IDENTIFYING GROUPS VULNERABLE TO VIOLENT EXTREMISM AND REDUCING RISKS OF RADICALISATION

RAPID EVIDENCE ASSESSMENT, SEPTEMBER 2019

The authors of this report are:

Jeremy Stickings

Magdulein Abaida

Maia Nikolaishvili

Shivit Bakrania

Annette Fisher

Conflicts of interest
There were no conflicts of interests in the writing of this report.

Contributions
The opinions expressed in this publication are not necessarily those of the Department of International Development. Responsibility for any views expressed herein remains solely with the individual authors.

Citation
This report should be cited as: Stickings, J; Abaida, M; Nikolaishvili, M; Bakrania, S; and Fisher, A. (2019) *Identifying Groups Vulnerable to Violent Extremism and Reducing Risks of Radicalisation.* London: Department for International Development.

A1.0 EXECUTIVE SUMMARY

This Rapid Evidence Assessment (REA) examines the body of evidence concerning the effectiveness of various approaches to the identification of specific groups vulnerable to violent extremism (VE) at a sub-national level – alongside existing intervention tactics aimed at reducing the risk of recruitment and radicalisation.

The UK Department for International Development (DFID) is working to strengthen the evidence base underpinning programme responses to violent extremism. As announced by the International Development Secretary on 12 April 2018, DFID intends to pilot new country-level programming – targeting specific communities and locations vulnerable to violent extremism. This REA is designed to strengthen DFID's understanding of the different methodologies and approaches used to identify and support those most at risk from VE recruitment and radicalisation; their relative strengths and weaknesses – and crucially, their impact and effectiveness (including unintended consequences and lessons learned). The specific research question guiding this REA is: **How effective are different approaches for identifying groups vulnerable to violent extremism, and how effective have interventions that stem from these been at reducing the risk of recruitment and radicalisation?**

Methodological Approach

The research question was split into two research sub-streams focussed on *identification* of populations vulnerable to violent extremism (including five sub-queries), and *interventions* that have been utilised to address violent extremism (including four sub-queries). The nine sub-queries are listed below:

1. **Identification**
   - What methodologies have been used to identify populations at risk of violent extremist recruitment and radicalisation?
   - How effective have these different approaches been at identifying the groups most at risk and the underlying risk determinants?
   - To what extent does the evidence differentiate women, girls, men and boys?
   - What are the key strengths and limitations associated with different identification methodologies?
   - Are there differences between effective methodologies in low- and middle-income countries compared with high income countries?
2. **Intervention**
   - What interventions have been used to address the vulnerable groups identified?
   - What indicators have been used to monitor progress and measure impact of these interventions and how robust are they to measure change?
   - What effect have interventions had (positive and negative)?
   - Are there differences between effective interventions in low- and middle-income countries compared with high income countries?

The REA focuses on English language academic articles published between 2005 and 2018. However, its geographic scope is global. The research team used a structured database search protocol, combined with expert consultation, to identify a total of 2,243 studies potentially relevant to the research question. This was narrowed to a final total of 38 studies based on a set of inclusion and exclusion criteria. The studies were then evaluated using a quality assessment framework, and subsequently coded in an Excel database (the full list is presented in Annex A). Of these 38 studies, 25 were rated 'low quality'; five were classified as being of 'moderate quality'; and eight were rated 'high quality' (based on DFID's March 2014 *How to* Note on "Assessing the Strength of Evidence"). Low quality studies have been excluded from analysis. A full description of the methodology used in this REA can be found in Section 2.

<u>**Key Findings**</u>

In relation to the research question overall, there is a medium sized body of evidence – with a total of 13 studies of high or medium quality that have been assessed in this REA. While all 13 studies identified by the systematic literature searches focus on the identification of populations at risk – three studies focus on *both* identification approaches *and* the effectiveness of interventions employing such approaches in reducing the risk of recruitment or radicalisation. Therefore, the body of evidence that addresses the second part of the research question is classed as limited. As such, it must be treated with caution, and no firm conclusions can be drawn from it.

Only four studies (of the 13) focussed specifically on developing countries – with the majority of research taking place in high-income contexts. The body of evidence is based on a variety of different research designs and methods that have been applied in a range of contexts; as such, this renders the consistency of the evidence somewhat inconclusive. All but one of the studies draw on primary data; with seven using quantitative methods, three utilising qualitative analysis, and one using mixed methodologies. One secondary qualitative study systematically reviewed a range of prior research.

Methodologies used for identification of groups vulnerable to violent extremism

This REA aims to identify indicators or markers for groups at risk of violent extremist recruitment and radicalisation (by contrast with those not, or less, at risk) – whether by virtue of background, personal circumstances, or modes of expression and behaviour. It also aims to determine markers around the degree of radicalisation – and seeks to build a deeper understanding of the process of radicalisation; sometimes touching on more than one of the aforementioned research sub-queries.

Five[1] studies utilise social media, and online sources more broadly, as a means of identifying the vulnerable. This is achieved by analysing internet and social media account such as Twitter and Facebook, often via algorithmic approaches.

There was one systematic review of existing evidence on the process of radicalisation and of interventions aimed at preventing radicalisation and extremism[2]. Another set of five studies use interview-based methods to assess key personal and social factors (including beliefs) – Two of these studied those already involved with extremist activities[3], two focussed on particular identified groups[4], and one used interview-based surveys to compare responses from a number of different groups in Toronto[5]. One further study carried out regression analysis to identify the association between attitudes toward terrorist activities and a set of individual characteristics, drawing on other data sets from surveys of attitudes, values and beliefs[6]. Another primary study drew on the criminal records of individuals identified as having involvement in extremist activity, in an effort to better understand features and pathways of their experience – and its relation to radicalisation[7].

Effectiveness of identification methodologies

The effectiveness of the identification approaches employed in the body of evidence is difficult to assess due to the lack of data on rates of offending, the limited reliability of some of the findings to date, and the experimental

---

[1] Fernandez et al (2018); Davey et al (2018); Frenett and Dow (2015); Ferrara et al (2016); Hung et al (2018).

[2] Christmann (2012).

[3] Botha and Abdile (2014); Zierhoffer (2014).

[4] Moaddel and Karabenick (2008), focussing on young adults (18-25) in Egypt and Saudi Arabia; and Bhui et al (2014) focussing on men and women aged 18–45 of Pakistani or Bangladeshi origin, of Muslim heritage, living in East London and Bradford

[5] Skillicorn et al (2012).

[6] Kiendrebeogo and Ianchovichina (2016).

[7] Basra and Neumann (2016).

nature of many assessment tools. Only three of the 13 studies expressly seek to analyse the effectiveness of their identification approaches. However, a range of key insights have arisen during the REA in relation to personal characteristics, behaviours and backgrounds of groups potentially at risk of radicalisation. Insights have also been gained into the processes and pathways by which individuals may become progressively at greater risk, or more radicalised. For instance, the results of one study[8], which examined a variety of countries[9] with low, middle and high income levels, suggest that, typically, an extremist who supports attacks against civilians is more likely to be young (under 33), unemployed, living in rural areas, and struggling to make ends meet. He or she is also more likely to be relatively uneducated (outside of religion), and more willing to sacrifice their own life for their beliefs.

Two studies,[10] which assessed high and middle income countries[11], argue that psychological aspects of youth – such as fatalism, perceived powerlessness, and feelings of insecurity – are factors that may correlate with religious fundamentalism; and that among Muslim men studied, those showing the most sympathy for violent protest and terrorism were more likely to report depression. Conversely, a greater number of social contacts, and being a recent migrant, were associated with more condemnation of violent extremism. Further, university education was not found to have any significant effect on fundamentalist attitudes. Analysis of the various studies involving Muslim groups also showed that radical attitudes appear absent among those individuals with moral and/or socio-political satisfaction; whereas radicalisation is indeed present among those who are *morally* (rather than economically) dissatisfied – seemingly driving them towards sympathy for terrorist groups. However, there are limits to identifying which individual extremist sympathisers are likely to progress to actually committing terrorist acts themselves; and other findings provide only partial correlations, rather than full explanations of what types of people are "at risk".

Gender differentiation in identification approaches

Of the 13 selected studies specifically focussed on identification, only five offered any gender analysis or gendered differentiation of outcomes[12]. Even these provide only a very superficial treatment of the issue. One study[13] attempts to understand attitudes and behaviours of young male and female Muslims in Egypt and Saudi Arabia, and finds significant differences between genders in both countries. In summary, young Saudi males are reportedly less fundamentalist than Saudi females; whereas the converse is reported in Egypt. Another study suggests that gender and marital status are not found to explain significant individual-level variations in attitudes towards extremism, although this is not viewed as conclusive[14].

Key strengths and weaknesses of different identification approaches

The authors themselves are often aware of limitations in both the reliability of the findings and the wider applicability of some of these methodologies.

---

[8] Kiendrebeogo, Youssouf, and Elena Ianchovichina (2016).

[9] Indonesia, Malaysia, Azerbaijan, Kyrgyzstan, Tajikistan, Algeria, Egypt, Iraq, Lebanon, Qatar, Tunisia, Yemen, Afghanistan, Bangladesh, India, Pakistan, Burkina Faso, Chad, Comoros, Guinea, Mali, Mauritania, Niger, Nigeria, Senegal, Somaliland and Tanzania.

[10] Moaddel and Karabenick (2008); Bhui, et al (2014).

[11] United Kingdom, Egypt, Saudi Arabia.

[12] Bhui et al (2014); Moaddel and Karabenick (2008); Christmann (2012); Davey et al (2018), Kiendrebeogo and Ianchovichina (2016)

[13] Moaddel and Karabenick (2008).

[14] Kiendrebeogo and Ianchovichina (2016).

Some risk assessment tools with strengths deriving from their focus on behaviours rather than demographic data are promising – although these observations of behaviour require a proximity to the perpetrator which is often not possible.

In the case of psychosocial methodologies (which explore the combined influence of psychological factors and the socio-economic context on an individual's physical and mental state), groups were selected by a statistically acceptable method of cluster analysis. Despite the limitations in identifying which extremist sympathisers are most likely to progress to committing terrorist acts, they do present a sequence of events consistent with the stages of radicalisation[15].

Those studies that look at online behaviours and semantic information face further limitations as they rely on data gathered from *public* social media pages. However, such methodologies do permit research at scale and overcome the need for proximity.

There are also issues around the practicality of acting on some of the findings, while yet others are limited by small data sets or a restricted piloting context (thereby pointing to the need for further research).

Differences between low, middle, and high-income countries

The majority of the 13 studies assessed high income countries (HICs)[16] – five exclusively, and seven as part of a global study, or grouped with selected lower or middle income countries (LICs/MICs). Almost all of the latter are Islamic countries.

Only one study is based in a LIC (Somalia), and one focuses on a MIC (Egypt) and a HIC (Saudi Arabia).

A key difference between methodologies in LICs/MICs compared with HICs is that those dealing with LICs/MICs employ more direct communication approaches (e.g. interviews with individuals, or observational assessment of their attitudes and characteristics). The literature from HICs tends to focus primarily on out-takes from social media platforms and electronic monitoring (often undertaken by US and European law enforcement).

Types of intervention methodologies applied

As noted above, there is a limited body of evidence (three studies)[17] addressing *both* the identification of vulnerable groups *and* interventions aiming to reduce risk of recruitment and radicalisation among the groups identified through the employed approaches. The three studies cover high, middle, and low-income countries. The first is global in nature, the second covers the US and the UK – while the third is focussed on the UK, the Netherlands *and* a range of Islamic countries. Two studies employed online outreach to interact with individuals who, through analysis of social media behaviours, the researchers had identified as "at-risk". The third study is a systematic review (exploring a range of in-person interventions).

Given the small number of sources, and the variety of different designs and methods applied in a range of contexts, there is no immediate, clear determination on consistency of evidence (as each study addresses questions not directly tackled by the others). Overall, the evidence base in relation to interventions is currently limited, with further research required. The authors of the three studies share the view that although PVE literature is developing rapidly, serious impact evaluations remain very limited. They further note that most PVE

---

[15] Stage 1: Pre-Radicalisation; Stage 2: Self-Identification; Stage 3: Indoctrination; Stage 4: Jihadization; Silber M, Bhatt A (2007) Radicalisation in the West: the Homegrown Threat. New York New York City Police Department. Cited in Bhui et al (2014).

[16] Four of the 13 documents were global in context, whilst the UK was the subject of four of the other studies, the USA and Netherlands in two, while Denmark, France, Germany, Egypt, Saudi Arabia, Canada, and Somalia, and a group of Islamic countries featured in one study each.

[17] Christmann (2012), Davey et al (2018) and Frenett and Dow (2015).

studies are theory-driven, and policy recommendations are frequently based on theoretical frameworks or conceptual models, rather than empirical evidence.

Though consisting of only three studies, the body of evidence reports and analyses several different types of intervention methods to prevent VE, as the systematic review[18] considers a small but quite varied range of different programmes.

One intervention type concerns pedagogical approaches and styles of teaching, or other engagement, to encourage tolerance of "value pluralism", critical thinking, and open-mindedness – enabling people to see and acknowledge different points of view. A similar approach involves capacity building or empowering at-risk groups to challenge narratives and assumptions that might increase risk of radicalisation, including through religious education to enable individuals to counter radicalising ideology.

The two remaining studies undertook online interventions that utilise Direct Messaging (DMs) to individuals openly expressing extremist sentiments on social media platforms and sought to dissuade them from following that path.

Types of indicators employed in monitoring and evaluating interventions with vulnerable groups

The three studies use multiple indicators to monitor results of interventions, but these were often not rigorously applied. Various indicators tend to be used depending on the target group and types of intervention. For instance, programmes using community outreach apply indicators based on instances of recognition or engagement with opposing viewpoints or moral dilemmas. Studies using online conversations between vulnerable individuals and staff apply indicators such as response rates and length of interactions as markers of positive impacts, alongside qualitative analysis of the conversations to 'detect instances which suggested that the conversation had possibly generated positive impact by presenting a candidate with an alternative point of view'[19]. Much of this focusses on the internalisation of counter arguments, with the assumption that it indicates empathy and the diversification of perspectives. An expressed desire to continue engagements with the programme's implementers is also seen as an indicator of positive impact.

Overall, both qualitative and quantitative indicators are shown to have relevance in assessing whether interventions have had positive and/or negative impacts.

Effectiveness of interventions

Within the three studies there is evidence of interventions having a positive impact in relation to improved attitudes, and an increased willingness to engage. However, as is the case when assessing the effectiveness of identification approaches, given the lack of data on offending (and the challenges around counterfactuals), clear conclusions on whether the interventions have indeed reduced violent extremist actions are not possible.

The two studies that utilise online outreach approaches suggest positive or neutral effects, while the systematic review indicates positive impacts from the interventions on which it reports. However, none of the studies use regression techniques or other methods that would allow for causal inference.

The three studies reviewed list various possible factors to determine the effectiveness of interventions, including social-psychological models; but largely fail to specify the interactions between them. The systematic review noted that capacity building – and interventions empowering young people and challenging ideologies – were most successful when delivered via outreach work that considered the social and institutional context of the relevant communities.

---

[18] Christmann (2012).

[19] Davey et al (2018).

As the majority of studies identified research that used a combination of various intervention methods, it is difficult to disaggregate the effects of these interventions. It does, however, appear that the methods listed above do in fact yield positive results (e.g. challenging extremist ideologies and offering alternatives) more often when applied in combination with one another.

High income, low- and middle-income country differentials in the evidence on interventions

While the number of studies is limited, the body of evidence does nevertheless draw on evidence from a range of geographical contexts. In the LICs/MICs, there is emphasis on employment and livelihoods (often incorporating elements aimed at building resilience, self-esteem and tolerance in various ways). In the HICs (where the sources are strongly orientated towards immigrant communities), the balance tends to be reversed, with a heavier emphasis on social media – and on tolerance, value pluralism, and "debiasing" through educational initiatives. Resilience, self-esteem and empathy are frequently addressed as objectives in and of themselves, rather than add-ons to economic initiatives.

Conclusions

An evidence base is certainly emerging around the effectiveness of various approaches to identifying groups vulnerable to violent extremism – and the outcomes of intervention methods aimed at reducing the risk of recruitment and radicalisation. However, the evidence is not yet sufficient to yield conclusive, actionable insights. The body of evidence that explores identification approaches is **medium** (13 studies); but when looking at the effectiveness of interventions based on these approaches, the evidence base is reduced to **limited** (consisting of only three studies). The body of evidence has a range of limitations related to specificity, restricted data sets, transferability of results – and a lack of focus on the question of effectiveness, and gender or geographic differentiation. As noted by one of the authors[20], there is as yet no general causal model or theory of the structural causes of extremism, and there is debate as to whether such a model is even achievable. The use of indicators to monitor progress and measure impact is also problematic. Most studies have multiple indicators, which are often difficult to disaggregate, and not very rigorously applied (with many also mixing a variety of approaches). Links of causality between these factors and radicalisation, or extremism, are very frequently unclear. Furthermore, the identification of particular characteristics being typical of extremists does not mean that all persons with those characteristics can be classified in such a way.

In spite of the limitations, the evidence is beginning to provide some useful insights around who is at risk, and how best to engage with them. It also provides a range of examples that demonstrate how questions around the identification of and intervention with at-risk groups can be tackled methodologically – whether through analysis of data (on and offline), or through discussion and personal engagement. When examined collectively, the body of evidence also begins to paint a picture of radicalisation as a process within an individual, affected by various contextual factors, interactions and influences; that is, a dynamic (rather than a fixed) feature of the individual. This is an important refocus (or widening of focus) where efforts look to prevent and reduce incipient or future radicalisation, though naturally it comes with a range of ethical and legal sensitivities that must be addressed carefully.

To progress our understanding of these complex issues, we should perhaps move away from the simple "what works?" question, and towards a more holistic, useful set of questions: "What works? For whom? In what circumstances? And how?". A particular issue remaining to be tackled in building a robust evidence base around the effectiveness of the various approaches to identification and risk reduction is the lack of focus on *effectiveness* in the existing research studies. A robust evidence base of studies that rigorously apply their hypotheses and indicators is required here, despite the methodological challenges that this entails. In addition, given the gaps in the existing evidence base around gender and geographical differences, concerted efforts will

---

[20] Christmann (2012).

need be made to better understand how these basic demographic attributes impact the likelihood of radicalisation – and the types of intervention to reduce risks that will be effective for each group.

Contents

CF	Collaborative Filtering
CSJ	Conflict, Security and Justice Group
CT	Counter-Terrorism
CVE	Countering Violent Extremism
DCLG	Department for Communities and Local Government
DFID	Department for International Development
EU CORDIS	EU-funded Community Research and Development Information Service
FCAS	Fragile and Conflict-Affected States
GAPS	Gender Action for Peace and Security
HIC	High-Income Country
HMG	Her Majesty's Government
INSiGHT	Investigative Search for Graph-Trajectories
ISIS	Islamic State
ISS Africa	Institute for Security Studies
LIC	Low-Income Country
MIC	Middle-Income Country
NGO	Non-Governmental Organisation
NLP	Natural Language Processing
ODA	Official Development Assistance
PVE	Preventing Violent Extremism
R4D	Research for Development Outputs
RCT	Randomised Controlled Trial
REA	Rapid Evidence Assessment
RED	Research and Evidence Division
RRFS	Revised Religious Fundamentalism Scale
RUSI	Royal United Services Institute
TORs	Terms of Reference
UNDP	United Nations Development Programme
UNSCR	United Nations Security Council Resolution
USAID	United States Agency for International Development
VE	Violent Extremism
VEO	Violent Extremist Offender

| VERA | Violent Extremist Risk Assessment |
| WPS | Women, Peace and Security Agenda |

1.0   INTRODUCTION

Over recent years, the Department for International Development (DFID) has substantially increased its emphasis on the generation and use of rigorous evidence to inform programme design and policy-making. A pre-requisite of evidence-informed policy and programming is access to the best research evidence (i.e. research evidence which can be relied upon to be rigorous, substantive and objective). It is expected that, by basing decisions on the most up-to-date information of 'what we know', the success, impact and value for money of policies and programmes will be enhanced.

In collaboration with the Conflict, Security and Justice Group (CSJ), the Research and Evidence Division (RED) is working to strengthen the evidence base underpinning programming responses to violent extremism (VE). One of the key challenges identified by practitioners is how best to allocate limited resources to support those at risk from violent extremism recruitment, and radicalisation.

Preventing violent extremism (PVE) programming ranges from broad-based interventions aimed at factors believed to indicate increased vulnerability to radicalisation and recruitment (e.g. addressing systemic issues of exclusion or societal grievances) – to highly localised, community-based interventions working with specific sections of the population. The evidence base on wider drivers of violent extremism is relatively well-advanced. However, identifying specific populations at risk is challenging – particularly when it comes to isolating how this differs between high, middle, and low-income countries. Existing literature suggests that violent extremism is diverse, has multiple pathways to entry, and cannot be predicted by one variable alone. There is no such thing as a homogenous extremist profile; those involved in violent extremism vary in terms of education, family background, and income. While there are some factors for which there is empirical evidence of a link between these profiles and violent extremism – how, where and when they matter depends on the context.

As announced by the International Development Secretary on 12 April 2018, DFID intends to pilot new country-level programming, targeted at specific communities and locations vulnerable to violent extremism. Communities in this context could be geographical, social or institutional. This Rapid Evidence Assessment (REA) is designed to strengthen DFID's understanding of the different methodologies and approaches used to identify and support those most at risk from VE recruitment and radicalisation; their relative strengths and weaknesses – and their impact and effectiveness (including unintended consequences and lessons learned). This body of insight will subsequently influence and inform the design and implementation of the pilot programme.

The overarching research question is: **How effective are different approaches for identifying groups vulnerable to violent extremism and how effective have interventions that stem from this been at reducing the risk of recruitment and radicalisation?**

REPORT STRUCTURE

Following this introduction, section two provides an overview of the assessment methodology we employed (including details on inclusion/exclusion criteria, types of studies and research design, quality appraisal, and synthesis of the evidence base). An assessment of the evidence is then presented in section three, with some headline conclusions offered in section four of the report.

2.0   METHODOLOGY

This section provides an overview of the methodology employed for this review. REAs synthesise the existing evidence on a particular topic, issue or question, using transparent methods to give the best possible, generalisable statements about what is known (Waddington et al., 2012:360). They are particularly useful for understanding the 'state of the art' in emerging areas and are used to account for both quantitative and qualitative evidence. With an understanding of the research question and the benefits of this approach, a modified version of this methodology that allowed literature to be solicited from both online sources and experts was selected, following consultation with DFID through the agreement of a research protocol. The methodology sought to respond to the research question set out in Section 1 related to available evidence on how specific groups vulnerable to VE can be identified and risks of radicalisation reduced, as outlined by DFID in the project's terms of reference (TORs).

This REA draws from an initial scoping exercise, which entailed some simple searches and an analysis of a sample of literature focusing on PVE interventions, and which was discussed with DFID prior to commencement of the substantive REA.

VE DEFINITION

This REA draws from the definition of VE offered in DFID's VE Operational Guide (2018). It states that: "Violent extremism refers to advocating, engaging in, preparing, or otherwise supporting ideologically motivated or justified violence to further social, economic, religious and political objectives. There are disagreements over the definition internationally". The same source emphasises that although extremism is defined in the HMG Counter Extremism Strategy as the "vocal or active opposition to fundamental values, including democracy, the rule of law, individual liberty and mutual respect and tolerance of different faiths and beliefs…the domestic definition of extremism is not easily translatable to foreign contexts. DFID has focussed on *violent* extremism".

2.1   SEARCH PROTOCOL

The research team used a structured database search protocol, combined with expert consultation, to select relevant texts (summarised at **Annex A**). The search process is transparent and uses a precise specification on the inclusion/exclusion criteria, keywords, search strings and data sources.

INCLUSION AND EXCLUSION CRITERIA

The criteria for considering studies to be included in the REA are stated in the table on the following page. This drew from the initial scoping and built on the thematic scope suggested by the framework above.

	Inclusion criteria	Exclusion criteria
Geographical location	Low-, middle-, and high- income countries (based on World Bank classification) OR regional[21]	None

---

[21] The geographic scope of this REA is global to enable comparisons between the effectiveness of approaches in high-income countries and low-/middle-income countries (HICS/MICS/LICS), as requested in the terms of reference. The initial scoping suggested that the majority of evidence available is focussed on HICs, with very little on MICs and LICs. Therefore, no geographical exclusions were applied.

	Inclusion criteria	Exclusion criteria
Language	English only[22]	Languages other than English
Publication date	2005 onwards[23]	Before 2005
Publication type	Electronically available peer reviewed journal articles, other academic research and grey literature (including studies published by non-academic institutions such as NGOs)	Electronically available books and book chapters, non-electronically available studies, operational and policy documents, toolkits and process reviews
Study aim	Focussed on the impacts of: (1) activities to identify populations at risk of violent extremism; and (2) interventions to prevent violent extremism amongst vulnerable populations	Studies not focussed on the impacts of interventions of: (1) activities to identify populations at risk of violent extremism; and (2) interventions to prevent violent extremism amongst vulnerable populations
Thematic coverage	Please refer to this section for full thematic coverage	Counter-terrorism, initiatives addressing radicalisation in prisons and the release and reintegration of prisoners into the community. Studies on the drivers of VE and radicalisation where they are not assessing the impacts or effectiveness of approaches or interventions
Research design	Primary empirical research using (qualitative, quantitative and mixed methods) and secondary synthesis – with an emphasis on systematic reviews of effects. See below for a further explanation of the types of research design considered	Theoretical and conceptual studies, reviews that are not systematic in nature (traditional literature reviews), other studies lacking an explanation of their methodology

2.2   TYPES OF STUDY AND RESEARCH DESIGN CONSIDERED

The research question agreed was: **'How effective are different approaches for identifying groups vulnerable to violent extremism and how effective have interventions that stem from this been at reducing the risk of recruitment and radicalisation?'**

---

[22] In order to limit the scope of the REA, and to ensure manageability, only English language studies have been included. The initial scoping showed that relevant evaluative studies have been either published or translated into English, hence including documents in other languages (e.g. French, Spanish or Arabic) added limited value.

[23] Only studies published between 2005-2018 were included. This incorporated the inclusion of documents pre-Arab Spring and reflects UK interventions in fragile and conflict affected states (FCAS) affected by extremism.

This research question was split into nine research sub-queries focussed on the identification (five sub-queries) of populations vulnerable to violent extremism, and the interventions (four sub-queries) that have been used to address violent extremism.

Research sub-queries

> **3. Identification**
> - What methodologies have been used to identify populations at risk of violent extremist recruitment and radicalisation?
> - How effective have these different approaches been at identifying the groups most at risk and risk determinants?
> - To what extent does the evidence differentiate women, girls, men and boys?
> - What are the key strengths and limitations associated with different identification methodologies?
> - Are there differences between effective methodologies in low- and middle-income countries compared with high income countries?
>
> **2. Intervention**
> - What interventions have been used to address the vulnerable groups identified?
> - What indicators have been used to monitor progress and measure impact of these interventions and how robust are they to measure change?
> - What effect have interventions had (positive and negative)?
> - Are there differences between effective interventions in low- and middle-income countries compared with high-income countries?

The research team used a structured database search protocol, combined with expert consultation, to identify a total of 2,243 studies potentially relevant to the research question. This was narrowed to a final total of 38 studies based on set of inclusion and exclusion criteria. The studies were then evaluated using a quality assessment framework and coded. Of these 38 studies, 25 were rated low quality and 13 classified as moderate or high-quality documents (eight documents classified as high quality and five classified as moderate quality) based on DFID's March 2014 *How To Note* on "Assessing the Strength of Evidence". Low quality studies have been excluded from analysis. The database of studies included in the REA is provided at **Annex A**.

The types of research design included were:

> **Quantitative methods:**
> - Studies using an experimental design – i.e. randomised controlled trials (RCTs)
> - Quasi-experiments that have taken steps to establish a reasonably credible counterfactual
> - Statistical analysis (including regression analysis and hypothesis testing)
> - Studies based on the use of descriptive statistics
>
> **Qualitative methods:**
> - Observational studies (e.g. cohort and/or longitudinal studies, case control designs and cross-sectional designs, and large n-studies)
> - Case study design (either single country or comparative case studies)
>
> **Mixed methods**
> - Studies that combine qualitative and quantitative approaches to design and analysis
>
> **Robust evidence syntheses:**
> - Any systematic review which collates all empirical evidence that fits the pre-specified inclusion criteria. This may use either quantitative meta-analysis or qualitative thematic analysis approaches to synthesis, or a combination of both, or rapid evidence assessments

The following table outlines the abbreviations used to denote the research methods applied:

Research Type	Quantitative (QUANT)	Qualitative (QUAL)	Mixed (MM)
Primary (P)	Experimental (EXP) Quasi-experimental (QEX) Statistical analysis (SA)	Observational (OB) Case Study Design (CS)	Mixed Methods (MM)
Secondary (S)		Systematic Review (SR)	

2.3   PROCEDURES

The REA's search strategy was based on systematic review principles, but with some restrictions and limitations to reflect the rapid nature of this assignment and to ensure manageability. The first phase involved a search of the available and agreed upon online academic and institutional databases for sources published in peer-reviewed journals and 'grey' literature[24].

Most of the queried bibliographic databases allowed for complex combinations of search terms, with operators, search limiters and search syntax. Accordingly, search strings were developed based on key thematic terms in the literature assessed during the initial scoping and included some search specifiers to account for the REA's inclusion criteria. Separate search string combinations were used for both elements of the primary research question and their sub questions.

Long-form searches strings

Column 1: base search terms	Column 2: Thematic specifiers	Column 3: Methodological specifiers
extremis? OR radicali?ation	**Identification:** identif* OR screen* OR trait* OR profil* OR characteristic* OR "risk assessment" OR scale OR vulnerab* OR criteria **Interventions:** program* OR intervention OR prevent* OR counter* OR "counter-radicalisation" OR incentiv* OR measurement OR indicator **Intervention types:** education OR resilience OR belonging OR certainty OR identity OR "civil society" OR empowerment OR "counter-narrative" OR messag* OR counsel* OR "psycho-social" OR "critical thinking" OR tolerance OR "non-violent" OR marginalisation OR "social capital"	impact OR effectiveness OR evaluation OR assessment OR analysis OR "systematic review" OR "rapid evidence assessment" OR synthesis OR trial OR survey

---

[24] Here, grey literature refers to non-peer reviewed reports and studies by think tanks, non-governmental organisations and international institutions.

Short-form search strings

Column 1: Methodological specifiers	Column 2: Research question specifiers	Column 3: Thematic specifiers
Evaluation OR impact OR effectiveness	Identify OR "risk assessment" OR vulnerability	extremism OR radicalisation or radicalisation OR recruitment
	Reduce OR prevent	extremism OR radicalisation or radicalisation OR recruitment

SOURCES INCLUDED IN THE SEARCH

The following databases were queried. They were chosen with the overall timeframe of the REA in mind and to ensure the manageability of the process.

Bibliographic indices

Google Scholar
JSTOR
ScienceDirect (Elsevier)

Research and independent institution websites

International Initiative for Impact Evaluation (3ie) Impact Evaluation Repository	RAND
United States Institute for Peace	Search for Common Ground
International Centre for the Study of Radicalisation	Global Center on Cooperative Security
Governance and Social Development Resource Centre	Reliefweb
ISS Africa	International Alert

Donor and development agency websites

Department for International Development (UK) (DFID) Research for Development Outputs (R4D)	**USAID**
UNDP	**World Bank documents and reports website**
EU CORDIS	

Systematic review databases

Cochrane Database of Systematic Reviews	Systematic Reviews (3ie repository)
Campbell Systematic Reviews	

2.4   SCREENING, SEARCHING, AND CODING

Using Boolean logic, queries for each research question were run in each database. Carrying out this process for each research question ensured all possible elements related to the research objective were obtained through the searches. In cases where long Boolean search strings were not accepted by the databases, queries were simplified to short strings or simply e.g. 'violent extremism' where the given search function did not allow anything more complex. Where possible all returned answers were sorted for 'relevance' using the databases' in-house criteria[25]. The returned texts were also reduced to evidence published after 2005 and to those that were available online (both open source and those behind paywalls).

As this process often returned thousands of potential sources for each query, the researchers only screened the titles and abstracts of the first 200 returned sources for relevance to the research questions. This is standard practice when time or budgetary constraints do not allow for all returned sources to be screened (Waddington et al. 2012; Hagen-Zanker and Mallett 2013). The aim is to screen the two hundredth returned source as thoroughly as the first, rather than to screen all of the returned sources poorly.

The searches of all the databases returned 127,735 documents, of these 3,422 were screened (the first 200 of each database) at the level of titles and abstracts and 3,059 were excluded for not meeting the inclusion criteria.

EXPERT INPUTS AND OTHER SOURCES

For assurance that the search methodology would not miss key information, we also consulted with three experts in the field to obtain further literature recommendations[26]. This led to the addition of 22 sources to the search results.

This gave us a total of 385 sources to be screened at the level of full text screening. Duplicates (12) and those documents which did not meet the inclusion criteria following initial consideration were then removed. This left the researchers with 243 texts to be further reviewed for inclusion/exclusion and coded.

FULL TEXT SCREENING STAGE

All references generated by the systematic searches have been stored using the reference management software Zotero. Researchers conducted this work under supervision from the thematic and methodological leads. Researchers then coded the studies according to country, income level, research design, and question focus (identification and/or intervention) as well as the type of intervention it applied to prevent violent extremism (see **Annex A**).

At the full-text screening stage, the research team applied coding guidelines to ensure that the included studies were relevant to the research question. Several validation points were taken into account during this stage. First, to be valid for inclusion a study must measure the impact and effectiveness of approach (to identification or intervention) on a stated outcome (e.g. the effectiveness of approaches or methodologies in identifying

---

[25] Upon enquiring within two leading British universities, the researchers learnt that the criteria used for such 'relevance' filters are generally not revealed as they are part of a databases' unique selling point.

[26] Dr. Lawrence McNamara, Senior Research Fellow at University of York; Professor Stuart Macdonald, Swansea University and Dr. Olayinka Ajala, Associate lecturer at University of York

populations vulnerable to violent extremism, or the impact of interventions used to address the risk of recruitment and radicalisation). For this specific study, an intervention could be a local project, a broader programme or the implementation of a policy. The intervention may be implemented by donors, host governments, local councils, non-governmental organisations or local civil society organisations. Second, studies that merely measure the impact of a phenomenon other than intervention were excluded. Hence, we excluded any study that reviews the effects of extremism or radicalisation on another outcome, such as development or poverty (see section 2.1 for full list of inclusion and exclusion criteria).

Through this process, 205 documents were excluded as they did not meet the inclusion criteria stated in section 2.1, leaving 38 studies eligible for review in full.

FULL TEXT REVIEW

This phase of the methodology consisted of grading the returned evidence for its relevance and rigour. To facilitate this, a standard grading template was devised to guide the researchers' assessment of the evidence. It asked a series of questions of each source, including its methodology, key themes, cross-cutting themes etc. This exercise was undertaken to inform both the REA's wider narrative and the construction of a detailed analysis sheet in **Annex A** of this report.

Quality scores were given to each source using the quality appraisal framework outlined in section 2.5. 'Low quality' sources were excluded at the end of the coding phase. Coding was undertaken according to the agreed coding framework, as attached in **Annex A**. The final total of included, high and moderate quality documents in the sheets are 13.

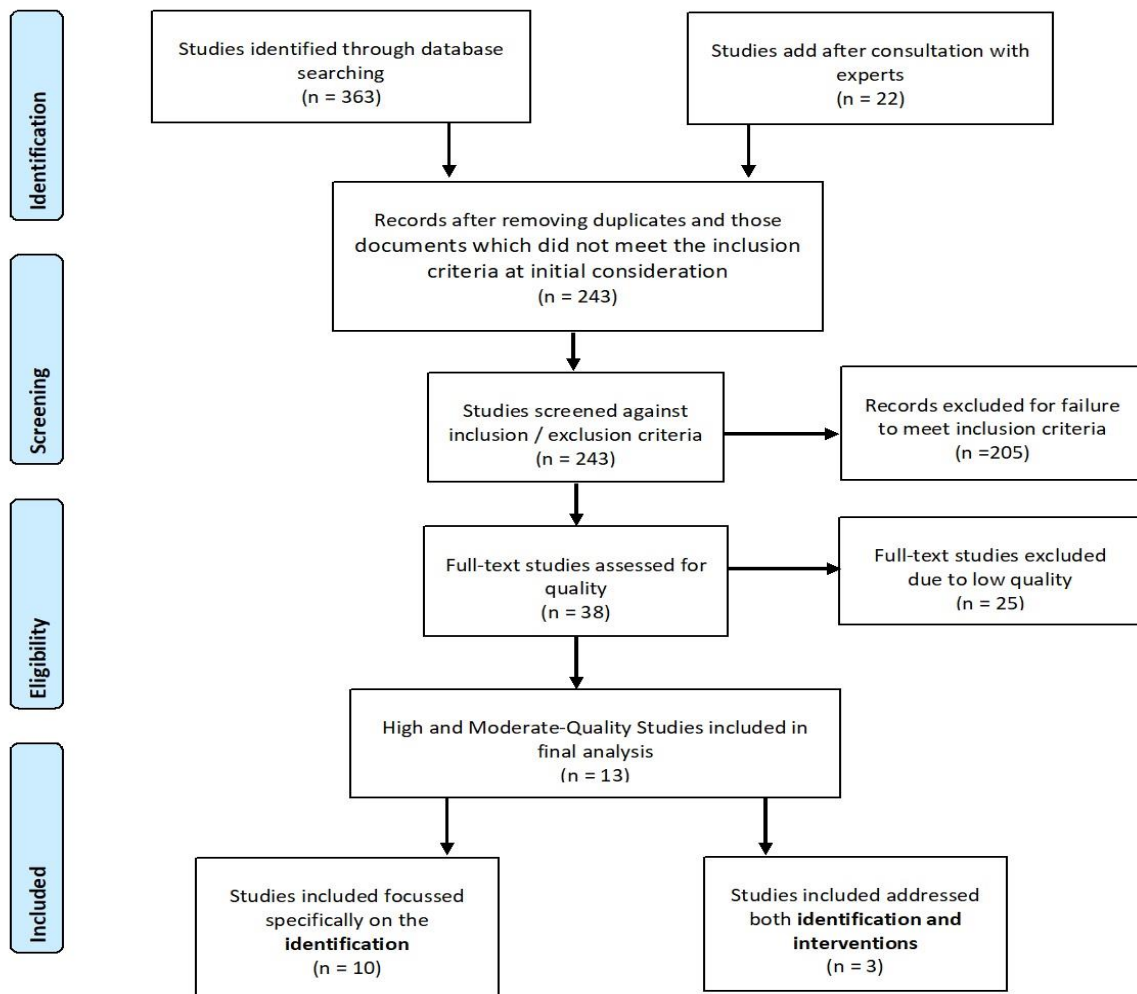The complete selection and evaluation process are mapped in the PRISMA diagram in Figure 1.

Figure 1 PRISMA diagram of analysis of evidence

2.5   QUALITY APPRAISAL

The initial scoping exercise suggested that rigorous evaluative literature on identification of groups vulnerable to VE and effective intervention approaches to reduce the risk of radicalisation was likely to be limited. PVE programmes often seek to understand the absence of violent extremism or to measure a negative. Therefore, the presence of counterfactual analyses (experimental and quasi-experimental methods) is limited. There are also a larger number of qualitative studies that seek to assess the contribution (rather than attribution) of interventions to outcomes.

Of the documents that met the inclusion criteria outlined in Section 2.4, a total of 13 were deemed of high or medium quality and used in this review. The assessment was based on the extent to which a study met the principles of research quality set out for the research question above and on the basis of DFID's How To Note (2014) (Box 1). This approach to quality appraisal recognises that there is no hierarchy of methods – qualitative designs can be equally suited to answering some research questions as experimental or quasi-experimental methods.

Box 1: Principles of Research Quality and Associated Questions

1. Conceptual framing: Does the study acknowledge existing research? Does the study pose a research question or outline a hypothesis?

2. Transparency: What is the geography/context in which the study was conducted? Does the study declare sources of support/funding? Does the study present, reference or link to data sources (interviews, surveys, databases)?

3. Appropriateness: Does the study identify a research design and method? Does the study demonstrate why the chosen design and method are well suited to the research?

4. Cultural sensitivity: Does the study explicitly consider any context-specific cultural factors that may bias the analysis/findings?

5. Validity: Does the study demonstrate measurement validity (is the standard used to assess effectiveness/impact appropriate to the issue at hand and the context)? Does it employ a technique capable of demonstrating that a specific intervention is capable of reducing/preventing extremism?

6. Reliability: to what extent are the measures used in the study internally reliable? Does the analysis show critical engagement with the wider literature?

7. Cogency: to what extent the author considers the study's limitations and/or alternative interpretations of the analysis or different arguments?

Each study was assigned a score of 1-3 against each principle, where 3 stands for 'no concerns', 2 for 'some concerns' and 1 for 'major concerns'. Each study was then assigned an aggregate score based on consideration of each of the principles outlined above to ensure consistency of approach across the studies – each principle carried equal weighting. Researchers recorded notes on each publication they assessed as they proceeded, which can also be found at **Annex A**.

To summarise the quality of evidence succinctly, this REA uses directional arrows as descriptors of the quality as recommended in the 2014 DFID note. The range of possible scores is 7-21. Studies scoring between 7 to 11 are regarded as low quality (↓), those between 12 to 15 as moderate quality (→), and those between 16 and 21 as high quality (↑).

As recommended by DFID's How To Note (2014), the studies were noted in the format '<research type>'; '<research method>'; '<research quality>' e.g. a primary study by Jones, who uses an experimental method, but the paper is only of a moderate quality, the citation is written as Jones (2005) [P;EXP; →].

ASSESSING THE BODY OF EVIDENCE

Our approach to grading the body of evidence assesses four characteristics within each evidence source, in line with DFID's 2014 How to Note:

I. The (technical) quality of the studies constituting the body of evidence (or the degree to which risk of bias has been addressed).

Low quality materials were excluded from the analysis, the technical quality of the studies included fall under two categories:

Box 2: Quality of the Body of Evidence and definition
- High: Many/the large majority of single studies reviewed have been assessed as being of a high quality, demonstrating adherence to the principles of research quality.   - Moderate: Of the single studies reviewed, approximately equal numbers are of a high, moderate and low quality, as assessed according to the principles of research quality.

II. The size of the body of evidence.

The research question did not receive a great attention in the literature. In our opinion the best term to describe the size of body of evidence is "medium". We have documented the number of studies that form the evidence base of the findings, see section 2.4 above.

III. The context in which the evidence is set.

The geography/context in which the studies were conducted was noted during the coding process.

IV. The consistency of the findings produced by studies constituting the body of evidence.

The consistency of the studies included fall under three categories:

- Consistent: A range of studies point to identical, or similar conclusions.
- Inconsistent (contested): One or more study/studies directly refutes or contest the findings of another study or studies carried out in the same context or under the same conditions.
- Mixed: Studies based on a variety of different designs or methods, applied in a range of contexts, have produced results that contrast with those of another study.

Evaluation	Quality	Size	Context	Consistency
Very strong	High	Large (15+ studies)	Global	Consistent
Strong	High	Medium (9-15 studies)	Global	Consistent
Medium	Moderate/High	Medium (9-15 studies)	Specific/Global	Inconsistent
Limited	Low/Moderate/High	Small (2-8 studies)	Specific	Inconsistent
No Evidence	Few to no studies exist (0-1)			

An assessment of the body of evidence can be found in Section 3 below.

3.0   ASSESSMENT OF THE EVIDENCE

Overall, there is a medium sized body of evidence (13 studies) included in this REA – relating to the identification of populations at risk, and the effectiveness of interventions designed to reduce the risk of recruitment and radicalisation that employ such identification approaches. Of the 13 studies, eight have been assessed to contain high quality evidence, and five contain evidence that has been categorised as moderate quality. The strength of evidence around the research question is accordingly considered 'medium' – though, as discussed in more detail below, consistency is notably mixed; with the evidence based on a variety of different designs and/or methods, applied in a range of contexts. However, there is no evidence to suggest that the studies have produced results which contrast with those of other studies, which provides at least one barometer of consistency.

Of the 13 studies, a total of ten focussed specifically on the first part of the research question relating to identification. The remaining three studies address *both* identification *and* intervention.

The methodologies used in the 13 studies are diverse. They implement a variety of different designs and methods – ranging from the combination of technological automation with human expert analysis, to the identification of common characteristics of radicalised individuals, psychometric tools and semantic graph-based approaches. Of the 13 studies included, four are global in nature, and the remaining nine are geographically context specific.

Five of the 13 studies focus on high-income countries, and six on a mix of high, middle, and low-income countries. Of the remaining two studies, one is based in a low-income context (Somalia), and the other in high and middle-income countries (Egypt and Saudi Arabia). A range of contexts are covered – from developed nations, through to fragile and conflict-affected, low-income countries like Somalia – as well as multi-country studies in Tajikistan, Chad and Yemen.

This review of key findings is structured in accordance with the nine specific research sub-queries defined in the terms of reference for the assignment; the same source may therefore be mentioned in several sections, covering the methodologies adopted, their effectiveness, and their strengths and limitations.

This section presents a summary assessment of the evidence. For greater detail, including summaries, evidence type and quality appraisal please refer to **Annex A**.

3.1   IDENTIFICATION

Ten of the 13 studies are focussed specifically and solely on the identification of populations at risk; comprising seven assessed to be of high quality, and three of moderate quality. Three further studies focus both on the identification of populations at risk *and* on the effectiveness of these interventions at reducing the risk of radicalisation. This section 3.1 looks at the evidence they provide on identification only, with the evidence on the interventions applied then discussed in section 3.2.

GEOGRAPHIES STUDIED

Three of the studies are global in nature, and ten are specific to particular geographical contexts. Of the context specific studies included, the UK features in four of the studies – and the USA, the Netherlands, and Egypt feature in two. Somalia, Canada, Saudi Arabia, Denmark, France, Germany, Indonesia, Malaysia, Azerbaijan, Kyrgyzstan, Tajikistan, Algeria, Burkina Faso, Chad, Afghanistan, Bangladesh, India, Pakistan, Iraq, Lebanon, Qatar, Tunisia and Yemen all feature in one of the included studies.

METHODS USED

The methods applied are summarised in the table below. A notable majority of studies utilise statistical analysis of primary data sets, while fewer draw on mixed methods and qualitative observations (based on various types

of interactions to reach their findings). Additionally, it is worth noting that only one secondary research study was selected.

Breakdown of selected studies by method
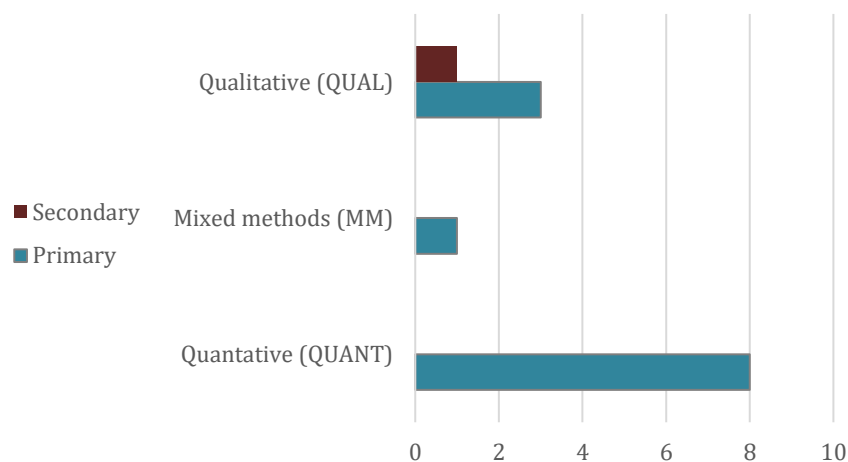
A further breakdown by high level research design shows that all of the eight quantitative (QUANT) studies adopt a statistical analysis (SA) approach; while the mixed methods (MM) study combines observational research (OB) with statistical analysis (SA). Further, the three primary research based qualitative studies all use observational research designs (OB); while the secondary study is a systematic review (SR). Further details are contained within **Annex A**.

3.1.1 WHAT METHODOLOGIES HAVE BEEN USED TO IDENTIFY POPULATIONS AT RISK OF VIOLENT EXTREMIST RECRUITMENT AND RADICALISATION?

Five studies[27] utilise social media, and online sources more broadly, as a means of identifying the vulnerable. This is achieved by analysing internet and social media accounts, such as Twitter and Facebook, often via algorithmic approaches.

Another set of five studies use interview-based methods to assess key personal and social factors (including beliefs). Two of these study individuals already involved with extremist activities[28], two focus on particular identified groups[29], and one compares responses from a number of different groups in one city to interview-based surveys[30].

One further study carried out regression analysis to identify the association between attitudes toward terrorist activities and a set of individual characteristics, drawing on other data sets from surveys[31]. Another primary

---

[27] Fernandez et al (2018); Davey et al (2018); Frenett and Dow (2015); Ferrara et al (2016); Hung et al (2018).

[28] Botha and Abdile (2014); Zierhoffer (2014).

[29] Moaddel and Karabenick (2008), focussing on young adults (18-25) in Egypt and Saudi Arabia; and Bhui et al (2014) focussing on men and women aged 18–45 of Pakistani or Bangladeshi origin, of Muslim heritage, living in East London and Bradford.

[30] Skillicorn et al (2012).

[31] Kiendrebeogo and Ianchovichina (2016).

study draws on the criminal records of individuals identified as having involvement in extremist activity, in an effort to better understand features and pathways of their experience – and its relation to radicalisation[32].

The final study, which is secondary research-based, carries out a systematic review of existing evidence on the process of radicalisation – and of interventions aiming to prevent radicalisation and extremism[33].

The methodologies used in each study are summarised in the following table:

Studies	Methodologies used
Hung et al (2018) [P; SA; ↑]	A semantic graph-based approach to the analysis of tweets; to capture patterns of semantic relations that discriminatingly characterise the radicalised stances of users – with the extracted patterns applied as features for radicalisation classifier training.
Zierhoffer (2014) [P; QUAL; →]	Evaluation of the viability of a threat assessment model developed to calculate the risk of targeted violence as a predictor of future violence by potential lone terrorists. It involves the analysis of thoughts and behaviour patterns that result in an attack on a particular target; applied retroactively to three known lone wolf terrorists.
Basra and Neumann (2016) [P; MM; ↑]	Risk assessment models of former jihadis as a predictor of terrorism in European countries like France, Denmark, Germany, the Netherlands, and the UK. Empirical examination of 79 European jihadists with criminal backgrounds, examining the relevance of their criminal pasts in relation to their terrorist futures.
Moaddel and Karabenick (2008) [P; SA; ↑] **Skillicorn et al (2012) [P; SA;→]** **Bhui et al (2014) [P; SA; →]**	Identification of characteristics associated with attitudes to terrorism, such as: age, gender, marital status, employment situation, education attainment, family demographics, importance of religion in one's life, willingness to sacrifice one's life for beliefs – and locality, as well as country-level common factors.
Botha and Abdile (2014) [P; QUAL; ↑]	Face-to-face interviews with self-confessed members of Al Shabab in Somalia; to develop a profile of typical al-Shabaab recruits.
Christmann (2012) [S; SS; ↑]	Development of typologies or profiles based on common characteristics of those at risk, using psychometric tools.
Frenett and Dow (2015) [P; QUAL;→] **Ferrara et al (2016) [P; SA; ↑]** **Davey et al (2018) [P; SA;→]**	Exploring different ways of using internet and social media accounts (including Twitter and Facebook); often employing algorithmic research and analysis tools to detect extremist users and/or determine indicators of risk of falling into the orbit of violent extremism.
Fernandez et al (2018) [P; SA; ↑]	Development of a computational approach for detecting and predicting the radicalisation influence a user is exposed to, grounded in the notion of 'roots of radicalisation' from social science models.

---

[32] Basra and Neumann (2016).
[33] Christmann (2012).

Studies	Methodologies used
Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑]	Identification of characteristics associated with attitudes to terrorism, such as: age, gender, marital status, employment situation, education attainment, family demographics, importance of religion in one's life, willingness to sacrifice one's life for beliefs – and locality, as well as country-level common factors. The study draws on information on attitudes toward extreme violence, and other relevant characteristics of 30,787 individuals from 27 developing countries.

The methodologies used in the body of evidence are designed to explore various points of enquiry within the scope of the research question of this REA. Some aim to identify indicators or markers for groups at risk of violent extremist recruitment and radicalisation (by contrast with those not, or less, at risk) – whether by virtue of background, personal circumstances, or modes of expression and behaviours. Others aim to determine markers of the degree of radicalisation, or to understand in more depth the process of radicalisation.

The studies are discussed in further detail below, in relation to the various points of enquiry mentioned above – though, again, it is emphasised that a number of studies touch on more than one of these angles.

DETERMINING INDICATORS OF RISK OF VIOLENT EXTREMIST RECRUITMENT AND RADICALISATION – CHARACTERISTICS AND BEHAVIOURS

Exploring different ways of using internet and social media accounts (including Twitter and Facebook), often employing algorithmic research and analysis tools, Frenett and Dow (2015) [P; QUAL; →] find that Facebook page 'likes' and Facebook group memberships are the most relevant factors in indicating risk. Additional markers include a user's cover photo, and the tone and content of their regular posts. Having collated the risk factors provided by the outreach providers, the project team create two distinct sets of search criteria; one for those considered at risk of falling into the orbit of violent far right extremism in North America, and another for violent Islamism in the United Kingdom.

Davey et al (2018) [P; SA; →] seek to combine technology-based algorithmic analysis with expert human assessment. They initially identify over 40,000 potentially "at-risk" users of pages associated with the far-right, and over 2,000 users of pages associated with Islamist extremism. A Natural Language Processing (NLP) algorithm examines user engagement with these pages, detecting instances of language that appear to be violent, aggressive and dehumanising – or keywords relevant to extremism; distilling the total pool of users identified. A sample of 1,600 is then selected to be examined in further detail. Other features of the approach include the manual verification of users and construction of profiles – as well as the assessment of individuals against a purpose-built risk matrix.

Another machine learning framework is utilised by Ferrara et al (2016) [P; SA; ↑] to "leverage a mixture of metadata, network, and temporal features to detect extremist users, and predict content adopters and interaction reciprocity in social media". The first task undertaken is the detection of extremist supporters, utilising a binary classification aimed at detecting Islamic State (ISIS) Twitter accounts, and separating them from those of regular users. This is followed by the creation of a set of 25,000 users randomly sampled among followers of ISIS accounts, which is leveraged to perform the prediction of extremist content adoption. The third task involves predicting whether a regular user will engage in online interactions with extremists. Ferrara et al, 2016 [P; SA; ↑] conclude their analysis by investigating the predictive power of different online actions.

Yet another group of studies seek to analyse the thoughts and behaviour patterns associated with the desire to attack a particular target. For example, Zierhoffer (2014) [P; QUAL; →] studies three terrorists to assess the value of a threat model as a predictor of terrorism. This approach is based on three principles: 1) targeted violence is the culmination of a process of thinking and behaviour that is deliberate and not impulsive; 2) there is interaction

among the potential attacker, a past emotional event, a current situation, and a target; and 3) understanding the behaviours of the individual as they progress from the development of the idea to the actual movements toward the target. Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑] define radicals as those willing to justify the targeting and killing of innocent civilians. They analyse data on attitudes, values and beliefs from 30,787 individuals from 27 developing countries (provided by the Gallup World Poll data set for 2006-12). Using a regression analysis they study the association between sympathetic attitudes toward terrorist activities and a set of individual characteristics, including: age, gender, marital status, employment situation, education attainment, family demographics, importance of religion in one's life, willingness to sacrifice one's life for beliefs – and locality, as well as country-level common factors.

UNDERSTANDING LEVELS AND TYPES OF RADICALISATION

Cluster analysis is used by Bhui et al (2014) [P; SA; →] on a sample of 608 men and women; all of Pakistani or Bangladeshi origin; aged between 18 and 45; and of Muslim heritage. The sample is divided into those most sympathetic to terrorist activities (or most vulnerable), those most condemning (most resistant), and a large intermediary group that acts as a reference or control group. Here, the links between factors such as number of social contacts, social capital, employment opportunities, and depressive symptoms are correlated with levels of sympathy for violent protest and terrorism.

A further breakdown of radicalisation typologies is attempted by Skillicorn et al (2012) [P; SA; →], which subjects some popular theories of radicalisation to empirical testing. The article distinguishes between three types of 'radical', according to the action in which they are engaged. Some engage in politically motivated violence ('terrorists'); others engage in non-violent but illegal, politically motivated acts ("radicals"); and a third group support individuals or groups who engage in politically motivated violence or other illegal acts ('activists'). A survey instrument was designed to capture the distribution of these attitudes across Ottawa's population – and the distribution of respondents relative to the questions: why do individuals end up in one of the three categories in the first place? Are there three different kinds of people who end up in these three different categories? Or are there 'stages' along a 'conveyor-belt' through which a given individual passes?

Conversely, a semantic graph-based approach is suggested by Hung et al (2018) [P; SA; ↑], which breaks down into four key steps: (1) extract named entities and their semantic concepts in tweets; (2) build a semantic graph per user, representing the concepts and semantic relations extracted from his/her posted content; (3) apply frequent sub-graph mining on the semantic graphs to capture patterns of semantic relations that discriminatingly characterise the radicalisation stances of users; and lastly (4) use the extracted patterns as features for radicalisation classifier training. Using data from public sources and government databases of individuals deemed at risk of radicalisation, the authors developed a technology that detects and tracks the behaviour patterns of radicalised individuals, as well as those who had transitioned to the use of violence; these patterns were then used to identify and detect 'at-risk' individuals who appear on a trajectory towards violent extremism. This technology utilises an investigative graph search for detection of patterns that flag individuals at risk of radicalisation, and subsequently follows their radicalisation trajectory over time.

Use of psychometric tools is another methodology utilised in the selected studies. For example, Christmann's (2012) [S; SR; ↑] systematic review of the radicalisation literature finds two different tools that measure features of radicalisation as a psychological construct: The Revised Religious Fundamentalism Scale (RRFS), and the Violent Extremist Risk Assessment (VERA). The RRFS measures attitudes towards one's religious beliefs, and not an adherence to any particular set of beliefs; thus, it has the potential for broad applicability. In contrast, VERA is designed to be used with (and limited to) people with either a history of extremist violence, or convictions for terrorist-related offences. The VERA thus appears applicable only to that small cohort that is in an operational phase, and already radicalised.

Natural Language Processing (NLP) and Collaborative Filtering (CF) algorithms are the basis of the computational approach used by Fernandez et al (2018) [P; SA; ↑] for detecting and predicting the radicalisation influence a user is exposed to; grounded in the notion of 'roots of radicalisation' from social science models. NLP and CF are said to automatically capture the different roots of radicalisation (individual, social and global) for each user, and represent them as keyword-based vector descriptions. In the context of this study, an active participant/user on a social media platform can perform two key actions: post new content or share content which has been created by another user. These actions have been translated into three distinct types of radicalisation influence: Micro-individual; Meso-social; and Macro-global[34]. During the three-month-long study, 112 pro-ISIS and 112 'general' Twitter users[35] were studied using CF to identify users who have a similar rating pattern – and these ratings are subsequently used to compute the prediction of future influences for a specific user. Users are flagged using a set of keywords (such as Dawla, Amaq, Wilayat etc.), and then filtered based on their use of imagery (e.g. ISIS flags, images of radical leaders like al-Baghdadi, Anwar Awlaki etc.) alongside their set of page or profile followers/followings. It is noted, however, that this approach does not aim to determine whether someone is being radicalised or not, but instead to provide a risk level for each user – based on the individual, social and global influences to which s/he is exposed on social media.

Another group of studies developed analytical tools both to support threat assessments (based on typologies and common characteristics of already radicalised or vulnerable individuals), and to understand aspects of the processes of radicalisation of these persons. This group includes a profile of typical al-Shabaab recruits; developed by interviewing 88 former fighters, plus intelligence personnel (Botha and Abdile, 2014 [P; QUAL; ↑]); profiling criminal histories (Basra and Neumann, 2016 [P; MM; ↑]); and assessing the extent to which Egyptian and Saudi youth rely on religious authorities as a source of knowledge about the semipolitical role of religion (Moaddel and Karabenick, 2008 [P; SA; ↑]). Botha and Abdile (2014) [P; QUAL; ↑] seek to profile a typical al-Shabaab recruit in terms of broad socialisation factors, including those stemming from age, family status (particularly during childhood), marital status, level of education, employment status, and the religious literacy of the individuals studied. Their views of authority, including government, are also assessed. Basra and Neumann's (2016) [P; MM; ↑] sample assess the intensity of criminality, age, level of violence, frequency of incarceration, and level at which crimes were conducted (local, national, and transnational). Moaddel and Karabenick (2008) [P; SA; ↑] draw on data from two comparable surveys of late adolescents and young adults (ages 18-25), which were conducted via face-to-face interviews in respondents' residences in Egypt and Saudi Arabia. Importantly, they were conducted by indigenous personnel with considerable experience in their respective countries. The paper also assesses the effects of various sources of information on fundamentalist attitudes, including: higher education, watching TV, and the Internet.

As a brief note in conclusion, the body of evidence overall provides an opportunity to look not only at static data on groups and individuals, but also at processes and dynamics of relationships and lives that may deepen insight into vulnerabilty to radicalisation.

---

[34] The micro-individual influence is best defined as the influence of posts created by an individual, whereas Meso-social influence is that generated by posts (from external sources) that the user has re-shared/retweeted. Each post that a user creates or shares always links (URLs) to an external site, such as YouTube, News sites, blogging platforms or pages etc. The overarching influence of these key sites and online channels (URLs) constitutes the macro-global influence.
[35] Obtained from the public dataset, Kaggle datascience.

The studies in this REA address the challenge of accessing data in this sensitive space in a number of ways. Some opting for 'remote' techniques that utilise information available on people's connections and beliefs through social media, or drawing from surveys and records collected for other purposes – while others (especially those in LIC/MIC settings) make use of direct interviews or surveys, which crucially are undertaken by interviewers with trusted profiles and understanding of cultural context.

Breakdown of selected primary research studies by lead data source type

- Survey or Interview
- Social media usage
- Records Analysis

5

5

2
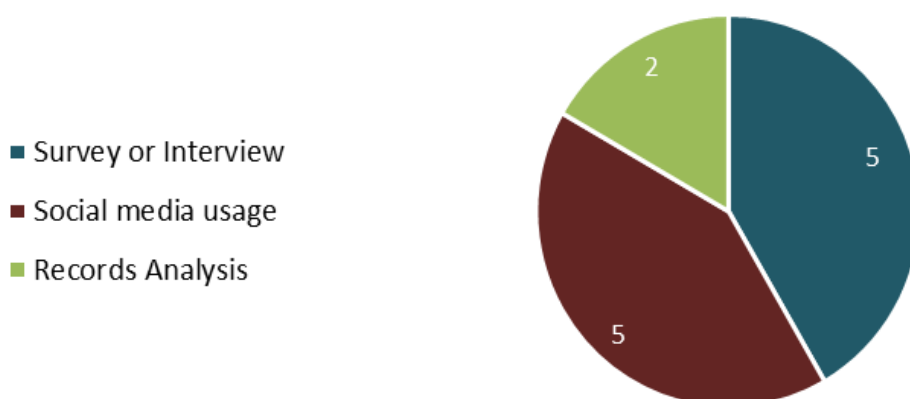
3.1.2 HOW EFFECTIVE HAVE THESE DIFFERENT APPROACHES BEEN AT IDENTIFYING THE GROUPS MOST AT RISK, AND THE UNDERLYING RISK DETERMINANTS?

The effectiveness of these approaches is difficult to assess due to the lack of data on offending rates, the limited reliability of some of the findings to-date, and the experimental nature of many of the assessment tools.

There are only three studies that expressly seek to analyse the effectiveness of their identification approaches. Two of these – Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑] and Bhui et al (2014) [P; SA; →] – identify particular characteristics as being typical of extremists. The third study – Botha and Abdile (2014) [P; QUAL; ↑] – takes a more holistic approach, exploring the broad political socialisation processes of individuals. More technological propositions, such as the use of algorithms, machine learning frameworks, and graph trajectories are also claimed to be promising by Ferrara et al (2016) [P; SA; ↑].

The authors themselves are often aware of the limitations in both the reliability of the findings and the wider applicability of some of these methodologies.

- Some **risk assessment tools** are promising, with strengths deriving from their focus on behaviours rather than demographic data. However, these observations of behaviour require a proximity to the perpetrator which is often not possible.
- In the case of **psychosocial methodologies**, there are often limits to showing which sympathisers with extremism are likely to progress to committing terrorist acts.

- Finally, limits also apply to the use of **online semantic information**, while other results should be interpreted only as partial correlations, unable to explain fully what makes people at risk.

There are also issues around the practicality of acting on some of the findings, and some of the studies are limited by small data sets or a restricted piloting context (thereby pointing to the need for further research). With this in mind, perhaps a more useful approach is to examine which of the studies have successfully pinpointed an "at-risk" group, or indeed a risk determinant.

Looking at risk in developing countries, results reported by Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑] from their regression analysis of survey data suggest that the typical extremist who supports attacks against civilians is more likely to be young (under 33), unemployed (and struggling to make ends meet), living in a rural area, and relatively uneducated. While likely not as religious as others surveyed, such individuals are more likely to be willing to sacrifice their own life for his or her beliefs. Gender and marital status are not found to significantly influence the individual-level variation in attitudes toward extremism. Although these results may vary in magnitude and significance across countries and geographic regions [36], they are robust to various sensitivity analyses. Contrary to common perceptions, extremely radicalised people, and those who mildly support attacks on civilians, tend to be less religious (20%) than other categories of respondents (75%).

Ferrara et al (2016) [P; SA; ↑] argue that the performance of their machine learning framework is extremely promising [37], with some significant features ranking very high in terms of predicting whether users will reciprocate online contacts initiated by extremists. Of the top 11 most significant online behaviour indicators, the study found that some of them – such as the ratio of retweets to tweets, the average number of hashtags adopted, the sheer number of tweets, and the average number of retweets generated by each user – systematically rank very highly in terms of predictive power. These insights shed light on the dynamics of extremist content production, as well as some of the network and timing patterns that emerge in these types of online conversations. The authors argue, therefore, in favour of developing computational tools capable of effectively analysing mass social data streams.

Bhui et al's (2014) [P; SA; →] cluster analysis finds that Muslim men displaying the most sympathy for violent protest and terrorism were more likely to report depression, and to report religion as important, than those who condemn violence. Conversely, those with a greater number of social contacts, and recent migrants, show a greater condemnation of violence. Being unavailable for work because of housewife roles, and disability, are also associated with opposition to violence. Importantly, however, this study does not show which individual sympathisers are likely to progress to terrorist acts. The authors note that their findings are unique to Europe and North America and are designed to address the current priorities of counter-terrorism – namely understanding and preventing the radicalisation of home-grown youth. Thus, it would not be prudent to generalise the findings to other contexts and types of terrorism; although similar research is of course feasible in other contexts.

The survey-based study by Skillicorn et al (2012) [P; SA; →] in Toronto, Canada, finds that radical attitudes appear absent among Muslims who report moral and/or socio-political satisfaction. It suggests that it is when moral dissatisfaction leads to changes in political and religious attitudes that radicalisation becomes a risk. Programmatic policy responses in the West have generally presumed that improving an individuals' life satisfaction, broadly understood as economic opportunity, will decrease the prevalence of radical attitudes. Yet,

---

[36] Indonesia; Malaysia; Azerbaijan; Kyrgyzstan; Tajikistan; Algeria; Egypt; Iraq; Lebanon; Qatar; Tunisia; Yemen; Afghanistan; Bangladesh; India; Pakistan; Burkina Faso; Chad; Comoros; Guinea; Mali; Mauritania; Niger; Nigeria; Senegal; Somaliland and Tanzania.

[37] Results in the different forecasting scenarios are stated to have achieved up to 93 % AUC (Area under the Curve) for extremist user detection, up to 80 % AUC for content adoption prediction, and finally up to 72 % AUC for interaction reciprocity forecasting.

this study suggests that radical attitudes appear among Muslims that are *morally* dissatisfied (as opposed to lacking in economic opportunity). The authors argue that this pathway is present among Muslims because religion and politics are often poorly distinguished. Any intervention here would require policy responses that focus on moral dissatisfaction, which the authors suggest most governments would be unwilling to do.

In investigating criminal backgrounds and the idea of a "crime-terror nexus", no uniform profile emerges; although it is possible to discern patterns. For example, Basra and Neumann's (2016) [P; MM; ↑] analysis of records and interview data from 79 European Jihadists with criminal backgrounds shows that the intensity of criminality amongst European jihadists varies – from 'one-time' criminals, to repeat offenders, and more sustained 'career criminals'. Given the age of those involved, it is likely that many were at the beginning of their criminal 'careers' at the time of the interview. They further find that the vast majority are low-level, local criminals; very few (six of 79 profiles) operated at a national or transnational level. Nonetheless, the prevalence of violent histories (65%) is notable. Prisons also play an important role, with the majority of the individuals in the sample (at least 57%) having been incarcerated on at least one occasion. The authors are confident that a quarter of those embraced jihadism in prison.

In complementary findings to Skillicorn et al (2012) [P; SA; →] regarding the role of moral satisfaction in vulnerability referenced earlier, Moaddel and Karabenick's (2008) [P; SA; ↑] interview-based study analyses religious fundamentalism among Egyptian and Saudi youth – and particularly the degree of reliance on religious authorities for knowledge and understanding[38]. The study argues that among the young Muslims interviewed, adherence to religious orthodoxy (as well as reliance on religious authorities as a source of knowledge) correlates with fundamentalist beliefs and attitudes. They also argue that the psychological states associated with youth – such as fatalism, perceived powerlessness, and feelings of insecurity – are a set of factors that may correlate with religious fundamentalism. In Saudi Arabia – where religious institutions have maintained a strong grip on culture and authority – the links between these institutions and aspects of religiosity to fundamentalism are found to be much stronger than in Egypt. This is perhaps because the authority of traditional religious institutions has been fragmented in Egypt following the influence of Western and secular cultures and because television and the internet are used as key sources of information. In neither country is university education found to have any significant effect on fundamentalist attitudes.

Taking a holistic approach, in which radicalisation is explained in terms of the broad political socialisation process of individuals – rather than stemming from a single root cause, or conditions conducive to terrorism – Botha and Abdile (2014) [P; QUAL; ↑] use interview-based techniques to profile 95 former al-Shabaab fighters in Somalia. They find that the profile of a 'typical recruit' is young (under 30), single, poorly educated (or only provided with religious education), unemployed or under-employed; with a high suspicion of government and a supportive stance toward social protest and revolt. Other note-worthy out-takes from Botha and Abdile's report include:
- 98% of respondents believe that the government only protects its own interests
- 96% believe that revolt is a legitimate course of action
- 64% joined al-Shabaab with a friend
- approximately two thirds of participants felt a sense of belonging in the group
- only 9% of interviewees joined al-Shabaab after their 30th birthdays
- marriage and having children (generally seen as signs of maturity) did not prevent respondents from joining
- only 9% of interviewees indicate that they would remain friends with those they did not agree with
- 60% indicate that they would not listen to advice from a friend

---

[38] For Moaddel and Karabenick, fundamentalism is denoted by a distinctive set of beliefs and attitudes toward one's religion, including obedience to religious norms, belief in the universality and immutability of its principles, the validity of its claims, and its indispensability for human happiness.

In the context of Twitter, results reported by Fernandez et al (2018) [P; SA; ↑] demonstrate the effectiveness of theory-based algorithms in detecting and predicting the level of radicalisation influence each social media user is exposed to. The micro and meso influences were computed for 224 users, based on their tweets and retweets. The macro influence was discarded from the analysis because 63% of the URLs contained links to other Twitter accounts that are now closed. The results show *inter alia* that individual (micro) and social (meso) influences of radicalisation are *both* higher for pro-ISIS users. Although this model does not necessarily determine whether someone is being radicalised or not, it does provide a risk level assessment tool for estimating the level of radicalisation *influence* (individual, social, and/or global) a user is potentially undergoing – which hints at the potential to utilise similar mechanisms to predict the level of radicalisation influence over a wider group of users.

The performance of the "radicalisation trend detection system", INSiGHT, proposed by Hung et al (2018) [P; SA; ↑] as a risk assessment assistance technology is demonstrated in the paper on a variety of datasets. It shows the tendencies of pro-ISIS and anti-ISIS users to mention different entities, places, and interventions. The results show a positive correlation regarding the impact of the use of semantic information to identify pro- and anti-ISIS stances. An additional analysis of the data is performed to identify signals of radicalisation. Results show that discussions between pro-ISIS users tend to mention entities and relations focussed on religion, historical events and ethnicity. Anti-ISIS user discussions tend to focus more on politics, geographical locations, and interventions against ISIS. Anti-ISIS users also tend to mention the entity ISIS with a higher frequency than pro-ISIS users. However, the authors argue that at this stage of research and analysis it is not possible to definitively determine a similarity score threshold by which analysts would be alerted for threats, or be able to screen for high-risk individuals. This would require data of non-violent radicals who exhibit some of the indicators under study – and such data is unavailable.

3.1.3 TO WHAT EXTENT DOES THE EVIDENCE DIFFERENTIATE BETWEEN WOMEN, GIRLS, MEN AND BOYS?

Evidence of gender differences within sub-national groups regarding vulnerability to VE, and insight around whether different identification approaches work better for men and women, is limited and inconclusive. Of the 13 selected studies only five[39] offered any gender analysis or gendered differentiation of outcomes specifically focussed on identification. Even these provide only a very superficial treatment of the issue.

Bhui et al's survey (2014) [P; SA; →] sampled 608 men and women in the UK of Pakistani or Bangladeshi origin and Muslim heritage (aged 18–45), but differential analysis is limited. Similarly, whilst Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑] suggest that gender and marital status are not found to explain significant individual-level variations in attitudes towards extremism, the topic was not given sufficient attention – so it's inadvisable to view this as conclusive.

Moaddel and Karabenick (2008) [P; SA; ↑] do make a substantive attempt to differentiate by gender in their interview-based study of young Muslims in Egypt and Saudi Arabia; where they found significant differences between men and women in both countries. They summarise that young Saudi males reported being less fundamentalist than Saudi females; whereas Egyptian males were found to be more fundamentalist than their female counterparts.

Christmann (2012) [S; SR; ↑] notes in brief findings in journalistic reports that suggest that the majority of female suicide bombers in Iraq are thought to have had family members killed by either multi-national or state

---

[39] Bhui et al (2014); Moaddel & Karabenick (2008); Christmann (2012); Davey et al (2018); Kiendrebeogo & Ianchovichina (2016).

forces in the country, triggering their own recourse to terrorism (though this is not contrasted with findings for males).

Davey et al (2018) [P; SA; →] find in their analysis of users of extremist social media pages that female Islamist candidates presented the youngest average age profile (81% under 30) of all the groups studied (namely male and female extreme right and Islamist candidates respectively). Conversely, female users of right-wing extremist pages had the oldest profile (81% older than 45).

3.1.4   WHAT ARE THE KEY STRENGTHS AND LIMITATIONS ASSOCIATED WITH DIFFERENT IDENTIFICATION METHODOLOGIES?

In this section, all thirteen studies are assessed. The strength of some of the recovered risk assessment tools lies in their focus on explicit behaviours rather than broad demographic data. However, a limitation is that the observations of behaviour require proximity to the subject being assessed. In the case of psychosocial methodologies, there are often limits to showing which of the individuals identified potentially as sympathisers (or who identify themselves as such) are likely to progress to terrorist acts. Online identification methodologies also come with limitations, linked to datasets and the social network structures within which extremist online language is employed. These are elaborated upon below.

METHODOLOGIES DETERMINING INDICATORS OF RISK OF VIOLENT EXTREMIST RECRUITMENT AND RADICALISATION

Each of the studies reviewed raises both challenges and advantages associated with their approach, and chosen data access methods. The interview-based approach of Zierhoffer (2014) [P; QUAL; →] provides a good example of observations of behaviour requiring proximity to the perpetrator; and thus presenting difficulties for the replication and scaling of the approach.

Kiendrebeogo and Ianchovichina (2016) [P; SA; ↑], using an analysis based on surveys of values and beliefs, conclude that because repeated cross-section data was used, it does not allow one to address individual-level unobserved heterogeneity – and consequently, their results should be interpreted as partial correlations. The researchers argue that in order to gain a better understanding of the causes behind radicalisation, there is a need for improved data collection, particularly with reference to panel data.

As for online identification methodologies, Davey et al (2018) [P; SA; →] note several issues with their own approach; suggesting that its reliance on user interactions with 'seed pages' (public and open Facebook pages) as sources for data may provide skewed results. In the case of Twitter, Fernandez et al (2018) [P; SA; ↑] caution that the majority of the datasets used in previous work lack solid verification. Another self-identified weakness of their approach concerns the difficulty of accounting for social network structures within which extremist online language is employed; including further interactions such as 'likes', 'replies' or even 'direct messages' that allow users to signal their agreement or create relationships. The authors argue that a fine-grained temporal analysis, following users' tweets over a long period, would better capture the radicalisation process. They also claim that in future work they will add violence-related algorithms to their detection method.

UNDERSTANDING LEVELS, TYPES, AND PROCESSES OF RADICALISATION

Equally, Hung et al (2018) [P; SA; ↑] (on the use of semantic information to identify pro- and anti-ISIS stances) argue that at this stage in the development of research and analysis, it is not yet possible to determine definitively a similarity score threshold by which analysts would be alerted for threats, or be able to screen for high-risk individuals. This would require data on non-violent radicals who exhibit some of the indicators for evaluation – such data is unavailable.

Bhui et al (2014) [P; SA; →] illustrate the limits to showing which of those people identified as sympathisers are likely to progress to terrorist acts – and also demonstrate that it would not be prudent to generalise the findings to other contexts and types of terrorism.

In the case of psychosocial methodologies (which explore the combined influence of psychological factors and the socio-economic context on an individual's physical and mental state), groups were selected by a statistically acceptable method of cluster analysis. Despite the limitations in identifying which extremist sympathisers are most likely to progress to committing terrorist acts, they do present a sequence of events consistent with the stages of radicalisation[40].

The systematic review by Christmann (2012) [S; SR; ↑] (undertaken for the Youth Justice Board for England and Wales's PVE programmes) also demonstrates the limits to the broader applicability of its findings. The author argues that tools piloted only in one context (with the exception of global social media platforms, such as Twitter and Facebook) should not yet be considered ready for identifying those at risk elsewhere. Overall, he concludes that the field is yet to develop a general causal model or theory of the structural causes of Islamic terrorism, and that there is disagreement as to whether this is in fact achievable.

3.1.5 ARE THERE DIFFERENCES BETWEEN EFFECTIVE METHODOLOGIES IN LOW- AND MIDDLE-INCOME COUNTRIES (LICS/MICS) COMPARED WITH HIGH INCOME COUNTRIES (HICS)?

Of the 13 studies, five concentrate exclusively on HICs (the UK, the USA, Canada, France, Germany, Denmark, and the Netherlands); one on HICs and MICs (Egypt and Saudi Arabia); another on a single LIC (Somalia); and six studies explore all three categories, or are focussed on online activity and data from public and government databases, and as such, are global in orientation.

Those dealing with LICs and/or MICs tend to rely more heavily on direct interviews with individuals; and the assessment of their attitudes and characteristics and are concerned mainly with youth and people living close to extremist groups. With HICs the approach is a little more diffuse; drawing heavily from information from social media platforms and electronic monitoring, often undertaken by US and European law enforcement.

3.2 INTERVENTIONS

Based on the set inclusion criteria and quality appraisal framework, just three studies explore interventions using identification approaches to reduce the risk of violent extremism. As such, there is a limited body of evidence. Due to the limited scope of the evidence base, caution is advised when drawing conclusions in relation to the intervention-focussed aspects of the research question.

As highlighted elsewhere in the report, two of the three studies were assessed as being of moderate quality, and one of high quality. Where stated, two of the three studies provide evidence which suggests positive or neutral effects, while the third indicates a positive impact of intervention. However, none of the studies use regression techniques, or other methods that would allow for causal inference – and are best understood as providing evidence of correlations.

As with the section on identification above, this review of key findings is structured in accordance with four specific sub-queries defined in the terms of reference for the assignment. The same source may therefore be

---

[40] Stage 1: Pre-Radicalisation; Stage 2: Self-Identification; Stage 3: Indoctrination; Stage 4: Jihadization; Silber M, Bhatt A (2007) Radicalisation in the West: the Homegrown Threat. New York New York City Police Department. Cited in Bhui et al (2014).

mentioned in several sections, covering the types of intervention attempted, indicators of progress and impact, and the overall effect of those interventions.

The authors of the three studies share the view that although PVE literature is developing rapidly, serious impact evaluations remain very limited. Christmann (2012) [S; SR; ↑] highlights that most PVE studies are theory driven, and policy recommendations are frequently based on theoretical frameworks or conceptual models, rather than empirical evidence. Thus, for instance, on the issue of value pluralism and "debiasing", it is openly stated that many current debiasing techniques may not be sufficiently robust to generate enduring attitudinal or behavioural changes – and most of the debiasing techniques developed thus far do not expose people to real world situations. Equally, improved "resilience" does not necessarily affect whether young people support violence and armed groups. Davey et al (2018) [P; SA;→] point out that improvements in online behaviour following an intervention, while important, do not necessarily represent behaviour change in the offline world. As such, these outcomes have yet to translate into a demonstrated reduction in terrorism and youth violence.

3.2.1 WHAT INTERVENTIONS HAVE BEEN USED TO ADDRESS THE VULNERABLE GROUPS IDENTIFIED?

The three studies in question used a number of different intervention typologies. All three used counter-narratives; two of them exclusively[41]. The third, which is a systematic review of primary studies, covers a range of intervention typologies – including education, religious education and identity, and counter-narrative[42].

Two of the three studies on interventions build on their analysis of social media use to determine indicators of vulnerability to extremism, and to identify cohorts deemed at risk. Having used a combination of algorithmic analysis with human expert analysis to identify users at risk of radicalisation, their profiles were passed on to the intervention providers who then engaged in one-to-one messaging with users.

Davey et al (2018) [P; SA; →] outline four types of intervention in their study. These are as follows:

- **Primary**: Work with individuals at risk of radicalisation, with activities such as one-to-one mentoring, psycho-social care, and group interventions.
- **Secondary**: Work with immediate families and friends closely connected to an individual known to have engaged with violent extremism; with activities such as family counselling, welfare, and general support.
- **Tertiary**: Activities with an individual's wider spheres of influence, such as institutions and the wider community – which often seek to engage, protect, or divert individuals who may be exposed to heightened risk factors.
- **Rehabilitation**: Working with high-risk groups (e.g. offenders while in prison, on remand, returnees, and foreign fighters) to disengage, and where possible, de-radicalise them away from extremism.

Their work, which was global in scope, focuses principally on primary intervention, through one-to-one, direct online outreach. They selected 814 individuals, who were assessed as displaying signs of radicalisation in their online behaviour, as candidates for online engagement.

Frenett and Dow's work (2015) [P; QUAL; →] aimed to test the viability of an approach based on directly messaging those openly expressing extremist sentiment online and seeking to dissuade them from following that path. In doing so, their project team designed a methodology for identifying candidates, overcoming the technological barriers, and measuring which messages were most effective in eliciting responses most likely to lead to longer-term engagement.

---

[41] Davey et al (2018); Frenett & Dow (2015).
[42] Christmann (2012).

Christmann (2012) [S; SR; ↑] provided a systematic review of scholarly literature to examine the process of radicalisation, particularly among young people. He found only a very limited number of programmatic interventions that had the explicit aim of addressing radicalisation within the literature, and few of these were thoroughly evaluated. Interventions explored in the document included efforts focussed on 'capacity building or empowering young people' and interventions that 'challenged ideology' through theological education. He noted that the use of outreach develops the high levels of trust necessary for interventions to be successful. He also notes findings on the importance of recruiting credible conversational partners to carry out the outreach – perceived as "us" and not as the hostile "them" – who carry authority and legitimacy, are equipped with profound ideological knowledge, and argue in a way that appeals to the radical. Christmann's study also states (p.40) that "the use of outreach methods for Muslim women also makes sense when considering the barriers (and self-imposed barriers) that some Muslim women tend to face (including their traditional role in Muslim society, adherence to religious practices which segregate the sexes, language difficulties, and low confidence), which can account for poor engagement with public institutions and wider society."

It should be noted that the *identification* methodologies specifically applied in relation to determining the interventions covered in Christmann's systematic review are not detailed.

3.2.2 WHAT INDICATORS HAVE BEEN USED TO MONITOR PROGRESS AND MEASURE THE IMPACT OF THESE INTERVENTIONS? AND HOW ROBUST ARE THEY AT MEASURING CHANGE?

This is a difficult question, as all three studies have different, and multiple indicators – often not rigorously applied (and many also mix a variety of approaches). However, some broad types of indicators tend to be used for the different groups or types of intervention. For instance, those initiatives focussing on value pluralism and education examine perspectives on opposing viewpoints and moral dilemmas; and online interventions rely on interactions with project staff, clicks and shares.

Direct engagement with individuals showing signs of radicalisation online is modelled by Davey et al (2018) [P; SA; →]. Engagements were assessed for positive impacts by their length, and qualitatively analysed to "detect instances which suggested that the conversation had possibly generated positive impact by presenting a candidate with an alternative point of view"[43]. Much of this focussed on the internalisation of counter arguments, with the assumption that it indicates empathy, and the diversification of perspectives. An expressed desire to continue engagements with the programme's implementers was also seen as an indicator of positive impact.

In Frenett and Dow's (2015) [P; QUAL; →] online engagement-based research, three metrics were used to assess the impact of online outreach: (1) initial response rates; (2) sustained engagements (conversations that included five or more messages between the candidate and intervention provider); and (3) indications of potential positive impact during the course of the conversations. Of the 154 profiles confirmed to be at risk, 76 messages were drafted and sent to candidates. Over 60% of the messages were actually seen by the candidate – a read rate significantly higher than standard unsolicited mailing campaigns; 59% responded directly or demonstrated a shift in behaviour (change in privacy settings or blocking the intervention provider who sent the message); and a majority were willing to engage in conversation (with 60% of candidates partaking in sustained engagement, 8% brief engagement, and 12% denial of involvement).

Christmann (2012) [S; SR; ↑] does not explicitly identify in his systematic review any indicators used to monitor progress against interventions themselves.

---

[43] Davey et al (2018), p.17.

3.2.3 WHAT EFFECT HAVE INTERVENTIONS HAD (POSITIVE AND NEGATIVE)?

Consistency is not easily determined given the variety of different designs and methods, applied in a range of contexts. Of the three studies in question, Christmann (2012) [S; SR; ↑] claims that the interventions assessed had a positive effect. The other two studies, Davey et al (2018) [P; SA; →] and Frenett and Dow (2015) [P; QUAL; →] claim positive or neutral effects (that is, some individuals engage positively while others do not). Therefore, where the intervention showed positive effects, it did not consistently achieve such effects.

In his systematic review, Christmann (2012) [S; SR; ↑] focussed on a couple of programmes targeted towards addressing Islamic radicalisation in the UK. He also performed an in-depth assessment of the Department for Communities and Local Government's (DCLG) rapid evidence assessment of community interventions to prevent support for violent extremism, *Preventing Support for Violent Extremism through Community Interventions: A Review of the Evidence* (Pratchett et al, 2010).

Christmann observed that the DCLG study shows promise for the use of outreach methods in encouraging participation in intervention programmes; often producing successful outcomes. This is perhaps because such methods focus on developing a relationship of trust with respondents. Building trust is essential for an intervention to be efficacious; particularly when targeting Muslim women – as they tend to face a variety of barriers – including, but not limited to, gender segregation (which may account for their restricted engagement with public institutions, and society as whole). As such, encouraging this demographic to participate in intervention activity means creating a safe, accessible space (aligned with cultural nuances and obligations). Further, the DCLG study also remarks that interventions encouraging young people to take part in education and training programmes (aimed at equipping participants with a 'profound ideological knowledge') are capable of successfully challenging radicalisation.

The DCLG study praised the Slotervaart Project in particular – a Dutch community-based programme aimed at building resilience to radicalisation (in a Netherlands borough with a significant immigrant population) – for successfully implementing outreach methods, and using sound mechanisms to achieved them. The project organised gatherings that discussed the Muslim identity, Islam, and a range of political issues. These gatherings were first open to religious-secular circles; and eventually widened the debate to include non-Muslim circles. According to the DCLG authors, the success of the Slotervaart Project can be attributable to:

- The "respectful, listening mode of interaction" with the groups targeted by the intervention
- A focus on also engaging with groups, agencies, and public bodies and the "wider community [around the targeted group] and those who were crucial to sustaining the communities' engagement with the project"

As previously stated, Davey et al (2018) [P; SA; →] used three measures to determine the effectiveness of online engagement: (1) initial responses rates – whether or not a candidate responded to an outreach attempt; (2) sustained engagement, which included conversations with five or more exchanges with a provider; and (3) identifying conversations that had qualitative indicators of potential positive impact – including suggestions that a conversation may have changed an individual's mind, admission that their online behaviour may be harmful to others, or requests to continue a conversation on another medium. In total, of 800 individuals identified, conversations were initiated with 569; of whom 112 responded to online outreach; 76 individuals engaged in a sustained conversation; with 8 of these sustained conversations exhibiting indicators of potential positive impact. On this basis, Davey et al concluded that although these numbers may seem small, they demonstrate that once individuals are engaged in a counter-conversation this can be leveraged to generate positive impact in a significant proportion of cases. Further, they suggest that a one in five response rate for intervention candidates replying to the initial message should be viewed as a measure of success; noting that this underlines the importance of achieving scale at the identification stage.

3.2.4 ARE THERE DIFFERENCES BETWEEN EFFECTIVE INTERVENTIONS IN LOW- AND MIDDLE-INCOME COUNTRIES, COMPARED WITH HIGH INCOME COUNTRIES?

While the number of studies is limited (and care is therefore needed in extrapolating from the findings), the body of evidence does draw on evidence from a range of geographical contexts. Of the three studies, one is focussed on the UK and the US; another focuses on the UK, the Netherlands and an unspecified list of Islamic countries; and the third is global in its coverage.

In the LICs/MICs, there is emphasis on employment and livelihoods – often incorporating elements aimed at building resilience, self-esteem and tolerance in various ways. In the HICs (where the sources are strongly oriented towards immigrant communities), the balance tends to be reversed – with a heavier emphasis on social media, tolerance, value pluralism, and "debiasing" through educational initiatives. Resilience, self-esteem and empathy are frequently addressed as objectives in and of themselves, rather than as add-ons to economic initiatives.

4.0   CONCLUSION

An evidence base is emerging around effective approaches to identifying those vulnerable to VE and interventions to reduce the risk but is not yet able to yield conclusive insights. The available research found by this REA of sufficient quality for inclusion in the evidence base is **medium** in relation to identification approaches (13 studies); but when looking at a subset (i.e. on effectiveness of interventions based on those approaches), the evidence base is **limited** (consisting of just three remaining studies).

Much of the existing literature is derived from limited or unrepresentative data sets, and/or of restricted or unclear wider applicability – and though PVE literature more generally is developing rapidly, serious impact evaluations remain very limited. As noted by one of the authors of the systematic review[44], there is as yet no general causal model or theory of the structural causes of extremism, and there is disagreement as to whether such a model is even achievable. Many studies list several possible factors, often including social-psychological models, but fail to specify the interactions between the listed factors in any detail.

Furthermore, links and causality between these various factors and radicalisation or extremism are very frequently unclear. Identification of particular characteristics as being typical of extremists does not mean that all persons with those characteristics can be classified in such a way. Furthermore, treatment of gender issues is generally superficial, and most sources contain no concerted analysis of the differential impact of PVE interventions on women/girls and men/boys, or of the existence and/or impact of any initiatives specifically aimed at women.

Use of indicators to monitor progress and measure impact is also problematic. Most studies have multiple indicators which are often difficult to disaggregate – and are not very rigorously applied anyway. But some broad types of indicators tend to be used for the different types of intervention, as summarised above. Online work delivered promising engagement with counter-narratives; but could not report any direct cognitive shifts. Importantly, in the case of psychosocial and other methodologies, there are often limits to showing which people identified as sympathisers are likely to go on to commit violent acts.

Much of the data and research found focusses wholly or in part on HICs, and in particular the evidence on interventions with groups in LICs/MICs is limited. However, some interesting contrasts have emerged between high and low-income countries – with efforts in LICs/MICs focussing on employment and livelihoods (often incorporating elements aimed at building resilience, self-esteem and tolerance) – while in HICs (where the

---

[44] Christmann (2012).

sources are strongly oriented towards immigrant communities) the balance tends to be reversed; with a heavier primary emphasis on tolerance, value pluralism, and "debiasing" through educational initiatives.

In spite of the limitations, the evidence provides some insights into both *who* is at risk and *how* best to engage with them. It also provides a range of examples of how questions around identifying and intervening with groups at risk can be tackled methodologically – through various types of analysis of online and offline data, or through discussion and personal engagement. The body of evidence – looked at collectively – also provides materials to understand and further investigate radicalisation as a process within an individual that is affected by various contextual factors, interactions and influences; a dynamic rather than a fixed feature of individuals. This is an important refocus where efforts seek to prevent and reduce incipient or future radicalisation. Though this comes with a range of ethical and legal sensitivities that must be addressed carefully.

To progress our understanding of these complex issues, we should perhaps move away from the simple "what works?" question, and towards a more holistic, useful set of questions: "What works? For whom? In what circumstances? And how?". A particular issue remaining to be tackled in building a robust evidence base around the effectiveness of the various approaches to identification and risk reduction is the lack of focus on *effectiveness* in the existing research studies. A robust evidence base of studies that rigorously apply their hypotheses and indicators is required here, despite the methodological challenges that this entails. In addition, given the gaps in the existing research base around gender and geographical differences, concerted efforts will need be made to better understand how these basic demographic attributes impact the likelihood of radicalisation – and the types of intervention that will be effective for each group. Such work could crucially aid progress by the research and policy community in understanding whether (and what aspects of) findings from one setting can be applied to another.

BIBLIOGRAPHY

1. "Countering Violent Extremism and Humanitarian Action." <u>Norwegian Refugee Council</u> June.2017: 1-6.
2. Abdile, Mahdi, and Anneli Botha. "Radicalisation and al-Shabaab recruitment in Somalia." <u>Institute for Security Studies Papers</u> Sept. 2014: 1-20.
3. Attree, Larry. "Shouldn't YOU be Countering Violent Extremism?" <u>Saferworld</u> Mar. 2017
4. Basra, Rajan, and Peter R. Neumann. "Criminal pasts, terrorist futures: European jihadists and the new crime-terror nexus." <u>Perspectives on Terrorism</u> 10.6 (2016): 25-40.
5. Bhui, Kamaldeep, Brian Everitt, and Edgar Jones. "Might depression, psychosocial adversity, and limited social assets explain vulnerability to and resistance against violent radicalisation?." <u>PloS one</u> 9.9 (2014): 1-10.
6. Christmann, Kris. "Preventing religious radicalisation and violent extremism: A systematic review of the research evidence." <u>Youth Justice Board</u> 2012:1-76.
7. Davey, J.; Birdwell, J.; Skellett, R. "Counter Conversations: A model for direct engagement with individuals showing signs of radicalisation online." <u>Institute for Strategic Dialogue</u> 2018: 1-31.
8. Fernandez, Miriam, Moizzah Asif, and Harith Alani. "Understanding the roots of radicalisation on twitter." <u>Proceedings of the 10th ACM Conference on Web Science</u> (2018): 27-30.
9. Ferrara, Emilio, et al. "Predicting online extremism, content adopters, and interaction reciprocity." <u>SocInfo</u> (2016): 22-39.
10. Frenett, Ross, and Moli Dow. "One to one online interventions: A pilot CVE methodology." <u>Institute for Strategic Dialogue</u> 2015: 1-27.
11. Hagen-Zanker, Jessica and Mallett, Richard. How to do a rigorous, evidence-focussed literature review in international development A Guidance Note. <u>ODI Working and Discussion Papers Sept 2013</u> (2013)
12. Hung, Benjamin WK, Anura P. Jayasumana, and Vidarshana W. Bandara. "INSiGHT: A system to detect violent extremist radicalization trajectories in dynamic graphs." <u>Data and Knowledge Engineering</u> 118 (2018): 52-70.
13. Kiendrebeogo, Youssouf, and Elena Ianchovichina. "Who supports violent extremism in developing countries? Analysis of attitudes based on value surveys." <u>The World Bank</u> June 2016: 1-59.
14. Skillicorn, David B., Christian Leuprecht, and Conrad Winn. "Homegrown Islamist radicalization in Canada: process insights from an attitudinal survey." <u>Canadian Journal of Political Science/Revue canadienne de science politique</u> 45.4 (2012): 929-56.
15. Zierhoffer, Diane M. "Threat Assessment: Do Lone Terrorists Differ from Other Lone Offenders?" <u>Journal of Strategic Security</u> 7.3 (2014): 48-62.

Section 1: Summary of bibliographic data on studies included in the REA

Item Nr	Author(s)	Year	Document Title	Journal/ Publication	Publisher	Pages	Issue	Volume	URL Online	Summary/Abstract
1	Bhui, K; Everitt, B; Jones, E	2014	Might Depression, Psychosocial Adversity, and Limited Social Assets Explain Vulnerability to and Resistance against Violent Radicalisation?	PLOS ONE	PLoS Journals	10	9	9	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105918	This study tests whether depression, psychosocial adversity, and limited social assets offer protection or suggest vulnerability to the process of radicalisation. Methods A population sample of 608 men and women of Pakistani or Bangladeshi origin, of Muslim heritage, and aged 18–45 were recruited by quota sampling. Radicalisation was measured by 16 questions asking about sympathies for violent protest and terrorism. Cluster analysis of the 16 items generated three groups: most sympathetic (or most vulnerable), most condemning (most resistant), and a large intermediary group that acted as a reference group. Associations were calculated with depression (PHQ9), anxiety (GAD7), poor health, and psychosocial adversity (adverse life events, perceived discrimination, unemployment). We also investigated protective factors such as the number social contacts, social capital (trust, satisfaction, feeling safe), political engagement and religiosity. Results Those showing the most sympathy for violent protest and terrorism were more likely to report depression (PHQ9 score of 5 or more; RR = 5.43, 1.35 to 21.84) and to report religion to be important (less often said religion was fairly rather than very important; RR = 0.08, 0.01 to 0.48). Resistance to radicalisation measured by condemnation of violent protest and terrorism was associated with larger number of social contacts (per contact: RR = 1.52, 1.26 to 1.83), less social capital (RR = 0.63, 0.50 to 0.80), unavailability for work due to housekeeping or disability (RR = 8.81, 1.06 to 37.46), and not being born in the UK (RR = 0.22, 0.08 to 0.65). Conclusions Vulnerability to radicalisation is characterised by depression but resistance to radicalisation shows a different profile of health and psychosocial variables. The paradoxical role of social capital warrants further investigation.
2	Hung, B. W. K.; Jayasumana, A. P.; Bandara, V. W.	2018	INSiGHT: A system to detect violent extremist radicalization trajectories in dynamic graphs	Data & Knowledge Engineering	Elsevier B. V.	19	0	0	http://www.sciencedirect.com/science/article/pii/S0169023X17303920	This paper proposes the development of a radicalization trend detection system as a risk assessment assistance technology that relies on data mined from public data and government databases for individuals who exhibit risk indicators for extremist violence, and enables law enforcement to monitor those individuals at the scope and scale that is lawful, and accounts for the dynamic indicative behaviours of the individuals and their associates rigorously and automatically. We frame our approach to monitoring the radicalization pattern of behaviours as a unique dynamic graph pattern matching problem, and develop a technology called INSiGHT (Investigative Search for Graph-Trajectories) to help identify individuals or small groups with conforming subgraphs to a radicalization query pattern, and follow the match trajectories over time. This paper demonstrates the performance of INSiGHT on a variety of datasets, to include small synthetic radicalization-specific datasets and a real behavioural dataset of time-stamped radicalization indicators of recent U.S. violent extremists.
3	Zierhoffer, D. M.	2014	Threat Assessment: Do Lone Terrorists Differ from Other Lone Offenders?	Journal of Strategic Security	University of South Florida Board of Trustees	48-62	3	7	https://www.jstor.org/stable/26465193	This study evaluates the viability of a threat assessment model developed to calculate the risk of targeted violence as a predictor of future violence by potential lone terrorists. There is no profile, to date, which would assist in the identification of a lone terrorist prior to an attack. The threat assessment model developed by Borum, Fein, Vossekuil, and Berglund and described in "Threat Assessment: Defining an approach for evaluating risk of targeted violence" (1999) poses ten questions about the patterns of thinking and behaviours that may precipitate an attack of targeted violence. Three terrorists are studied to assess the model's value as a predictor of terrorism. It is assessed for its use within law enforcement, during an investigation of someone brought to attention as a possible terrorist and for family members or friends who suspect potential terrorist behaviour. Would these questions encourage someone to report a friend to prevent a possible attack? This threat assessment model provides a foundation for future research focused on developing a structured risk assessment for lone terrorists. In its present form, the questions can assist both citizens and law enforcement personnel in identifying the patterns of thought and behaviour potentially indicative of a lone terrorist.
4	Basra, R.; Neumann, P. R.	2016	Criminal Pasts, Terrorist Futures: European Jihadists and the New Crime-Terror Nexus	Perspectives on Terrorism	Terrorism Research Institute	25-40	6	10	https://www.jstor.org/stable/26297703	The prevalence of criminal backgrounds amongst European jihadists is remarkable. Whether amongst 'foreign fighters' that have travelled to Syria and Iraq, or amongst those involved in terrorism in Europe, criminal pasts are common. Yet, they remain unexamined. This article presents a unique empirical examination of 79 European jihadists with criminal backgrounds, examining the relevancy of their criminal pasts in relation to their terrorist futures. The results fall into four themes. Firstly, jihadism can affect a criminal's radicalisation process in two ways: it can offer redemption from past sins, or it can legitimise crime. Secondly, prisons offer an environment for radicalisation and networking amongst criminals and extremists. Thirdly, criminals develop skills that can be useful for them as extremists, such as access to weapons and forged documents, as well as the psychological 'skill' of familiarity with violence. Finally, white-collar and petty crime is often used to finance extremism. The results challenge conceptions on radicalisation, and can affect counter-terrorism responses.
5	Moaddel, M.; Karabenick, S. A.	2008	Religious Fundamentalism among Young Muslims in Egypt and Saudi Arabia	Social Forces	Oxford Journals, Oxford University Press	1675-1710	4	86	https://www.jstor.org/stable/20430825	Religious fundamentalism is conceived as a distinctive set of beliefs and attitudes toward one's religion, including obedience to religious norms, belief in the universality and immutability of its principles, the validity of its claims, and its indispensability for human happiness. Surveys of Egyptian and Saudi youth, ages 18-25, reveal that respondents with higher levels of fundamentalism are more likely to rely on religious authorities as the source of knowledge about the socio-political role of Islam, support religious law, be fatalistic, and feel insecure. They are alsoless likely to watch TV. Saudi females are more fundamentalist than males, but in Egypt, the opposite held true. Country-specific effects are present, and there are implications for future research.
6	Skillicorn, D. B.; Leuprecht, C.; Winn, C.	2012	Homegrown Islamist Radicalization in Canada: Process Insights from an Attitudinal Survey	Canadian Journal of Political Science / Revue Canadienne de science politique	Canadian Political Science Association and the Société Quebecoise de science politique	929-956	4	45	https://www.jstor.org/stable/23391062	Theories of radicalization make implicit predictions about variation among attitudes in the communities from which radicals are drawn. This article subjects some popular theories to empirical testing. We designed a survey instrument to measure attitudes to issues widely believed to be relevant to radicalization and deployed it among Muslim communities in Ottawa. The results are remarkably inconsistent with patterns of variation in attitudes predicted by popular theories of radicalization. Instead, they show variation of attitudes along three independent dimensions: social/economic/political satisfaction/dissatisfaction, moral/religious satisfaction/dissatisfaction, and a dimension that appears to be associated with radicalization. This suggests that governments may have less policy leverage to mitigate radicalization than generally supposed.

Section 1: Summary of bibliographic data on studies included in the REA

Item Nr	Author(s)	Year	Document Title	Journal/ Publication	Publisher	Pages	Issue	Volume	URL Online	Summary/Abstract
7	Ferrara, E.; Wang, W.; Varol, O.; Flammini, A.; Galstyan, A.	2016	Predicting Online Extremism, Content Adopters, and Interaction Reciprocity	Social Informatics	Springer Link	22-39	0	0	https://link.springer.com/chapter/10.1007/978-3-319-47874-6_3	We present a machine learning framework that leverages a mixture of metadata, network, and temporal features to detect extremist users, and predict content adopters and interaction reciprocity in social media. We exploit a unique dataset containing millions of tweets generated by more than 25 thousand users who have been manually identified, reported, and suspended by Twitter due to their involvement with extremist campaigns. We also leverage millions of tweets generated by a random sample of 25 thousand regular users who were exposed to, orconsumed, extremist content. We carry out three forecasting tasks, (i) to detect extremist users, (ii) to estimate whether regular users will adopt extremist content, and finally (iii) to predict whether users will reciprocate contacts initiated by extremists. All forecasting tasks are set up in two scenarios: a post hoc (time independent) prediction task on aggregated data, and a simulated real-time prediction task. The performance of our framework is extremely promising, yielding in the different forecasting scenarios up to 93 % AUC for extremist user detection, up to 80 % AUC for content adoption prediction, and finally up to 72 % AUC for interaction reciprocity forecasting. We conclude by providing a thorough feature analysis that helps determine which are the emerging signals that provide predictive power in different scenarios.
8	Abdile, M.; Botha, A.	2014	Radicalisation and al-Shabaab recruitment in Somalia	N/A	Institute for Security Studies	20	266	0	https://www.ingentaconnect.com/content/sabinet/ispaper/2014/00000201 4/00000266/art00001	Researchers interviewed former al-Shabaab fighters and identified a complex array of reasons for why they joined the organisation. Interviewers developed a profile of typical al-Shabaab recruits and identified factors facilitating their recruitment, including religious identity, socio- economic circumstances (education, unemployment), political circumstances and the need for a collective identity and a sense of belonging. The reasons for al-Shabaab's rise are discussed and recommendations are made to the Somali government, countries in the region and international organisations and donors on how to counter radicalisation and recruitment to al-Shabaab.
9	Christmann, K.	2012	Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence	N/A	Youth Justice Board for England and Wales	77	0	0	http://www.justice.gov.uk/downloads/research-and-analysis/yjb/preventing-violent-extremism-systematic-review.pdf	The purpose of this systematic review is to examine the scholarly literature on the process(es) of radicalisation, particularly among young people, and the availability of interventions to prevent extremism. The review was undertaken to inform the national evaluation of the Youth Justice Board for England and Wales' (YJB) preventing violent extremism programmes within the youth justice system, and as such, represents one of the research outputs from that study. The review found that the evidence base for effective preventing violent extremism interventions is very limited. Despite a prolific output of research, few studies contained empirical data or systematic data analysis. Furthermore, although a growing body of literature investigating the radicalisation process is emerging, the weight of that literature is focused upon terrorism rather than radicalisation. As such, the evidence is concerned with that smaller cohort of individuals who, once radicalised, go on to commit acts of violence in the pursuit of political or religious aims and objectives. This introduces a systematic bias in the literature, away from the radicalisation process that precedes terrorism, including radicalisation that does not lead to violence. Despite these limitations, the systematic review found that Islamic radicalisation and terrorism emanate from a very heterogeneous population that varies markedly in terms of education, family background, socio-economic status and income. Several studies have identified potential risk factors for radicalisation, and, among these, political grievances (notably reaction to Western foreign policy) have a prominent role.
10	Davey, J.; Birdwell, J.; Skellett, R.	2018	Counter Conversations: A model for direct engagement with individuals showing signs of radicalisation online	N/A	Institute for Strategic Dialogue	32	0	0	http://www.isdglobal.org/wp-content/uploads/2018/03/Counter-Conversations_FINAL.pdf	This report outlines the results of a programme trialling a methodology for identifying individuals who are demonstrating signs of radicalisation on social media, and engaging these individuals in direct, personalised and private 'counter-conversations' for the purpose of de-radicalisation from extremist ideology and disengagement from extremist movements. This is the first programme globally which has trialled the delivery of online interventions in a systematised and scaled fashion.
11	Frenett, R.; Dow, M.	2015	One to One Online Interventions. A Pilot CVE Methodology	N/A	Institute for Strategic Dialogue	15	0	0	https://www.isdglobal.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf	This report outlines the results of a pilot project undertaken by the Institute for Strategic Dialogue (ISD) in partnership with Curtin University and members of the global Against Violent Extremism (AVE) network. The project assesses the viability of direct online interventions with those at risk of falling into the orbit of violent extremist organisations. This report details the results of the pilot, outlines lessons learned and invites other organisations to critique and replicate these results. As the aim of this project was to test a methodology which could be replicated and utilised by other organisations, it was decided to avoid costly bespoke software to identify and engage those expressing extremist ideas online and instead rely on tools freely available to all. The pilot aimed to test this methodology across ideologies and geographies. Data was gathered on an on-going basis throughout this project, with the aim of providing a strong foundation for future projects.
12	Kiendrebeogo, Y.; Ianchovichina, E.	2016	Who Supports Violent Extremism in Developing Countries? Analysis of Attitudes Based on Value Surveys	N/A	World Bank Group, Middle East and North Africa Region, Office of the Chief Economist	61	WPS9 671	0	http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-7691	What are the common characteristics among radicalized individuals, willing to justify attacks targeting civilians? Drawing on information on attitudes toward extreme violence and other characteristics of 30,787 individuals from 27 developing countries around the world, and employing a variety of econometric techniques, this paper identifies the partial correlates of extremism. The results suggest that the typical extremist who supports attacks against civilians is more likely to be young, unemployed and struggling to make ends meet, relatively uneducated, and not as religious as others, but more willing to sacrifice their own life for his or her beliefs. Gender and marital status are not found to explain significantly individual-level variations in attitudes towards extremism. Although these results may vary in magnitude and significance across countries and geographic regions, they are robust to various sensitivity analyses.
13	Fernandez, M.; Asif, M.; Alani, H.	2018	Understanding the Roots of Radicalisation on Twitter	WebSci '18: 10th ACM Conference on Web Science	ACM	10	0	0	http://dl.acm.org/citation.cfm?doid=3201064.3201082	In this paper we propose a computational approach for detecting and predicting the radicalisation influence a user is exposed to, grounded on the notion of 'roots of radicalisation' from social science models. This approach has been applied to analyse and compare the radicalisation level of 112 pro-ISIS vs.112 "general" Twitter users. Our results show the effectiveness of our proposed algorithms in detecting and predicting radicalisation influence, obtaining up to 0.9 F-1 measure for detection and between 0.7 and 0.8 precision for prediction. While this is an initial attempt towards the effective combination of social and computational perspectives, more work is needed to bridge these disciplines, and to build on their strengths to target the problem of online radicalisation.

Item Nr	Author(s)	Year	Income level and geography				Research Design							Gender	Conflict Sensitivity	Research question focus		
			High Income	Middle Income	Low income	Country	Experimental (Randomized Control Trail)	Quasi-experimental	Statistical analysis	Systematic Review	Mixed methods	Qualitative	Add Other RD description here if applicable	Is there any gender analysis or gendered differentiation of outcomes?	Conflict-sensitivity/do no harm (where there is explicit evidence that conflict sensitivity has been considered as part of implementation)	Study focus on methodology or approach used to identify populations at risk	Study focus on interventions to reduce the risk of recruitment and radicalisation	Both
1	Bhui, K; Everitt, B; Jones, E	2014	x			UK			x				Sample of 608 men and women of Pakistani or Bangladeshi origin, of Muslim heritage, and aged 18–45	Yes	No	x		
2	Hung, B. W. K.; Jayasumana, A. P.; Bandara, V. W.	2018				Global			x				Semantics. Entities are extracted from the tweets of users' timelines (e.g. "ISIS", "Syria", "United Nations") and expanded with their corresponding semantic concepts (e.g. "Jihadist Group", "Country", "Organisation") and relations (e.g., Military intervention against ISIL, place, Syria) by using DBpedia. Frequent sub-graph mining is applied over the extracted semantic graphs to capture patterns of semantic relations that help discriminating the radicalisation stances of users. These patterns are then used as features (so-called semantic features in our work) for detecting the radicalisation stances of users on Twitter. The trial of the technique uses a dataset of 1,132 European Twitter users (together with their timelines) equally divided in pro-ISIS and anti-ISIS.	No	No	x		
3	Zierhoffer, D. M.	2014	X			USA						x	Risk Assessment model: Three terrorists are studied to assess the model's value as a predictor of terrorism.	No	No	x		
4	Basra, R.; Neumann, P. R.	2016	x			Denmark, France, Germany, Netherlands, UK					x		A multi-lingual team of ICSR researchers compiled a database containing the profiles of 79 recent European jihadists with criminal pasts. All data was gathered from open sources, such as newspaper articles, court documents, and government reports. A series of interviews with current and former counter-terrorism officials, in order to assess the implications of the findings, and check database entries. To assess individuals' criminal histories, the researchers coded for time spent in prison (as well as the number of stays); criminal convictions; involvement with firearms; types of crime, such as violent (for example, assault, robbery) or petty crime (for example, trespassing, theft), drug dealing, trafficking, and white-collar crime (for example, identity theft or financial fraud).				x	

Item Nr	Author(s)	Year	Income level and geography				Research Design							Gender	Conflict Sensitivity	Research question focus		
			High Income	Middle Income	Low income	Country	Experimental (Randomized Control Trail)	Quasi-experimental	Statistical analysis	Systematic Review	Mixed methods	Qualitative	Add Other RD description here if applicable	Is there any gender analysis or gendered differentiation of outcomes?	Conflict-sensitivity/do no harm (where there is explicit evidence that conflict sensitivity has been considered as part of implementation)	Study focus on methodology or approach used to identify populations at risk	Study focus on interventions to reduce the risk of recruitment and radicalisation	Both
5	Moaddel, M.; Karabenick, S. A.	2008	x	x		Egypt, Saudi Arabia			x				The data presented here are based on two comparable surveys of late adolescents and young adults (ages 18-25), which were conducted in Egypt and Saudi Arabia in the spring and summer of 2005. Surveys required approximately 45 minutes on average and were conducted in face-to-face interviews in respondents' residences. Importantly, they were conducted by indigenous personnel with considerable experience in their respective countries. The Egyptian sample included three cities and rural surroundings: Alexandria; El-Minya, and Cairo. The Saudi survey (administered July 10-25, 2005) also included three cities and rural surroundings: Jeddah, Riyadh and Dammam-Khobar.	Yes	No	x		
6	Skillicorn, D. B.; Leuprecht, C.; Winn, C.	2012	x			Canada			x				The Ottawa Radicalization Survey was conducted by COMPAS Research. Respondents constituted a representative sample of 506 residents of Ottawa who were either Muslim («=455), Canadian or foreign born or of Christian Arab origin (n=51).	No	No	x		
7	Ferrara, E.; Wang, W.; Varol, O.; Flammini, A.; Galstyan, A.	2016	x	x	x	Global			x				Millions of tweets by 25,000 Twitter users suspended from extremism and 25,000 regular users	No	No	x		
8	Abdile, M.; Botha, A.	2014			x	Somalia						x	88 respondents were interviewed, another seven interviewees interviewed off the record. The study aims to explain radicalisation from the perspective of individual, self-professed members of al-Shabaab. It will explain radicalisation in terms of the broad political socialisation process rather than from the perspective of a single root cause, or conditions conducive to terrorism.	No	No	x		

Item Nr	Author(s)	Year	Income level and geography				Research Design							Gender	Conflict Sensitivity	Research question focus		
			High Income	Middle Income	Low income	Country	Experimental (Randomized Control Trail)	Quasi-experimental	Statistical analysis	Systematic Review	Mixed methods	Qualitative	Add Other RD description here if applicable	Is there any gender analysis or gendered differentiation of outcomes?	Conflict-sensitivity/do no harm (where there is explicit evidence that conflict sensitivity has been considered as part of implementation)	Study focus on methodology or approach used to identify populations at risk	Study focus on interventions to reduce the risk of recruitment and radicalisation	Both
9	Christmann, K.	2012	x	x	x	UK, Netherlands, Islamic countries				x			Studies were eligible for inclusion if they had some direct bearing upon either the process of radicalisation or an intervention or programme aiming to prevent radicalisation and extremism. The search strategy used key research questions and sub-questions to supply the initial keywords. Three-part search strategy: electronic searches; personal communications, hand searches. This generated 2,613 citations, with a total of eventually 310 included in the review.	Yes	No			x
10	Davey, J.; Birdwell, J.; Skellett, R.	2018	x	x	x	Global			x				Combining technological automation with human expert analysis; initially identifying over 42,000 potentially at-risk users of pages associated with the far-right, and over 2,000 users of pages associated with Islamist extremism; NLP algorithm to examine user engagement with these pages; selected a sample of 1,600 to be examined; manual verification; assessment against a purpose-built risk matrix. They were then reduced to 814 individuals who were assessed as displaying signs of radicalisation in their online behaviour and, therefore, good candidates for online engagement by the intervention team.	Yes	No			x
11	Frenett, R.; Dow, M.	2015	x			UK, USA						x	80 Facebook profiles identified as at risk of falling into violent Far Right in North America and 80 from the violent Islamist movement in the UK. Intervention providers then engage users in one-on-one messaging.	No	No			x

| Item Nr | Author(s) | Year | Income level and geography | | | | Research Design | | | | | | | Gender | Conflict Sensitivity | Research question focus | | |
			High Income	Middle Income	Low income	Country	Experimental (Randomized Control Trail)	Quasi-experimental	Statistical analysis	Systematic Review	Mixed methods	Qualitative	Add Other RD description here if applicable	Is there any gender analysis or gendered differentiation of outcomes?	Conflict-sensitivity/do no harm (where there is explicit evidence that conflict sensitivity has been considered as part of implementation)	Study focus on methodology or approach used to identify populations at risk	Study focus on interventions to reduce the risk of recruitment and radicalisation	Both
12	Kiendrebeogo, Y.; Ianchovichina, E.	2016	x	x	x	Indonesia, Malaysia, Azerbaijan, Kyrgyzstan, Tajikistan, Algeria, Egypt, Iraq, Lebanon, Qatar, Tunisia, Yemen, Afghanistan, Bangladesh, India, Pakistan, Burkina Faso, Chad, Comoros, Guinea, Mali, Mauritania, Niger, Nigeria, Senegal, Somaliland, Tanzania							A univariate ordered probit (UOP) regression is used to identify the association between attitudes toward terrorist activities and a set of individual characteristics, including age, gender, marital status, employment situation, education attainment, family demographics, importance of religion in one's life, willingness to sacrifice one's life for beliefs, and locality, as well as country-level common factors; drawing on information on attitudes toward extreme violence and other characteristics of 30,787 individuals from 27 developing countries.	Yes	No	x		
13	Fernandez, M.; Asif, M.; Alani, H.	2018	x	x	x	Global			x				Computational approach to measure and predict radicalisation influence using a keyword-based representations of the roots of radicalisation and on a combined lexicon of radical terminology; the approach has been applied to analyse and compare the radicalisation level of 112 pro-ISIS vs. 112 "general" Twitter users. Guided by two research questions: 'How can we translate the different aspects of social theories of radicalisation into computational methods to enable the automatic identification of radicalised behaviour?' And; 'How the incorporation of theoretical perspectives into computational approaches can help us to develop effective radicalisation detection and prediction approaches?'	No	No	x		

Section 3: Summary of methodologies, findings and assessment of quality of studies included in the REA

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
1	Bhui, K; Everitt, B; Jones, E	2014	Other (The association between sympathy for extreme acts and depression, psychosocial adversity and limited social capital)	Cluster analysis of sample into most sympathetic (or most vulnerable), most condemning (most resistant), and a large intermediary group that acted as a reference group.				Those showing the most sympathy for violent protest and terrorism were more likely to report depression and to report religion to be important. Higher social capital and being a migrant were associated with more condemnation of violent protest and terrorism. A larger number of social contacts and being unavailable for work because of housewife roles and disability were also associated with condemnation. Importantly, this study does not show which people who are sympathisers are likely to progress to terrorist acts. The authors argue that it would not be prudent to generalise the findings to other contexts and types of terrorism, although similar research is feasible in other contexts.	Moderate
2	Hung, B. W. K.; Jayasumana, A. P.; Bandara, V. W.	2018	Other	The proposed semantic graph-based approach breaks down into four main steps, (1) extract named entities and their semantic concepts in tweets, (2) build a semantic graph per user representing the concepts and semantic relations extracted from her posted content, (3) apply frequent sub-graph mining on the semantic graphs to capture patterns of semantic relations that discriminatingly characterise the radicalisation stances of users, and lastly (4) use the extracted patterns as features for radicalisation classifier training.				Results show a positive impact on the use of semantic information to identify pro and anti ISIS stances. An additional analysis is performed over the data to identify signals of radicalisation. Results show that pro-ISIS users' discussions tend to mention entities and relations focused on religion, historical events and ethnicity. Anti-ISIS users' discussions tend to focus more on politics, geographical locations, and interventions against ISIS. Anti-ISIS users also tend to mention the entity ISIS with a higher frequency than pro-ISIS users. However, the authors argue that at this point of research and analysis it is not possible to definitively determine a similarity score threshold by which analysts would be alerted for threats and screen for high-risk individuals. This would require data of non-violent radicals who exhibit some of the indicators under study.	High Quality
3	Zierhoffer, D. M.	2014	Other	A threat assessment is defined as "…a set of investigative and operational activities designed to identify, assess, and manage persons who may pose a threat of violence to identifiable targets." It involves the analysis of thoughts and behaviour patterns that result in an attack on a particular target. This approach to assessing the threat of targeted violence is based upon three principles: 1) targeted violence is the culmination of a process of thinking and behaviour that is deliberate and not impulsive; 2) there is interaction among the potential attacker, a past emotional event, a current situation, and a target; and 3) understanding the behaviours of the individual as they progress from the development of the idea to the actual movements toward the target. The level of threat is assessed through ten questions. However, in this study the questions were applied retroactively to three known lone wolf terrorists.				The authors argue that the tool provides a solid foundation for further research into a potential risk assessment model to identify lone terrorists. The strength of the sketched model lies in its focus on behaviours rather than demographic data which may allow police forces with limited budgets to identify those about to act. The weakness of the model is that the observations of behaviour require proximity to the potential perpetrator. It is also worth noting that the study's conclusions are made on data from only three terrorists.	Moderate

Section 3: Summary of methodologies, findings and assessment of quality of studies included in the REA

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
4	Basra, R.; Neumann, P. R.	2016	Repressive actions	Profiling criminal pasts and incarceration histories, biographical information, and involvement in jihadism. To assess individuals' criminal histories, they coded for time spent in prison (as well as the number of stays); criminal convictions; involvement with firearms; types of crime, such as violent (for example, assault, robbery) or petty crime (for example, trespassing, theft), drug dealing, trafficking, and white-collar crime (for example, identity theft or financial fraud).				No uniform profile emerges from the sample, though it is possible to discern patterns. The intensity of criminality varies, from 'one-time' criminals, to repeat offenders, and more sustained 'career criminals'. Given the age of those involved, it is likely that many stood at the beginning of their criminal 'careers'. The vast majority are low-level, local criminals. There are very few that operated on a national or transnational level. Whilst the majority were at some point involved in petty crime (68 percent), the prevalence of violent histories (65 per cent) is notable. Prisons play an important role, with the majority of the individuals in our sample (at least 57 percent) having been incarcerated on at least one occasion. In twelve cases (15 percent of the total, or 27 percent of those who spent time in prison), we are confident that individuals embraced jihadism in prison, though most of them continued (and intensified) their radicalisation after being released. The authors conclude that more attention also needs to be paid to prisons and to countering 'petty' and organised crime (as well as the people engaging in it). Institutional silos – for example, the separation between countering crime and countering terrorism, or between counter-terrorism and 'criminal' police, customs, and other agencies – need to be broken down.	High Quality
5	Moaddel, M.; Karabenick, S. A.	2008	Other	This survey of Egyptian and Saudi youth aims to reveal the extent of reliance on religious authorities as a source of knowledge about the semipolitical role of religion. The paper argues that adherence to religious orthodoxy by young Muslims as well as reliance on religious authorities as a source of knowledge correlates with fundamentalist beliefs and attitudes. It also argues that the psychological state of youth such as fatalism, perceived powerlessness and feelings of insecurity are set of factors that may correlate with religious fundamentalism. The paper then assesses the effects of higher education, watching TV and using the internet as sources of information on fundamentalist attitudes.				The analysis of Islamic Fundamentalism in the world of young Muslims in Egypt and Saudi Arabia revealed that it was linked positively to the epistemic role of religious authorities, Islamic orthodoxy, fatalistic attitudes and the feeling of insecurity in both countries, and negatively to the frequency of watching TV. In Saudi Arabia, where religious institutions have maintained a strong grip on culture and are seen as important authorities, the linkages of these institutions and aspects of religiosity to fundamentalism are much stronger than in Egypt. This may be because in Egypt the authority of traditional religious institutions are fragmented as a result of the development of Western and secular cultures. There were significant gender differences in both countries: young Saudi males reported being less fundamentalist than did Saudi females, whereas Egyptian males are more fundamentalist than Egyptian females. In neither country did university education have any significant effect on fundamentalist attitudes. This suggest that higher learning in Egypt and Saudi Arabia contributes little to the openness of mind that would be conducive to reducing religious-centrist attitudes and enhancing egalitarian orientations toward the followers of other faiths.	High Quality

Section 3: Summary of methodologies, findings and assessment of quality of studies included in the REA

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
6	Skillicorn, D. B.; Leuprecht, C.; Winn, C.	2012	Other	This article subjects some popular theories of radicalisation to empirical testing. The article distinguishes among three types of radical that distinguish themselves by the type of action in which they are engaged: They engage in politically motivated violence ('terrorists'). They engage in non-violent but illegal politically motivated acts ("radicals"). They support individuals or groups who engage in politically motivated violence or other illegal acts ('activists'). The survey instrument was designed to capture both the distribution of attitudes across a population and the distribution of respondents relative to the entire set of questions: why do individuals end up in one of the three categories in the first place? Are there three different kinds of people who end up in these three different categories? Or are there 'stages' along a 'conveyor-belt' through which a given individual passes? What are the drivers of the transition involved? What motivates an individual to cross boundaries, either passing from non-radical to radical, or from one category of radical to another? What are the barriers to these transitions? Why do so few people become radicalised? Why does the number of individuals in categories differ?				The survey finds that radical attitudes appear absent among Muslims with moral and/or social/political satisfaction. The survey suggests that it is when moral dissatisfaction leads to changes in both political and religious attitudes that radicalisation becomes a risk. Programmatic policy responses in the West have generally presumed that improving individuals' life satisfaction, broadly understood as economic opportunity, will decrease the prevalence of radical attitudes. Yet, this study suggests that radical attitudes appear among Muslims that are morally dissatisfied, driving them towards religion and sympathy for terrorist groups. The authors argue that this pathway is present among Muslims because they do not often distinguish between religion and politics. This entails policy responses that focus on moral dissatisfaction, which the authors suggest most Western states would be unwilling to enact.	Moderate Quality
7	Ferrara, E.; Wang, W.; Varol, O.; Flammini, A.; Galstyan, A.	2016	Other	A machine learning framework that leverages a mixture of metadata, network, and temporal features to detect extremist users, and predict content adopters and those likely to interact with extremists. To do this, it carries out three analytical tasks. I: 'Detection of extremist supporters.' The first task consists of a binary classification aimed to detect ISIS accounts and separate them from those of regular users. II: 'Predicting extremist content adoption.' The second task is to randomly sample among followers of ISIS accounts to support the prediction of extremist content adoption. A positive instance of adoption is defined as when a regular user retweets some content s/he is exposed to that is generated by an ISIS account. III: 'Predicting interactions with extremists.' The last task is predicting whether a regular user will engage with this extremists if they are contacted by them.				OF the top 11 most significant online behaviours, the study found that some of them, such as the ratio of retweets to tweets, the average number of hashtags adopted, the sheer number of tweets, and the average number of retweets generated by each user, systematically rank very high in terms of predictive power. These insights shed light on the dynamics of extremist content production as well as some of the network and timing patterns that emerge in these type of online conversation. Based on this evidence, the authors argue in favour of developing computational tools capable of effectively analysing massive social data streams, to detect extremist users, to predict who will become involved in interactions with radicalized users, and to determine who is likely to consume extremist content.	High Quality

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
8	Abdile, M.; Botha, A.	2014	Lack of confidence (economic), Other (religion and proximity to extremist group)	A profile of typical al-Shabaab recruits was developed by interviewing 88 former fighters, plus another seven intelligence personnel off the record.				A profile of a typical recruit was developed. The empirical data showed that the group targets adolescent and young adults (only 9% of interviewees joined after their 30th birthdays, 34% grew up without a father, while 16% grew up without a mother but the majority had both parents present in their lives. The majority were also single (47% but correlation with young age) but marriage and having children did not prevent interviewees from joining al-Shabaab. Only 9% of interviewees indicated that they would remain friends with those they did not agree with, while a further 60% indicated that they would not listen to friends' advice. Interviewees who indicated that they would listen to others (40%), referred to as elders, parents, religious leaders, community members and friends. 40% had no education and 43% religious education. Although the group was roughly split between those that were employed and unemployed when they joined. 64% joined with a friend and roughly two thirds felt a sense of belonging in the group. Whilst 98% believed the government only protects its own interests and 96% that revolt is legitimate.

The recruit profile is not the main focus of the study, drivers of radicalisation are also discussed at length but not included here as it meets the exclusion criterion. | High Quality |

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
9	Christmann, K.	2012	Grievances, Other	Two different psychometric tools were found that offered some relevance to measuring features of radicalisation as a psychological construct. The Revised Religious Fundamentalism Scale and the Violent Extremist Risk Assessment (VERA). The former measures attitudes towards one's religious belief and not an adherence to any particular set of beliefs, using 12 items. Thus, it has the potential for broad applicability. In contrast, the latter is designed to be used with (and limited to) people with either a history of extremist violence or convictions for terrorist-related offences. Consisting of 28 risk factors, it aims to determine whether or not the individual under test has an identified target; whether the violence has an ideological, religious or political motivation; and finally, if the person is acting as part of a group or alone. As such, the VERA appears applicable only to that small cohort that is in (or near) an operational phase and hence, already radicalised and, in some countries, breaking the law.	Education, Religious education, Identity, Counter-narrative	The review only found two studies of programmes (outreach and engagement projects in London) that aimed to address Islamic radicalisation in the UK. It also found a teaching resource pack. To make up for this, the review drew heavily upon the DCLG 2010 rapid evidence assessment of community interventions to prevent support for violent extremism assessing. It assessed which interventions work best in relation to tackling extremism, particularly extremism in the name of religion. It covered interventions that use education or training to build the capacity of and empower young people, often with a view to challenging extremist ideology through a focus on debating theology. This included some interventions that empower women to debate and discuss theological issues.	Positive	The review's authors argue that as the Revised Religious Fundamentalism Scale has only been extensively piloted in the United States among undergraduate university students and their parents, it should not yet be considered ready for identifying those at risk elsewhere. Furthermore, they argue that the utility of VERA remains unclear as it is only intended for those individuals already deemed merely 'at risk' of radicalisation and subject to interventions. Overall, they conclude that the field its yet to develop a general causal model or theory of the structural causes of Islamic terrorism and that there is a general disagreement as to whether such a model is achievable. More prevalent were studies that listed several possible factors, usually social-psychological models, but failed to specify the interactions between the listed factors in any detail. The DCLG review clearly advocated the adoption of interventions focussed on capacity building and empowering young people, and those that challenge ideology through a focus on theology and the use of education/training. These interventions were argued to be most successful when delivered through outreach work that focused attention directly towards the relevant communities. Otherwise, the main learning points from studies of interventions outside the UK concern the need for those engaging with radicalised individuals to carry authority and legitimacy, and to be equipped with profound ideological knowledge.	High Quality

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
10	Davey, J.; Birdwell, J.; Skellett, R.	2018	Other	A multi-stage identification methodology and process combining technological automation with human expert analysis. It initially identified over 40,000 potentially at-risk users of pages associated with the far-right, and over 2,000 users of pages associated with Islamist extremism. NLP algorithms were then used to examine user engagements with these pages, detecting instances of language that appeared to be violent, aggressive and dehumanising, or content keywords relevant to extremism. This distilled the total pool of users identified and allowed for the selection of a sample of 1,600 to be examined in further detail. Rankings and scores that used measures for the number of extremist-related pages the user posted on, the frequency of posts, and the extent to which their comments included violent or dehumanising language were given. The study also made use of the manual verification of users and construction of profiles, and assessed individuals against a purpose-built risk matrix.	Counter-narratives	Direct engagement with radicalising individuals by mentors and 'intervention providers'; about 70% of candidates selected for online intervention were engaged by intervention providers, who initiated conversations through Facebook Messenger.	Positive, Neural	Engagements were assessed for positive impact by their length and qualitatively analysed to 'detect instances which suggested that the conversation had possibly generated positive impact by presenting a candidate with an alternative point of view.' Much of this focussed on the internalisation of counter arguments, with the assumption that it indicates empathy and the diversification of perspectives. A expressed desire to continue engagements with the programme's implementers was also seen as an indicator of impact. Three metrics were ultimately used to consider the impact of online outreach: initial response rates, sustained engagements (conversations that included five or more messages between the candidate and intervention provider), and indications of potential positive impact during the course of the conversations. About 20% of candidates responded. Islamist candidates were more likely to respond (26%) than extreme right candidates (16%). Sustained engagement rates were achieved with 71% Islamist candidates and 64% extreme right candidates. In 10% of the sustained conversations it was suggested the programme had a positive impact. Limitations of identification methodology: The reliance on users interactions with 'seed pages' – public and open Facebook pages – may have skewed results. The authors also add that 'It is important to set clear expectations for what online engagements can achieve at the outset. In the same way that an individual's journey to radicalisation is not the sole product of online influence, their exit and diversion will likely require more than just an online conversation. While possible, it is unlikely that an online conversation alone will lead to significant and measurable disengagement and de-radicalisation.'	Moderate Quality

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
11	Frenett, R.; Dow, M.	2015	Other	Facebook pages liked and Facebook group membership were found to be the most relevant factors to indicate risk; additional markers: a user's cover photo and the tone and content of their regular posts. The project team collated the risk factors provided by the outreach providers and created two distinct sets of search criteria - one for those considered at risk of falling into the orbit of the violent far right extremism in North America and another for those considered at risk of falling into violent Islamism in the United Kingdom.	Counter-narrative	Testing the viability of an approach based on directly messaging those openly expressing extremist sentiment online and seeking to dissuade them from following that path. Working on the assumption that former extremists would be the most credible messengers, 10 former extremists (five former far-right from North America and five former Islamist extremists from the UK) were identified, one woman and four men in each ideological grouping. In total 154 profiles were identified; once identified profiles were passed to "intervention providers" to verify; over 90% of the profiles identified were confirmed as being 'at risk'.	Positive, Neutral	Of the 154 profiles confirmed to be at risk, 76 messages were drafted and sent to candidates (a number of factors account for a large difference). Over 60% of the messages were actually seen by the candidate – a read rate significantly higher than standard unsolicited mailing campaigns; 59% responded directly or demonstrated a shift in behaviour (change in privacy settings or blocking the provider); a majority were willing to engage in conversation (60% sustained engagement, 8% brief engagement, 12% denial of involvement). The data set is considered by the authors to be too small to build policy upon. The time period is also too short to effectively measure any long-term shifts in belief system or behaviour. A number of interactions gave indications that achieving a long-term adjustment in behaviour may be possible. This was evident when certain candidates [no percentage indicated] stated they felt they could trust their outreach provider and that the interaction led them 'to think deeply' for the first time. Accordingly, they invite others to replicate and improve on their methods.	Moderate Quality

Item Nr	Author(s)	Year	Typology of Key Themes of Identification	Identification methodology types	Typology of Intervention	Intervention types	Effects of Intervention	Outcome/indicator types	Quality Appraisal
			From the list in the coding guidelines: key theme/s most focussed on by the methodology or approach to identification	Please insert some details about the key methodologies or approaches used to identify populations at risk	From the list in the coding guidelines please indicate which type/s of intervention the document most pertains to (list more than 1 if needed)	Please insert some details about the key intervention types from the document text	Where explicitly stated, whether interventions have had positive, neutral or negative effects (more than one option can be selected).	Please insert some details about the key outcome/indicator types	According to the quality appraisal methodology, is the study of low, moderate or high quality (choose one)
12	Kiendrebeogo, Y.; Ianchovichina, E.	2016	Education, Lack of confidence, Other	The study looks for common characteristics among radicalized individuals, defined as those willing to justify the targeting and killing of innocent civilians. These individuals hold extreme views and represent the group of potential terrorism supporters, namely those who may sympathize with and collaborate with extremists. They may be at a higher risk than others to eventually become terror recruits, although they may not be involved directly in terrorist activities. The study analyses the association between attitudes toward terrorist activities and a set of individual characteristics, including age, gender, marital status, employment situation, education attainment, family demographics, importance of religion in one's life, willingness to sacrifice one's life for beliefs, and locality, as well as country-level common factors. The analysis is based on the information on attitudes from the Gallup World Poll data set for 2006-12.				The results suggest that the typical extremist who supports attacks against civilians is more likely to be young (under 33), unemployed and struggling to make ends meet, relatively uneducated, and not as religious as others, but more willing to sacrifice their own life for his or her beliefs, and that they tend to live in rural areas. Gender and marital status are not found to explain significantly the individual-level variation in attitudes toward extremism. Although these results may vary in magnitude and significance across countries and geographic regions, they are robust to various sensitivity analyses. Contrary to common perceptions, extremely radicalized people and those who mildly support attacks on civilians tend to be less religious (20%) than other categories of respondents (75%). These results are robust to various estimation techniques, approaches for dealing with non-responses to the radicalization question, and additional controls. They also hold for a global sample that includes both non-OECD and a few OECD countries and across geographic regions, although the effects may vary in magnitude and significance across countries and regions. As repeated cross-section data was used, the study does not deal with individual-level unobserved heterogeneity. Thus, the results should be interpreted as partial correlations. The researchers argued that to gain a better understanding of the causes behind radicalization, there is a need to improve data collection, particularly of panel data.	High Quality
13	Fernandez, M.; Asif, M.; Alani, H.	2018	Other	A computational approach for detecting and predicting the radicalisation influence a user is exposed to, grounded on the notion of 'roots of radicalisation' from social science models. The approach is based on Natural Language Processing (NLP) and Collaborative Filtering (CF), that automatically captures the different roots of radicalisation (micro - individual, meso - social and macro - global) for each user and represents them as keyword-based vector descriptions. The approach, whose goal is to automatically detect and predict the level of radicalisation influence a user is subjected to, does not aim to determine whether someone is being radicalised or not, but to provide a risk level for each user based on the individual, social and global influences to which she is exposed to in social media.				The results show the effectiveness of the proposed algorithms in detecting and predicting radicalisation influence, obtaining up to 0.9 F-1 measure for detection and between 0.7 and 0.8 precision for prediction. As expected, individual and social influences of radicalisation are both higher for pro-ISIS users (the global influence was discarded from the rest of analysis when it was discovered that 63% of those URLs were closed). The Mean Absolute Error (MAE) value is low in all cases. A low value of MAE indicates the effectiveness of the models, since it assesses the mean of the absolute differences between the ratings and the predicted values. A self-identified weakness of the study's approach concerns the difficulty of accounting for social network structures within which extremist online language is employed, such as further interactions such as the 'likes', 'replies' or even 'direct messages' that Twitter uses often use to signal their agreement or create relationships. The authors argue that a fine grained temporal analysis, following users' tweets over a period, would better capture the radicalisation process.	High Quality