaustive and likely unsatisfying route, we would suggest a more pragmatic route that prioritizes data collection and flexible definitions. Each small facet of a pre-determined definition has the potential to sway results in any given direction. Instead, a project could list all of the potential facets of a definition and dummy code (e.g. code as yes or no) each one. Rather than checking whether an individual had recidivated against a pre-determined definition, coders would answer separately whether there was a parole violation (this is the closest approximation to re-engagement), a re-arrest, a re-offense, new charges, reconvictions, reincarcerations, and so on. They could code whether the recidivism was a terrorism offense or not, and whether it occurred within 1, 3 or 5 years and so on. The outcomes of such a pragmatic data driven process would show how re-engagement/recidivism rates (and their risk factors) differ according to different definitional permutations.

Next, we return to the five key conclusions we derived and consider the implications for threat assessment and management of terrorism and targeted violence.

1. Risk Assessment Instruments Can Help Predictions

The emphasis here is ‘help.’ Risk assessments are not foolproof because risk is thoroughly dynamic, subject to ebb and flow, information will be missing or lost between agencies, and there are only finite resources. This message is important for the general public, policymakers and politicians alike because the expectation is often that 0% risk is the only acceptable outcome in

terms of terrorism risk assessment and management. Given the oftentimes more sophisticated, predatory and crafty behavior associated with targeted violence versus other forms (which are often impulsive and substance abuse driven), the expectation that an instrument will provide better results for the former than decades of research have demonstrated for the latter is fanciful. On top of that, risk assessment predictions, generally, work best for high volume incidents. It was striking to see meta-analyses refer to the sometimes poor performance of sexual violence recidivism prediction as being due to that crime type's 'low base rate.' Compared to targeted violence, it is not a low base rate. So, this again should temper our expectations about prediction and targeted violence and this provides a large challenge for policy-makers to effectively communicate both the advantages and limits to such approaches.

2. Risk Assessment Performance Can Be Moderated by Several Design Features

Traditionally the focus of general violence risk assessment instruments has been on the differential importance of static and dynamic risk factors. This has happened in parallel in the study of radicalization where focus has shifted away from root cause explanations to more group/network-level and dynamic influences. Given the social nature of terrorist engagement, it is likely that additional factors require consideration for this problem area as well. Issues related to beliefs and attitudes are often overlooked elsewhere but may relate to the facets of an individual we would typically associate with being radicalized. Social context and capability factors are also often overlooked because of the simple nature of most forms of violence but these factors move closer toward the individual's immediate social network, and capability for violence. More work is required to consider how this impacts threat assessments, and whether this type of outcome is readily observed and what types of systems are needed to watch for changes over time.

3. Blending Insights from Different Instruments Can Improve Predictive Accuracy

Existing violent extremism and targeted violence instruments differ in their focus. The risk specifications of some tools vary in focus, such as assessing vulnerability to radicalization (IMV), Jihadist extremist tendencies (EMI), and online radicalization (OVEST). The MLG and TRAP-18 are limited to "group violence" and "lone-actor" terrorism respectively. Others are slightly broader e.g. ERG (risk of re-offending), and VERA (violent extremism). Table 4 notes the percentage of risk factors for each tool, further demonstrating the different types of focus (Gill et al., 2021). It is likely some assessment cases could benefit from multiple instruments being looked at.

Table 4 Distribution of Risk Factor Types by Violent Extremism Instrument

Tool	Static Risk Factors	Dynamic Risk Factors	Beliefs, Attitudes, Behaviors	Social Context & Capability	Protective Factors
IVP	0%	31.25%	25%	43.75%	0%
EMI	0%	0%	100%	0%	0%
OVEST	7.69%	7.69%	30.77%	30.77%	23.07%
ERG22+	13.67%	4.55%	59.09%	27.27%	0%
MLG	30.43%	17.39%	21.74%	30.43%	0%
MLG2	23.81%	19.05%	23.81%	33.33%	0%

VERA	12.5%	8.33%	37.5%	20.83%	20.83%
VERA2	13.79%	0%	51.72%	17.24%	17.24%
TRAP-18	16.66%	27.77%	44.44%	11.11%	0%

However, a bigger challenge likely awaits the development and validation of threat/risk assessment instruments in this space. Repeatedly, it has been demonstrated that radicalization and terrorist engagement should be viewed as a process with various discrete phases along the way. These discrete phases depict multiple types of ‘risk’ related to extremism and terrorism. These phases labelled differently across various theoretical models and include Precht’s (2007) phases of pre-radicalization, conversion and identification, conviction and indoctrination, and action; Borum’s (2003) terrorist mind-set model which includes the stages of grievance, injustice, target attribution and devaluation; Wiktorowicz’s (2004) radicalization model which includes the stages of cognitive opening, religious seeking and frame alignment, socialisation and joining; Moghaddam’s (2005) ‘staircase to terrorism’ model involving psychological interpretations of material conditions, perceived options to fight unfair treatment, displacement of aggression, moral engagement, solidification of categorical thinking and the perceived legitimacy of the terrorist organization and the terrorist act and sidestepping inhibitory mechanisms; Silber and Bhatt’s (2007) radicalization model involving pre-radicalization, self-identification, indoctrination and jihadization. What constitutes a risk factor for vulnerability to radicalization, may not necessarily represent a risk factor for terrorist engagement. Similarly risk factors may play differential roles at different points of the phases. Threat/risk assessment instruments in this space should therefore be mindful of what phase of the process they are considering the likelihood of harmful outcomes occurring.

4. Prediction Is Not Everything

Fortunately for violent extremism threat/risk assessment, prediction is not everything, because the science lags far behind the other forms of violence outlined in the preceding sections. Few instruments here have demonstrated the ability to discriminate between persons of interest and those that are not, few have demonstrated inter-rater reliability and certainly none have been tested to the extent that a meta-analysis is achievable.

Fortunately, SPJ approaches are currently in vogue in this space and widely seen as the best practice amongst those who have responsibility for risk assessment in a variety of clinical, forensic and custodial settings.

Each of the instruments in this space either explicitly or implicitly (e.g. IVP) identifies itself as a SPJ approach. However, the grounds by which they self-identify as SPJ differ. The likes of IVP, EMI, OVEST, and VERA encourage the first two steps of SPJ (e.g. gather information and resources; determine the relevance of factors to individual risk) but fall far short of providing advice on formulation, scenario planning, and management strategies. IVP, for example, provides “useful and relevant information to practitioners about issues they should consider” and encourages decisions being made “in conjunction with practitioners’ professional judgement and expertise, and consultation with line management and agency policy” (Cole et al., 2010: 2). IVP falls short of providing the full-SPJ process of formulation, scenario planning and management strategies. The ERG22+ goes a step further than IVP and uses formulation but not scenario planning.

The TRAP-18 and MLGs are the closest match in terms of the degree to which they mirror SPJ development guidelines. This is particularly so with the MLG whose developers were also heavily involved with Historical Clinical Risk – 20 (HCR-20), the Guidelines for Stalking Assessment and Management (SAM), and other highly rated risk assessment tools for other types of violence. The TRAP-18 manual suggests formulations and management approaches. However, the guidance provided is comparatively very limited, with just 1 page in the guidance document, and appears underdeveloped compared with the MLG.

Unlike actuarial approaches, SPJ approaches do not include fixed guidelines on how to calculate level of risk and instead allow practitioners to review all available evidence to identify any potential risk (or if applicable, protective) factors. SPJ approaches then typically direct a user to code the relevance of an individual indicator. By relevance, we mean the degree to which the item motivates, disinhibits or destabilises the risks posed by the individual being assessed (e.g. its functional relation to offending). SPJ then typically involves a generation of a statement of understanding about the case (e.g. a formulation). It is imperative to synthesise the quality of formulations and whether (a) formulation quality impacts on outcomes, (b) formulations conducted remotely differ to those informed by contact with the individual, (c) training and experience impact the quality of formulations. Additionally, we know very little right now about how risk assessments inform management approaches. Do the presence/volume of risk factors impact upon subsequent intervention intensity? Are interventions emanating from formulations evidence-based, tailored to culture and individual characteristics, and do they draw on building protective factors? Questions such as these are paramount for threat/risk assessment and management, particularly for an understanding of what works, for whom, and in what contexts. Such questions highlight the need for a project such as this current one which aims to both synthesize what is known but also create new knowledge via process and outcome evaluations through engagement with practitioner communities.

5. Assessments Require Multivariate Explanations of Risk

Empirical research demonstrates targeted violence is usually the culmination of a complex mix of personal, political and social drivers that crystalize at the same time to drive the individual down the path of violent action. Whether the violence comes to fruition is usually a combination of the availability and vulnerability of suitable targets that suit the heady mix of personal and political grievances and the individual's capability to engage in an attack from both a psychological and technical capability standpoint. Many individual cases share a mixture of unfortunate personal life circumstances coupled with an intensification of beliefs that later developed into the idea to engage in violence. What differed was how these influences were sequenced. Sometimes personal problems led to a susceptibility to ideological influences. Sometimes long-held ideological influences became intensified after the experience of personal problems. This is why we should be wary of mono-causal master narratives. The development of these behaviors is usually far more labyrinthine and dynamic. Threat/risk assessment instruments need to embrace this complexity rather than shy away from it with pre-determined algorithms.

SPJ-informed assessments aim to make sense of an individual's behavior in the context of their unique circumstances. The 'risk' someone poses is subject to change. It is driven by situations and

circumstances. This is particularly true for adolescent populations, who are subject to major changes in their lives. An instrument should help assessors determine the nature and degree of risk an individual poses for certain kinds of behavior given their current and future circumstances. By understanding the individual's trajectory into a problematic behavior, we might gain an insight into the factors that might exacerbate or stem the issue. Assessments can inform our understanding of future likely scenarios, and the tailored management plans that are required to avert harm. We need to look at each case on its individual merits by weighing and integrating the evidence, and gaining an overall understanding of the individual, their behavior, intentions and motives. We seek to understand whether there is a risk of harm, its nature, scale, severity and imminence. Through this, we seek to prevent potentially harmful outcomes. To do so, we require priorities for intervention that guide risk management. Identifying these priorities is contingent on getting a thorough understanding of the individual.

Assessors should not weigh the presence of each risk factor equally. If an individual has a lot of risk factors but little proclivity or skills for designing a bomb, the risk of a bombing is likely low. On the other hand, there might be cases where only one factor is present but is highly influential in driving harmful risk. The purpose here of an instrument should be to help assessors exercise their best judgement.

In conclusion, the synthesis suggests there are several issues that require immediate focus from practitioners and academics alike in the coming years. First, we need to consider how users and their organizations interact with instruments. The meta-analyses and systematic reviews we synthesized throughout this paper are overtly focused on the instruments, their psychometric properties and predictive capabilities. However behind every risk assessment is an individual practitioner, behind them is an organization, and in front of them both are individuals in need of assessing, many requiring some form of intervention. Practitioners will differ in experience, professional background, and degrees of training. Institutions will differ in terms of caseload, supervisory capacity, and data access. Those assessed as 'high-risk' will likely have greater wrap-around interventions which, in theory, should disrupt the onset of offending (and makes the overt focus on prediction even more curious). Yet none of these issues are synthesized thus far yet are likely far more consequential for real life harm reduction than an instruments' predictive powers, and especially so when low base rate events are the subject of enquiry. Future work on this project will endeavor to do so through a systematic review of the human factors side of risk assessment and management.

Second, the move toward SPJ instruments is a welcome one. They often perform just as well as actuarial instruments on prediction but offer a lot more insight for management. However, they need to be developed specifically for certain types of harmful outcomes. Any radicalization, terrorism, or targeted violence instrument needs to make best use of available science, and work closely with practitioners (e.g. through Delphi studies, interviews and focus groups) regarding the types of information that typically may be available at different points of an assessment and intervention. This is likely to be highly context specific and any learning from other domains and countries needs to reflect on what is possible within the United States. There are many single case study descriptive accounts of targeted violence threat assessment cases, and future work by the project team will develop frameworks to understand (dis)similarities and context-specific factors that operate.

Finally, the vast majority of the research reviewed here focuses on recidivism where individual cases are referred into a system because of some criminal wrongdoing. However, threat assessment works predominantly in the pre-crime space and referrals arrive from a range of other sources in the community which may have different levels of veracity behind them. Unfortunately, little is currently known about the systems of triage that determine the suitability of a referral, and that prioritize competing referrals in situations where information is often incomplete and decisions are time-sensitive.

References³

- Asscher et al., (2011). The relationship between juvenile psychopathic traits, delinquency and (violent) recidivism: A meta-analysis. *The Journal of Child Psychology and Psychiatry*, 52(11): 1134-1143.
- *Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, 19(4), 442-461.
- *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.
- Cole, J., Alison, E., Cole, B., & Alison, L. (2010). Guidance for identifying people vulnerable to recruitment into violent extremism. *Liverpool, UK: University of Liverpool, School of Psychology*.
- Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, 29(5), 617-658.
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11(3), 347.
- Douglas, K. S., Ogloff, J. R., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services*, 54(10), 1372-1379.
- Douglas, K. S., Ogloff, J. R., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of consulting and clinical psychology*, 67(6), 917.
- *Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. *Bmj*, 345.
- *Geraghty, K. A., & Woodhams, J. (2015). The predictive validity of risk assessment tools for female offenders: A systematic review. *Aggression and violent behavior*, 21, 25-38.

³ Those marked with a * were part of the 23 meta analyses identified in the sift

- *Gill, P., Kyprianides, A., and Marchment, Z. (2021) A Rapid Evidence Assessment of Predictors of Recidivism in Violent Offending and Violent Extremism. Report for UK Government
- Gill, P., Zolghadriha, S., and Salman, N. (2021) Testing the Validity, Reliability and Equity of Terrorist Risk Assessment Tools. Report for Public Safety Canada
- *Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological assessment*, 21(1), 1.
- Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. *Forensic case formulation*, 83-106.
- *Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal justice and behavior*, 39(9), 1148-1171.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, 34(3), 400-409.
- Logan, C., & Lloyd, M. (2019). Violent extremism: A comparison of approaches to assessing and managing risk. *Legal and criminological psychology*, 24(1), 141-161.
- Monahan, J. (2008). Structured risk assessment of violence. *Textbook of violence assessment and management*, 17-33.
- *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.
- *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(2), 255-270
- *Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*, 36(4), 329-353.

- *Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, 26(1), 156.
- Pedersen, L., Rasmussen, K., & Elsass, P. (2010). Risk assessment: The value of structured professional judgments. *International Journal of Forensic Mental Health*, 9(2), 74-81.
- *Pusch, N., & Holtfreter, K. (2018). Gender and risk assessment in juvenile offenders: A meta-analysis. *Criminal Justice and Behavior*, 45(1), 56-81.
- *Rossdale, S. V., Tully, R. J., & Egan, V. (2020). The HCR-20 for predicting violence in adult females: A meta-analysis. *Journal of forensic psychology research and practice*, 20(1), 15-52.
- *Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and human behavior*, 31(5), 449.
- *Schwalbe, C. S. (2008). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior*, 35(11), 1367-1381.
- *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.
- *Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? A systematic review and meta-analysis. *Plos one*, 8(9), e72484.
- *Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review*, 33(2), 287-316.
- *van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, 30(2), 179.
- *van Der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment
- *Viljoen, J. L., Mordell, S., & Beneteau, J. L. (2012). Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99. *Law and human behavior*, 36(5), 423.

- *Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes: A meta-analysis. *Criminal Justice and Behavior*, 33(3), 279-304.
- Wibbelink, C. J., Hoeve, M., Stams, G. J. J., & Oort, F. J. (2017). A meta-analysis of the association between mental disorders and juvenile recidivism. *Aggression and violent behavior*, 33, 78-90.
- Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of quantitative criminology*, 36(3), 407-447.
- Yu, R., Geddes, J. R., & Fazel, S. (2012). Personality disorders, violence, and antisocial behavior: a systematic review and meta-regression analysis. *Journal of personality disorders*, 26(5), 775-792.

Appendix 1

We conducted a keyword search of titles and abstracts in ProQuest and PsycINFO for papers published from database inception until 28th November 2020, restricted by English language. For ProQuest, we used the following databases: PRISMA, Public Health Database Information, Psychology and Social Science Premium Collection. For PsycINFO, we used the following databases: Journals, PsycArticles Full Text, PsycBooks, PsychExtra, PsycInfo, PsycTests, and Social Policy and Practice.

We used the same key words for both databases. Key words searched for issues related to recidivism (recidiv* OR re?offen* OR desist* OR re?imprisonment OR re?conviction OR repeat offending OR recur* OR re?arrest OR re?incarceration OR violation OR relapse OR reoccur* OR re?institution* OR setback OR repetit* OR rehabili* OR recall), violence (violen* OR crim* OR offen* OR antisocial OR delinqu* OR aggress* OR abuse OR conflict OR threat OR harm OR assault OR force OR intent* OR serious bodily harm OR victim OR violation), risk or protective factors (risk OR protect* OR predict* OR moderat* OR mediat* OR need OR factor OR cause OR determinant OR indicator OR explain OR predecessor OR associated OR precede OR impact OR antecedent OR promot*), and publication types (meta* OR systematic review OR rapid evidence assessment OR scoping review OR evidence map OR rapid review OR research syntheses* OR Campbell). We also conducted forward and backward citation searches to supplement the search.

The inclusion criteria sought evidence syntheses on predictors/inhibitors of violent recidivism. For the purposes of the sift, we:

- used the HCR-20's definition of violence: "Actual, attempted or threatened infliction of bodily harm on another person. Bodily harm includes both physical and serious psychological harm so long as it substantially interferes with the health or well-being of an individual. Psychological harm includes fear of physical injury and other emotional, mental or cognitive consequences of the act in question."
- understood recidivism as a person's relapse into criminal behavior, often after the person receives sanctions or undergoes intervention for a previous crime.
- defined a risk factor as something that is considered to contribute (directly or indirectly) to the perpetration of violence. This can be specific to the individual, their upbringing, their context, their experiences.
- broadly defined a protective factor as something that inhibits (directly or by buffering risk factors) the perpetration of violence.
- focused on evidence syntheses such as meta-analyses, and systematic reviews.

For the purpose of the search and sift, we were purposefully broad in the definitions of what constitutes recidivism and what a risk/protective factor for it is. The purpose of this broadness was to give us the ability to flesh out the contours of definitional boundaries across relevant evidence syntheses.

Figure 1 depicts the search process. Four researchers sifted through 2,854 studies, and excluded 2,767 for not meeting all of the definitions above, based on an assessment of the title and abstract.

87 full papers were analysed by one member of the team. 21 were excluded because they were some forms of evaluation on a single type of intervention, not an instrument. 33 were excluded because they solely focused on specific risk factors, not ones that were prescriptively belonging to a named risk assessment instrument. 8 were excluded because they did not distinguish violent recidivism from general recidivism, although the abstract suggested otherwise. 2 were excluded because although the abstract was in English, the full paper was not.

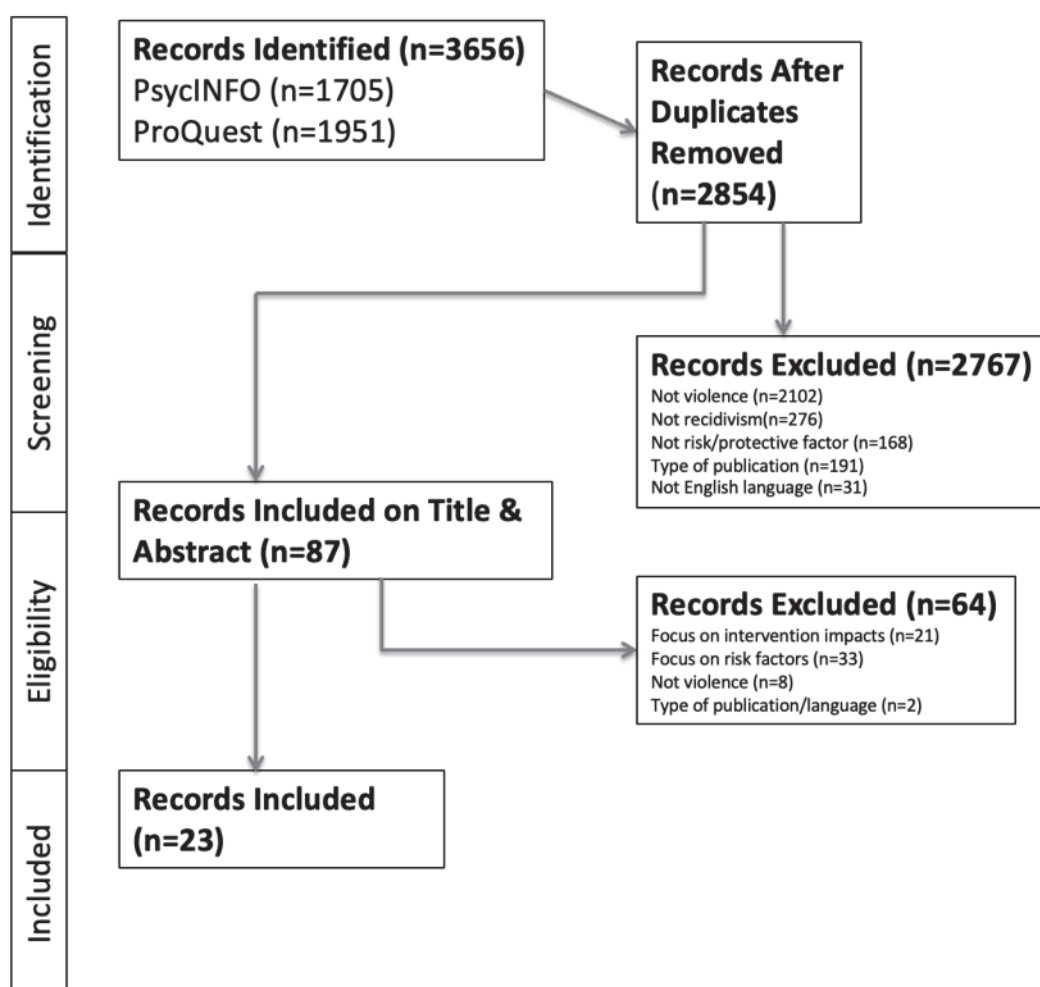


Figure 1: Systematic Search Process & Rates of Attrition

The remaining papers and those identified in the forward citation search were analysed and key sources of information (e.g. study design features, variables, effect sizes) were extracted by a second member of the team. To allow for a standardised comparison, we converted all Cohen's d , Odds Ratio, and Standard Normal Deviate z and Pearson's Correlation r to Area-under-Curve using one of two online effect size converters (<https://www.escale.site>, <https://www.polyu.edu.hk/mm/effectsizefaq/calculator/calculator.html>).