



Evaluating Programmes to Prevent and Counter Extremism

FULL REPORT
NOVEMBER 2020

James Lewis
Dr Sarah Marsden
Dr Simon Copeland

Evaluating Programmes to Prevent and Counter Extremism

FULL REPORT

James Lewis, Lancaster University
Dr Sarah Marsden, Lancaster University
Dr Simon Copeland, Lancaster University

This report is one of a series exploring *Knowledge Management Across the Four Counter-Terrorism 'Ps'*. The project looks at areas of policy and practice that fall within the four pillars of CONTEST. For more information visit:

www.crestresearch.ac.uk/projects/the-four-counter-terrorism-ps

With grateful thanks to Prof Adrian Cherney for his contribution as expert academic advisor.

This research was funded by the Centre for Research and Evidence on Security Threats - an independent Centre commissioned by the Economic and Social Research Council (ESRC Award: ES/N009614/1) and which is funded in part by the UK Security and intelligence agencies and Home Office.

www.crestresearch.ac.uk



TABLE OF CONTENTS

KEY POINTS	4
INTRODUCTION	5
INTRODUCING P/CVE PROGRAMMES	6
EVALUATING P/CVE INTERVENTIONS	7
CHALLENGES FACING P/CVE EVALUATIONS	9
PRACTICAL CHALLENGES.....	14
RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES	16
ASSESSING THE EVIDENCE BASE	25
READ MORE	26

KEY POINTS

In contrast to comparable areas, such as gang-related interventions, programmes for preventing or countering violent extremism (P/CVE) have neither a strong knowledge base nor established standards for evaluating their impact. Research has instead sought to draw international comparisons across interventions and identify transferable lessons from other fields.

However, very little research has systematically tested or applied these approaches to P/CVE programmes. Key challenges facing researchers in this area are only just beginning to be addressed.

- The overwhelming majority of P/CVE programmes have not been subject to formal evaluation. Where evaluations have taken place, they can fall short of the standards of transparency, independence and rigour typical of related fields.
- Greater sharing of internal evaluations would strengthen the field and enhance the speed at which progress is made, as well as avoiding parallel research cultures developing in the open and closed source literature. Pooling expertise on evaluation methods across the different bodies involved in delivering P/CVE interventions, and more publicly accessible evaluations would also strengthen work in this area.
- Key challenges facing P/CVE interventions include the absence of an appropriate counterfactual, or an understanding of what would have happened in the absence of an intervention, and the small numbers of people who are supported through these programmes. Quasi-experimental designs have been used in comparable fields such as gang-related interventions and have the potential to overcome these challenges.
- There are ethical and security challenges when selecting an appropriate control group against which to evaluate the impact of P/CVE programmes. It is relatively straightforward to identify a control group for primary prevention methods that are aimed at larger populations. It is far harder to generate control groups for those at risk, or involved in extremism, as this would

typically involve denying individuals access to support to determine if an intervention was effective. Switching-replications designs can potentially overcome this issue.

- There is an absence of robust data against which to triangulate the findings of P/CVE evaluations. However, lessons can be learned from evaluations of gang-related interventions, which commonly use more than one evaluation method to triangulate their findings.

INTRODUCTION

The evidence base on how to effectively evaluate P/CVE programmes is extremely limited. Several recent studies outline principles of best practice based on literature from other fields. However, there are few examples of how these principles have been used to evaluate P/CVE programmes in practice.

A wide range of evaluation methodologies have been discussed in the literature on P/CVE interventions, but their utility is poorly understood. Although several candidate approaches have been identified based on their use in comparable fields, little is known about whether they work in the context of P/CVE. This report should therefore be considered exploratory. Much more research is needed into how evaluation methods from other fields can be adapted to P/CVE. Programme designers can learn lessons from comparable areas of work, but it is important not to overstate the similarity between extremism and other issues. Any evaluation must attend to the unique range of individual, social and structural issues that can contribute to the journey into and out of extremism.

This report provides an overview of the types of P/CVE programmes that have been developed; reviews the methods used to evaluate them; and outlines the challenges facing evaluation efforts alongside a review of how research has sought to overcome them. The most significant limitation is the lack of evaluative work carried out to date. Most research is descriptive, and although there is a good understanding of the challenges facing the field, few studies have successfully addressed them.¹

This report is primarily based on academic literature published from 2017 onwards. Due to the limitations of this research it also draws on literature from outside this period, grey literature, and work from comparable fields, such as evaluations of medical and gang-related interventions. Much P/CVE research focuses on international efforts, including attempts to draw

comparisons between countries. Countries that are the focus of more in-depth work include the UK, Australia and the Netherlands.

¹ The lack of a strong evidence base underpinning P/CVE interventions is highlighted by *Feddes & Gallucci (2015)* and *Schuurman (2018)*. The most recent review of literature on P/CVE programmes found that only 38 of 112 publications included clear outcomes, and within these, anecdotal analyses were the most common means of evaluation (*Pistone et al., 2019*). Of 107 studies and reports, an earlier paper identified 24 studies that included some correlational findings from P/CVE interventions and non-experimental designs (*Mastroe & Szmania, 2016*).

INTRODUCING P/CVE PROGRAMMES ²

P/CVE programmes are extremely diverse, targeting a wide range of policy settings through a variety of types of intervention.

Programmes to prevent or counter violent extremism (P/CVE) encompass a range of initiatives that aim to address the drivers of violent extremism and the factors that enable it. There is significant diversity in P/CVE interventions. They work with diverse client groups and operate at different scales; from long-term development work through to short courses that last a few hours. Although delivered across the world, a small number of countries have produced the greatest amount of evaluation-oriented research, foremost among which is the UK, then Germany, the Netherlands and the USA.

Interventions are commonly described using a public health model of prevention. Primary prevention seeks to tackle the ‘root causes’ of extremism; secondary prevention is directed at the processes of radicalisation and those considered at risk of involvement in extremism; while tertiary prevention is targeted at those already engaged in extremism. Within these broad tiers, specific intervention programmes fall into four categories:

- 1. Empowerment or resilience** Programmes that aim to develop knowledge, skills or change participants’ attitudes, including community forums, prejudice reduction projects, youth work, critical thinking programmes, and human rights-based education.
- 2. Policy programmes** Led by government, statutory bodies or larger-scale community bodies, these initiatives include PREVENT, citizenship education, community policing projects or community reporting programmes.

Interventions operate at different levels and target a range of populations:

Individual – targeting individuals

Relational – concerned with increasing the ability of peers or gatekeepers to provide support

Group – seeking to influence community-based systems or policies

National – policies focused on the economy, education, social welfare and health

Global – relating to global political or religious initiatives

3. Deradicalisation

Interventions working with those already involved in extremism aiming to rehabilitate and reintegrate.

4. Counter-narratives

Projects offering an alternative narrative to those disseminated by extremist groups delivered online or in person.

Through these interventions, P/CVE initiatives aim to achieve a range of goals including to procure information and intelligence about possible threats; prevent or disrupt violent extremism and terrorist attacks; protect people, critical infrastructure and property; prepare for attacks and their consequences; and pursue those suspected of involvement in violent extremism.

² This discussion draws on *Pistone et al.’s (2019)* review of the literature which categorises intervention types and different kinds of programmes. Other work categorising P/CVE interventions includes *Stephens et al. (2019)*

EVALUATING P/CVE INTERVENTIONS ³

P/CVE evaluations are characterised by weak research designs and internal evaluations, and most take a largely descriptive approach.

Despite the increasing use of P/CVE interventions globally, their impact, effectiveness, and economic return is often unclear. The overwhelming majority of P/CVE programmes have not been formally evaluated. Where evaluations have taken place, internal self-evaluation has been the principal form of assessment. While potentially valuable, internal evaluations are often perceived to lack independence and can result in work that describes a programme rather than offering a robust assessment of its impact.

Reviews of P/CVE evaluation suggest their number and quality are increasing. However, few programmes employ robust methods to interpret impact and outcomes. The quality of CVE evaluation has been persistently weak; a systematic review of work published before 2014 assessed none to be high quality and only 37 per cent as medium quality.

In part due to the challenges facing evaluation in this field, evaluations of P/CVE interventions have often adopted weaker research designs. Only 12 per cent of the 135 participant samples in 55 studies published before 2014 drew on empirical data. The majority of these employed a relatively weak research approach, typically using a cross-sectional design with a single instrument. Of 48 intervention evaluations released between 2013 and 2018, 33 had no clear evaluation approach.

Evaluations with a recognisable research design have adopted a number of approaches which are described

below alongside examples from different types of P/CVE programme:

ADDITIONALITY ⁴

- This typically involves taking a single measurement to understand what happens once an intervention has been carried out. Because this research design does not take baseline measures or use counterfactuals, it cannot easily determine if an intervention caused an observed outcome. Additionality-focused studies are the most prevalent form of evaluation approach in P/CVE. Their diversity means that a wide range of constructs is assessed across most types of programmes, examples include:
 - a. **Empowerment/resilience:** knowledge about terrorism and radicalisation, or the impact of mentors on reducing risk of involvement in extremism
 - b. **Policy programmes:** role of police training in improving reporting rates and extent of community engagement, or changing levels of social inclusion
 - c. **Deradicalisation:** recidivism rates
 - d. **Counter-narratives:** perceptions of costs and benefits of involvement in extremism.

QUASI-EXPERIMENTAL DESIGNS ⁵

Involving different groups, where the allocation of participants to groups is not random. There are a number of types of design nested under this approach

³ This section draws on research exploring specific international and UK-based P/CVE interventions as well as review articles of international P/CVE efforts published between 2013 and 2020. For a review of methods, research designs, and a list of indicators that could be used in P/CVE evaluations, see *Cherney et al. (2018)*. *Impact Europe* run a database of data collection and evaluation designs. The systematic review of work published before 2014 was conducted by *van Hemert et al. (2014)*; the review of 55 studies by *Feddes & Gallucci (2015)*; and *Bellasio et al. (2018)* reviewed evaluations released between 2013 and 2018.

⁴ For a review of different P/CVE constructs used in additionality-focused studies, see *Brett & Kahlmeyer (2017)*; *Broadbent (2013)*; *Schuurman & Bakker (2016)*; and *Berman (2019)*, who focuses on counter-narratives.

⁵ *Mazerolle et al. (2020)* set out an approach for determining the relative strengths of quasi-experimental designs. Several studies explore the use of value complexity-based approaches for assessing the impact of P/CVE interventions on levels of empowerment and resilience in the UK (*Liht & Savage, 2013*) and international contexts (*Savage et al., 2014*; *Holmer & Bauman, 2018*; *Aly et al., 2014*; *Marrone et al., 2020*; *Boyd-MacMillan, 2016*; *Kurtz et al., 2016*). Authors such as *Williams et al. (2016)*; *Aldrich (2014)*; and *Mitts (2017)* have outlined quasi-experimental designs for policy programmes, and *Webber et al. (2018)* do the same for deradicalisation.

EVALUATING P/CVE INTERVENTIONS

Evaluating Programmes to Prevent and Counter Extremism

which can involve testing the same group before and after the intervention; cross-sectional comparisons of control and experimental groups; and a combination of before and after testing between groups that receive the intervention and those which do not (difference-in-differences method) – there are also different approaches to matching participants across groups. A diverse range of outcomes is assessed across different types of P/CVE initiative, including:

- a. **Empowerment/resilience:** levels of integrative complexity, or an individual's ability to reason and think in a way that incorporates multiple, differing perspectives and tolerate ambiguity; moral disengagement; attitudes to reintegrating at-risk youth and the value of diversity; awareness of radicalisation risk factors and willingness to respond; employability and social status
- b. **Policy programmes:** capacity to build community-statutory body collaborations; whether there has been a reduction in the extent of online extremist material; civic behaviour, attitudes towards the West and cultural identities
- c. **Deradicalisation:** attitudes towards extremist ideology, legitimacy of political violence, and changing levels of personal significance.

LONGITUDINAL ⁶

Studies assessing change over a longer period of time are less common but have been used across different intervention types, including:

- a. **Empowerment/resilience:** seeking to identify changes in levels of prejudice; attitudes toward ideologically motivated violence, self-esteem, empathy and perspective-taking
- b. **Deradicalisation:** assessing the extent of integration among terrorism offenders.

RANDOMISED CONTROL TRIAL (RCT) ⁷

Very few studies have adopted an RCT model which involves the random allocation of participants to different conditions. One of the only studies to use this type of method undertook a randomised controlled effect evaluation to assess the impact of a primary intervention using former extremists. The programme sought to change young people's attitudes about the legitimacy of political violence, political tolerance, self-efficacy, awareness of extremist recruitment tactics, and confidence responding to extremism. RCTs for secondary or tertiary interventions are lacking.

Within these, a wide range of methods has been deployed including interviews, focus groups, desk-based research, network analysis, case studies, data mining, data informatics, ethnography, surveys, and stakeholder analysis. Such methods are used in the context of a number of analytical approaches including studies which use a theory of change, realistic evaluation, Multi-Attribute Utility Technology, and contribution analysis. Evaluations have also focused on economic factors including cost-benefit analyses and cost-effectiveness assessments.

⁶ For longitudinal approaches, see *Savoia et al. (2019)*; *Feddes et al. (2015)*; and *Cherney & Belton (2019a)*.

⁷ While RCTs are often seen as the 'gold standard' research design, non-experimental and quasi-experimental designs might produce useful data in some circumstances. As *Baruch et al. (2018)* note, "Methodologies for CVE evaluations do not fit sensibly into a simple hierarchy capped by randomised control trials." The RCT referenced here was conducted by *Parker and Lindkilde (2020)* who randomly assigned 2,156 individuals into a treatment and control group. Half (control) were asked attitudinal questions before the workshop to establish a baseline, and the other half (treatment) were asked these questions after the workshop had ended. The only other RCT identified was an evaluation of a training and economic incentive programme for young people in Afghanistan (*Kurtz et al., 2018*).

CHALLENGES FACING P/CVE EVALUATIONS ⁸

KEY POINTS

Evaluations of P/CVE interventions face analytical and practical challenges:

- Key analytical challenges relate to the variety of individual and contextual factors that impact engagement and disengagement processes; challenges in identifying appropriate counterfactuals; difficulties in defining the specific outcomes that an intervention can feasibly seek to achieve; and the absence of agreed metrics for evaluating success.
- Key practical challenges relate to the logistical, ethical and security challenges that researchers face accessing data, and identifying appropriate control groups; challenges in constructing baselines against which to measure the impact of interventions; and the absence of robust administrative data (e.g. official statistics on success rates) against which to triangulate the results of evaluations.

The literature outlines a series of potential methods for addressing some of these challenges, such as using mixed-methods designs to triangulate data across different methods, or making better use of quasi-experimental designs. However, some of these challenges, particularly relating to the construction of counterfactuals and the selection of control groups, are much more difficult to address.

A series of widely acknowledged challenges face efforts to evaluate P/CVE programmes. These comprise analytical and practical issues and affect

Introducing evidence-based approaches and an evaluation culture [to P/CVE] is in many senses more akin to a social movement than to a scientist with a narrow set of tools and techniques, in that it involves the creation of a collective identity, shared norms, a concern for changing the status quo, practical actions and the creation of organisational capacity.

(Baruch et al. 2018)

most P/CVE initiatives including empowerment and resilience programmes, policy work, deradicalisation initiatives, and counter-narrative work. Part of the challenge in addressing these issues is the lack of an evaluation culture within P/CVE that supports evidence-based approaches.

ANALYTICAL CHALLENGES

This section outlines specific analytical challenges relating to the complexity of engagement and disengagement in violent extremism and the selection of appropriate evaluation metrics. It also presents potential solutions to these challenges drawn from existing research. Two fundamental and more difficult analytical challenges relating to identifying counterfactuals and the use of quasi-experimental designs are explored in greater depth in a later section.

THE COMPLEXITY OF VIOLENT EXTREMISM ⁹

P/CVE programmes have been used across a wide variety of cultural, social, and political contexts. Different settings have contrasting understandings

⁸ Challenges facing evaluations of P/CVE are highlighted by several studies (e.g. *Feddes & Gallucci, 2015; Khalil & Zeuthen, 2016; Baruch et al., 2018; Koehler, 2017; Holmer & Bauman, 2018; Koehler & Fiebig, 2019*).

⁹ *Ris & Ernstorfer (2017)* discuss the different levels and scales of P/CVE interventions and the importance of taking account of varied and complex socio-political factors when evaluating these programmes by applying lessons learned from the monitoring and evaluation of peacebuilding interventions to this field. *Baruch et al. (2018)* discuss the use of public health models to identify 'tracers'. *Gielen (2018)* discusses the utility of realistic evaluation. For a worked example of a realistic evaluation of a P/CVE intervention, see *Gielen & Dijkman (2019)*. *Khalil & Zeuthen (2016)* discuss the importance of avoiding spillover between P/CVE and other workstreams, while a report produced for DFID by *Kelly (2019)* also outlines the limits to any crossover between CVE and humanitarian work in the context of the conflict in Syria and Iraq, and some of the risks in linking these agendas.

CHALLENGES FACING P/CVE EVALUATIONS

Evaluating Programmes to Prevent and Counter Extremism

of concepts and approaches to P/CVE, which makes it difficult to develop evaluation approaches that are cross-culturally applicable. All interventions should be tailored to their context, adapting good practice and considering the potential unintended consequences that might emerge from transposing interventions across contexts.

The multiple pathways towards extremism and the range of individual, social, and political factors that might be relevant in individual cases make it difficult to narrow down the specific causes that P/CVE interventions seek to address. P/CVE interventions can be used to affect structural, individual, or community-level drivers of violent extremism. Programme designers and evaluators must therefore be clear about the specific issues that an intervention is seeking to address, and what the intervention can realistically achieve, in order to understand what success looks like. Different pathways into extremism have different indicators of risk that must be accounted for when designing interventions or evaluations.

- One way to take account of different drivers and trajectories would be to draw on public health evaluations that seek to identify proximal

attributable risks; causes of risks; and causes of causes. For P/CVE this could involve obtaining weapons; engaging with radical websites; and having low perceived power. These metrics could be used as ‘tracers’ to evaluate impact.

- Alternatively, realistic evaluation could be used to explore the contextual and individual factors that contribute to the success, or failure, of a P/CVE intervention (see table below).

P/CVE interventions must not be decontextualised from the wider socio-political context in which they operate, and from socio-political factors that contribute to violent extremism. Interventions must take account of socio-political contextual factors which can affect whether and how programmes work:

- Research which applies lessons learned from peacebuilding interventions to the field of P/CVE outlines the need for interventions to take account of structural factors that might contribute to involvement in violent extremism, such as unemployment, discrimination and inequality.
- Recognising the role that these broader factors play in driving engagement in violent extremism, and

REALISTIC EVALUATION OF A DUTCH EXIT PROGRAMME

Gielen (2018) sets out the process of using realistic evaluation to understand ‘what works, for whom, how, and in what circumstances’ in an exit programme for female jihadists. This approach recognises the different types of individualised and contextualised drivers of radicalisation, and involves:

1. Building a theoretical model about what might work, for whom, and how, based on a review of existing P/CVE or related interventions. This ‘C-M-O model’ includes Contextual and individual factors that might contribute to success, Mechanisms that can support the journey out of violent extremism, and specific intervention Outcomes.
2. Conducting multi-method data collection with relevant stakeholders.
3. Analysing the data to explore how the theoretical model relates to individual cases asking: ‘Which interventions and underlying mechanisms led to the successful or unsuccessful disengagement of this particular young female jihadist? What contextual factors made it possible to activate those mechanisms? To what extent did the outcomes fit or differ from the theoretical model developed during the first phase of the realistic evaluation?’
4. Developing a refined programme theory based on the data analysis which could be used to develop additional hypotheses for further testing and future validation.

how the failure to address them might inhibit the success of interventions is crucial. Doing so helps to build theories of change that accurately reflect the complexity of violent extremism; understand the limits of individual interventions in addressing the diverse range of factors that contribute to violent extremism; and develop a more holistic approach to tackling the underlying drivers of this phenomenon.

There is a lack of clarity around what types of intervention are CVE-relevant. It is important to avoid P/CVE interventions that spill over into development work or other analogous policy areas that are not directly relevant to violent extremism, particularly when trying to address structural drivers of violence.

EVALUATION METRICS AND OUTCOMES ¹⁰

There are no agreed metrics of success for P/CVE interventions. Programme designers and evaluators need to consider which behavioural and/or attitudinal metrics are most appropriate for a specific intervention, and how to interpret chosen indicators of success. This process is hindered by a lack of commonly agreed methods for recording and measuring behavioural changes, while assessing emotional or attitudinal change remains difficult because the evidence of this kind of impact may not be immediately obvious:

- Medical interventions use patient-centred approaches for evaluating impact, whereby they assess the impact of an intervention at the individual patient level by exploring functional, social, psychological, and emotional outcomes. A transferrable approach for P/CVE might be to explore whether participants gained a better sense of their identity, or in the case of convicted extremists a better understanding of their offence, through engaging with an intervention.

A diverse range of constructs could be used to measure

the impact of P/CVE interventions, but they have yet to be adequately defined or formalised. More attention should be paid to developing construct validity so there is a recognised way of defining and assessing key constructs, such as support for violent extremism; radicalisation risk; deradicalisation; and disengagement.

It is difficult to develop constructs to measure the risk of an individual becoming involved in violent extremism, and thus for evaluating whether a primary intervention has reduced this risk. Identifying and measuring attitudes that contribute to the risk of involvement in extremism is challenging given the low numbers involved. Several studies have suggested using proxy measures, such as employment or coping skills, but the link between these metrics and specific P/CVE-relevant outcomes is often unclear.

¹⁰ *Baruch et al. (2018)* discuss the importance of construct validity in this context, while both *Koehler (2017)* and *Cherney (2020)* discuss the distinction between deradicalisation and disengagement outcomes. Koehler also discusses the issues with using recidivism rates as a metric for measuring these outcomes. For a practical example of CFA, see *McGill et al. (2015)*. For the use of CFA to measure attitudes towards violent extremism, see *Trip et al. (2019)*, *Bhui et al. (2020)* and *Fodeman (2020)*. Several studies discuss proxy measures and existing measurement frameworks that they believe could be adapted to evaluate P/CVE interventions, including *Cherney et al. (2018)* and *Baruch et al. (2018)*. The latter study discusses the applicability of patient-centred approaches to P/CVE, while the description of a patient-centred research design presented here is drawn from *Carr (1996)*. The importance of testing and validating Theories of Change is discussed by *Ris & Ernstorfer (2017)*. *Horgan & Braddock (2010)* outline the potential benefits of MAUT as an evaluation tool for deradicalisation programmes, and *Marsden (2015)* uses this approach to analyse interviews with Counter-Extremism Unit probation officers.

CHALLENGES FACING P/CVE EVALUATIONS

Evaluating Programmes to Prevent and Counter Extremism

Table 1: Identifying intervention outcomes and metrics

Cherney et al. (2018) provide an overview of different P/CVE objectives, outcomes, and potential indicators for measuring success against these metrics. The table below breaks down examples of different outcomes and indicators that they identify, and which could be used to measure the impact of primary and secondary interventions.

Overall objective	Outcome	Example indicators
1. Communities and individuals are resilient to violent extremism.	1.1. Individual	Critical thinking skills; Coping skills; Sense of belonging; Self-efficacy; Strong cultural identity and openness to other sources of belonging; Wellbeing; Social participation; Social, problem-solving and conflict resolution skills.
	1.2. Environment	Civic participation; Opportunities for education, training and employment; Engagement between communities and government; Sense of marginalisation; Experience of discrimination; Supportive social networks in community.
	1.3. Communities	Awareness and understanding of violent extremism; Trust in government; Perceived community safety; Social cohesion; Perception of community harmony; Inter-communal tensions; Positive perception of country as-a-whole; Sense of national identity; Community capacity and willingness to respond to crisis.
	1.4. Ideologies	Recall of CVE-related media campaigns; Exposure to extremist messaging; Media discussion of inter-communal relations; Media presentation of racial/ other stereotypes.
	1.5. Recruitment	Intentions of joining violent extremist group; Estimated current membership of extremist organisations.
2. Individuals at risk of becoming violent extremists divert and do not engage in violence.	2.1. Identification	Community awareness of violent extremism; Willingness to report suspicious behaviour and to voice concerns.
	2.2. Community-led support	Community awareness of violent extremism; Willingness to challenge radical extremist views; Willingness to support diversity in the community; Community awareness of government CVE initiatives and programmes.
	2.3. Government-led support	Number and coverage of services providing diversion-related services; Willingness among community members/ organisations to refer to government CVE interventions.
	2.4. Government-led intervention/diversion	Participants' commitment to participation in programme; Evaluation of service providers delivering intervention services and their capabilities; Number of CVE initiatives.

There is no agreed approach for defining or measuring deradicalisation or disengagement, which makes evaluating the impact of more targeted secondary or tertiary interventions challenging. For example, recidivism rates may seem like a useful metric for assessing the success of deradicalisation and disengagement programmes, but these figures are difficult to interpret given that base rates of reoffending are not yet fully understood.

- Several studies have explored the use of Confirmatory Factor Analysis (CFA) to test relationships between constructs and those behaviours or attitudes that are seen as contributing to them. CFA has increasingly been used to test hypothesised links between specific behaviours and attitudes and support for violent extremism in a variety of different contexts. Similar approaches could potentially be applied to study the impact of interventions on constructs such as

support for violent extremism; radicalisation risk; deradicalisation; or disengagement.

- Developing a robust theory of change is an important step in identifying intervention objectives and associated metrics. Each P/CVE relevant measure is explicitly linked to how the intervention aims to bring about change. Theories of change need to be tested and refined through ongoing evaluation and by learning from relevant research that has explored the drivers of violent extremism.

Individual outcomes can also be broken down using Multi-Attribute Utility Technology (MAUT). This also involves linking specific metrics to intervention objectives, but a key strength of this approach is that it can be used to compare how well different interventions score on individual attributes (metrics), and thus to understand the relative effectiveness of different approaches.

USING MULTI-ATTRIBUTE UTILITY TECHNOLOGY (MAUT) TO EVALUATE P/CVE INTERVENTIONS

Horgan and Braddock (2010) and Marsden (2015) outline how MAUT could be used to evaluate the impact of tertiary P/CVE interventions by proceeding through a series of steps:

1. The objects to be evaluated, and the functions of the evaluation, are defined. Theoretically, MAUT would be used to (a) monitor the performance of tertiary interventions; (b) identify any problems associated with interventions; and (c) select which attributes are most important for building an effective intervention.
2. Relevant stakeholders are identified. This might include government officials or influential citizens or organisations who have the power to shape the programme design.
3. Relevant attributes relating to the intervention are identified by stakeholders.
4. The relative importance of each attribute is assessed, before attributes are organised into a hierarchical 'value tree' through which specific attributes are grouped into overarching constructs that represent specific intervention aims and objectives.

PRACTICAL CHALLENGES

This section outlines practical challenges in evaluating P/CVE interventions relating to establishing base rates against which to measure intervention outcomes and the challenges in coordinating activities across different policy contexts. Where possible, it also discusses potential solutions drawing on relevant research. Two of the most significant practical challenges, namely the use of control groups, and the challenges of triangulating research data, are explored in more depth in the final section of this report as these challenges have yet to be adequately addressed by existing studies.

an intervention might be assessed to work. The extent to which a particular attitude or behaviour can be used to interpret risk is linked to how prevalent it is among the wider population. The challenge of identifying risk factors is particularly acute for primary and secondary interventions working with those at risk of involvement in extremism. Here, the threshold for intervention, and the baseline for behaviour which indicates that an individual is at risk, does not benefit from a strong evidence base. More research is needed to understand the construct validity of those factors that might indicate risk (see box below).

ESTABLISHING BASE RATES ¹¹

Base rates, or the frequency with which specific factors are found in the wider non-offending population, are difficult to establish, which impacts whether and how

USING QUANTITATIVE DATA TO CONSTRUCT BASE RATES

Clemmow et al. (2020) identified a variety of different approaches that could be used to establish base rates of attitudinal and socio-demographic factors that might contribute to radicalisation, including comparing different types of terrorists; and comparing terrorists with non-violent extremist offenders.

The authors compared those who engaged in terrorism or who held 'attitudinal affinity' with a violent extremist cause with members of the general population by comparing a pre-existing dataset of 125 lone actor terrorists against a survey of the general population (n=2,108). Through this approach, they were able to identify potential risk factors and protective factors among the general population.

The findings of such a study could be used to establish baselines against which to assess the success of secondary and tertiary interventions, as well as for designing primary interventions that address individual factors that are prioritised based on their prevalence among the general and offending populations. However, given that there is no simple linear relationship between individual characteristics or beliefs and involvement in extremism, these factors should not be seen as predictive of radicalisation.

¹¹ *Ris & Ernstorfer (2017)* have explored how the peacebuilding literature might overcome challenges relating to data collection. The discussion of base rates is drawn from *Clemmow et al. (2020)*, with *Ris & Ernstorfer (2017)* also discussing how base rates are dynamic and might change over time, creating further analytical challenges.

DIFFERENT POLICY CONTEXTS

12

The exceptionalism associated with P/CVE programmes means that commonalities with other types of intervention may be overlooked. Several studies have outlined how the evaluation of P/CVE interventions and programmes can learn from evaluations in other fields, such as medical, gang-related or peace-building interventions. Drawing on principles established in these fields would help to improve standards of monitoring and evaluation in P/CVE. When new methods or approaches are proposed, identifying comparable areas from other fields can help inform evaluation methods (worked examples which take this approach are set out in the following section).

A lack of policy coherence may limit opportunities for coordination and collaboration between actors, and across different policy areas. To increase policy coherence in contexts where individual projects

may have discrete theories of change, and particular actors may have specific objectives, the means by which individual projects contribute to overall policy objectives should be made as explicit as possible (see box below).

The high-profile nature of P/CVE work, and the policy urgency that can accompany it, means that unrealistic expectations about the potential impact of interventions may develop. Objectives can be ambitious, but they must also be realistic, and programme development and evaluation should be accompanied by clear explanations about what might be realistically achieved.

LINKING PROJECT AND PROGRAMME OBJECTIVES

The recent evaluation of the Building a Stronger Britain Together programme highlighted how the links between individual projects and overall policy aims could be made explicit by incorporating specific project objectives into an overarching theory of change, and by interviewing both providers and beneficiaries of individual projects to explore how each project contributed to the overall impact.

While P/CVE initiatives have different aims, evaluation strategies such as the one employed to evaluate Building a Stronger Britain Together could theoretically be adapted to evaluate primary intervention projects. This evaluation used a number of different data sources including surveys, qualitative interviews, project monitoring data and overall metrics such as the number of organisations supported, and the value of grant funding issued through the programme.

This type of coordination across policy areas could provide a foundation to consider how P/CVE efforts affect and are affected by other policy agendas. It may also help to develop a more holistic approach to tackling specific issues across different policy areas, and in turn to inform evaluation strategies.

12 *Davies et al. (2017)* explore how lessons from gang-related interventions can be applied to P/CVE; *Ris & Ernstorfer (2017)* discuss the applicability of the peacebuilding literature to this space and reflect on issues of policy coherence and unrealistic objectives; and *Baruch et al. (2018)* discuss the potential utility of public health evaluation principles. The evaluation of Building a Stronger Britain Together was conducted by *Ipsos MORI (2019)*. This evaluation demonstrates the importance of linking project objectives with policy goals, and of incorporating the views of different actors in evaluations (e.g. intervention providers, policy teams, beneficiaries).

RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES

Although the challenges associated with evaluating P/CVE programmes are well known, workable strategies for resolving these issues are not widely discussed in the literature. This section presents worked examples relating to four particularly difficult challenges: establishing a counterfactual; triangulating data; using control groups; and applying quasi-experimental research designs to P/CVE.

ESTABLISHING A COUNTERFACTUAL ¹³

KEY POINTS

- Identifying a robust counterfactual for P/CVE interventions is challenging. When evaluating P/CVE programmes, the ideal outcome involves identifying and proving a counterfactual, or ‘measuring a negative’. This would demonstrate that violence or radicalisation would have otherwise occurred had there not been an intervention. This challenge is most acute for pre-emptive programmes such as Channel, where there is no guarantee that individuals would have become violent without intervention.
- Counterfactual analysis offers a potential solution to this issue. Where experimental or quasi-experimental designs are not feasible, counterfactual analysis offers a structured method for establishing alternative scenarios against which the impacts of a P/CVE intervention or programme can be measured.

Evaluations of P/CVE interventions measure whether outcomes should be attributed to the intervention rather than other factors. To isolate the specific impact of an intervention, a counterfactual, or an understanding of what might have occurred if that programme had not

taken place, is required. Commonly, counterfactuals are provided by a control or comparison group who do not receive the treatment or intervention. However, as the next section discusses in more depth, there are ethical, practical, and national security challenges in trying to identify control groups for P/CVE interventions.

An alternative approach is to use counterfactual analysis, which provides a structured method to aid in ‘measuring a negative’ or establishing an alternative against which a programme’s impact can be measured. It is a useful approach where experimental or quasi-experimental designs are not practical.

Counterfactual analysis requires taking a known sequence of events and considering how they might have differed at critical points in ways which would have led to different outcomes. While it can never be conclusively established that these alternative scenarios would have occurred, counterfactual analysis can help create structured, transparent, and evidence-led comparisons in data-poor settings. The validity of counterfactuals is established by interpreting plausibility, likelihood, and result.

Examples showing how counterfactual analysis can help to assess the contributions of medical, planning and development, and aid programmes, demonstrate its potential for the P/CVE field. However, this potential has yet to be exploited. Processes for establishing counterfactuals can be drawn from the academic literature in comparable fields such as counterterrorism and intelligence analysis.

¹³ Counterfactual analysis has been used to assess different types of programmes across a range of fields (Baer & Fleming, 1976; Bae & Jun, 2003; Höfler, 2005). The process for undertaking counterfactual analysis presented here is adapted from Hendrickson’s (2009) guide for its use in security and intelligence assessments.

CASE STUDY: COUNTERFACTUAL ANALYSIS IN SECURITY AND INTELLIGENCE ASSESSMENTS

STAGE 1: SELECT POSSIBLE ALTERNATIVE SCENARIOS

The first stage of counterfactual analysis is to fully explore the range of possible scenarios that might have occurred in place of those which did. When evaluating a P/CVE programme this might include alternative scenarios where a participant did/did not commit an act of terrorism as well as the range of associated actions that precede these events. Establishing a plausible range of outcomes provides the necessary boundaries for thinking about any impacts of the programme being evaluated.

Step 1: Maximise the range of possible scenarios

Envisage as many alternative scenarios as practically possible. Exclusions should only be made where there is a specific reason for doing so. Scenarios should not be limited to what has happened previously.

Step 2: Weigh the plausibility of scenarios

Context-relevant information should be used to exclude scenarios that contain highly improbable events. Scenarios that are merely unlikely to have occurred should remain under consideration.

Preference should be given to sequences of events that reach the scenario outcome faster with fewer combinations of independent events. The shorter and less complex this sequence, the greater the amount of real information used in making projections, and the smaller the margin for error. With each alteration, or event added, the potential for error increases.

Preference should be given to sequences where the first counterfactual event has a higher probability of occurring. Again, this reduces the potential for errors in projections.

STAGE 2: DEVELOP SCENARIOS

After an alternative scenario has been selected, the time gap between known events and this scenario occurring must be revisited and evaluated. As many counterfactuals require projections that stretch over months, years, or even decades, the events included in this period are crucial to understanding the final outcomes. It is, therefore, necessary to determine any events likely to occur that are independent of the projected scenario but liable to impact upon it. For example, the known emergence of a new conflict in a specific region might increase the likelihood of a participant in a P/CVE programme reengaging in violence if they have a known history of travelling to such areas.

Step 3: Include events known or likely to occur during the timeframe of interest

Events should be included if they are known or are likely to occur and the probability of them doing so is not influenced by the projected scenario.

Events known or thought likely to occur during the timeframe of interest should be included where the probability of them happening is increased in the projected scenario.

Step 4: Include unknown events that form the basis of the projected scenario

Building on events known or likely to occur, the next step is to fill in those that are entirely projective or unknown. Moving sequentially from the beginning to the end of the counterfactual scenario, events that might occur at each juncture should be assessed and weighed. This ensures that any relationships that might influence the overall sequence of events are not overlooked and that those included logically follow from one another.

STAGE 3: ESTABLISH SCENARIO OUTCOMES

Counterfactual scenarios may end in several ways. The final stage is to evaluate the possible outcomes to select those that are most likely given the scenario and

RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES

Evaluating Programmes to Prevent and Counter Extremism

chain of events developed. This includes revisiting the original counterfactual question.

Step 5: Maximise the range of possible outcomes

Envisage as many possible outcomes that result from the scenario developed. No outcomes should be initially excluded without a concrete reason to do so. Caution must be exercised to avoid only considering outcomes that largely reflect what has happened in the past.

Step 6: Weigh the probabilities of the possible outcomes

The probabilities of each possible outcome should be weighed and ranked. Not all outcomes will have the same likelihood of occurring and there may be considerable overlap between the scenarios.

Step 7: Re-evaluate the original counterfactual question

Re-evaluate whether the final counterfactual addresses the original question as to whether the intervention has produced a discernible result. Assess the reasoning and the assumptions made in reaching the final outcome and determine what has and has not been established.

ACCESSING AND TRIANGULATING EVALUATION DATA ¹⁴

KEY POINTS

- Access to reliable and relevant data is an ongoing challenge for P/CVE evaluations. Evaluators face significant practical, ethical, and security issues seeking to use data that is not publicly available,

and which may be distributed across a range of agencies.

- Triangulating multiple data sources and taking a creative approach towards data collection can potentially help to overcome these problems,

Ethical and security challenges for P/CVE evaluation

Individuals supported by P/CVE interventions may be vulnerable due to their proximity to extremist actors.

Collecting and storing evaluation data can pose risks to programme staff, interviewees, partners and respondents.

The number of participants in tertiary P/CVE programmes is often small. Evaluations must ensure their identities remain protected.

Data collection that includes potentially traumatic or sensitive topics risks emotional or psychological harm to respondents.

including the ethical and security challenges listed in the box below.

A range of data can be used to evaluate P/CVE programmes, though data collection often poses significant challenges. Practical and ethical issues that are specific to P/CVE interventions include:

- The use of primary empirical data remains uncommon due to limitations in accessing classified information, and data protection and privacy legislation.
- Evaluators face important ethical and security issues, particularly when seeking to use data that is not publicly available, while data may be distributed across a range of agencies, making it harder to collate.

¹⁴ The most common approaches to collecting data on P/CVE programmes are reflected across a number of studies that have scoped the field (*Feddes & Gallucci, 2015; Schuurman & Bakker, 2016; Bellasio et al., 2018*). The need for triangulation is stressed in more robust work on gang-related interventions (*Davies et al., 2017*) and extremism (*Feddes & Gallucci, 2015; Cherney & Belton, 2019a*). For a detailed discussion of different types of triangulation, see *Bamberger (2012)* and *Farmer et al. (2006)*. The German study that adopted a multi-method approach to triangulate research findings is discussed by *Feddes & Gallucci, 2015* (the original study was published in German). The Dutch initiative is also identified by *Feddes & Gallucci* but was evaluated by *Schuurman & Bakker (2016)*. The US-based multi-method evaluation was conducted by *Williams et al. (2016)*. Other measures for assisting in data collection are drawn from studies that focus on carrying out evaluations in challenging contexts, such as *Lindekilde (2012); Davies et al. (2017)*, who discuss issues relating to 'dark figures'; and *Ris & Ernstorfer (2017)*, who discuss survey designs that avoid associating participants with sensitive information, and partnering with third party organisations. For studies that have used list experiments, including several designed to capture illegal or unacceptable behaviour, see a bibliography produced by *Imai (undated)*. *Clemmow et al. (2020)* have also used this approach to capture sensitive opinions in a survey of public attitudes around extremism. For a discussion of proxy indicators, see *Cherney et al. (2018)*.

- Accessing data from the beneficiaries of secondary and tertiary P/CVE interventions is particularly challenging, as national security concerns can often limit independent researchers from accessing individuals supported through such programmes.
- Given the small sample sizes, there is often a lack of robust administrative data for P/CVE interventions, such as centrally held figures on the total numbers of complete/incomplete cases, or data capturing how effective beneficiaries found the intervention to be, against which to triangulate any evaluation data.

The most common approaches to gathering data are observation and qualitative interviews, although surveys, focus groups, questionnaires, and data mining have also been used. The under-reporting of activities or attitudes associated with radicalisation creates 'dark figures' which have significant biases and limitations as they do not give a full picture of the prevalence of specific behaviours or attitudes among the population being analysed. Evaluations of peacebuilding interventions, and programmes targeting hard-to-reach populations, such as gangs, demonstrate how these issues can be overcome:

- **Triangulation:** Robust evaluations require data triangulation, or multi-method collection, something that can help mitigate the limitations of available data. Combining different forms of data, for example, by collecting qualitative interviews with intervention facilitators and official law enforcement statistics, provides a more complete picture that captures nuances that might be lost if using a single method.
- **Proxy indicators:** Data that indirectly helps to understand the impact of an intervention. This might include interviews with family members, peers, or probation workers, rather than participants. Measuring objective goals, like finding employment or returning to education, is another means of interpreting change. This might include or combine qualitative data such as interviews with family members, and quantitative data, like school attendance rates.
- **Survey design:** Using questionnaires that 'avoid associating respondents and researchers with potentially dangerous information'. One such

method includes the use of list experiments, in which sensitive questions are concealed within a long list of less sensitive questions. This might involve presenting one half of a sample with a list of views that includes one racist or extremist opinion and presenting the other half with the same list minus the sensitive opinion. By asking both samples how many of these statements they agree with, it is possible to estimate the number of people who agree with the sensitive opinion by comparing these two groups.

- **Third-party monitoring:** Typically includes using local civil society or research organisations to carry out data collection. Third-party monitoring may be used to increase or diversify data collection. This can be beneficial when P/CVE interventions, such as 'counter-narrative' programmes, are delivered remotely. The credibility and local knowledge of these organisations can help mitigate the security concerns that come with collecting data in difficult contexts and can assist in accessing hard-to-reach populations. Effective partnerships include dedicated feedback mechanisms for regular communication between organisations and evaluators.
- **Integrating anecdotal evidence:** Anecdotal evidence is commonly used to interpret the impact of P/CVE interventions. The robustness of such data can be increased by implementing systems to amalgamate and compare anecdotal understandings or integrate them into formal evidence-gathering processes.

Triangulation is a particularly valuable way of ensuring that evaluation findings are robust. There are a number of ways to build triangulation into research designs (see case study below), and several authors have used mixed methods approaches to triangulate the findings of P/CVE evaluations:

- The German EXIT programme collected quantitative data provided by the intervention provider including numbers of completed cases, current cases, and cases that ended before the programme was finished, and conducted a process evaluation using qualitative analysis of documents and interviews. This multi-method approach enabled the researchers to explore the underlying processes of the intervention, and link processes

RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES

Evaluating Programmes to Prevent and Counter Extremism

to the quantitative outcome data in a way that evidenced how the intervention contributed to those outcomes.

- A Dutch reintegration programme used three rounds of qualitative interviews to conduct a process and impact evaluation. The interviews charted the development and implementation of the programme over time (interpreting process) and explored the programme's perceived effectiveness (assessing impact) by asking project staff whether the programme had been successful in reducing recidivism among offenders; if the programme was an effective tool for monitoring offenders; and whether the programme was seen as useful by prosecutors.
- A US-based community-led CVE programme used data collection tools over multiple waves of fieldwork to collect qualitative (focus groups and interviews) and quantitative (survey) data from Muslim community leaders, Muslim communities, law enforcement, and county officials. Themes emerging from the first wave were verified by different members of the research team and were used to inform subsequent waves of fieldwork, which tested and verified these themes.

CASE STUDY: TRIANGULATING EVALUATION DATA USING MIXED METHODS ¹⁵

The use of a mixed-methods approach, which 'involves the systematic integration of different kinds of data, usually drawn from different evaluation designs', is one way to increase the validity and reliability of research findings. However, practical advice on how to triangulate different methods is limited.

Several authors advocate for using a '**triangulation protocol**' for combining qualitative and quantitative data. First developed for healthcare evaluations, this involves multiple stages.

Stage 1: Sorting

Sort findings from different sources / methods into

categories to determine areas of overlap / divergence.

Stage 2: Convergence coding

Identify themes from each data source / method, and compare the findings to characterise the degree and type of convergence using a coding frame of agreement; partial agreement; silence (i.e. a code is found in data collected from one source / method but not from the other source(s) / method(s) used); dissonance (i.e. the data collected from different sources / methods is contradictory on specific points).

Stage 3: Convergence assessment

Review all codes to provide an overall assessment of the convergence between data sources / methods.

Stage 4: Completeness assessment

Identify similar and unique contributions to the research question made by each data source / method to ensure that interpretations of the data and overall findings are a complete representation of the data.

Stage 5: Researcher comparison

Compare individual assessments of convergence or dissonance and completeness of the full set of findings across multiple researchers / research teams and identify and address different interpretations.

Stage 6: Feedback

Circulate triangulated results to research team / stakeholders for feedback.

A slightly different approach to triangulating mixed-methods data is the use of a **mixed-methods matrix**. The matrix includes rows which relate to individual cases (e.g. specific projects within a P/CVE intervention, or individual participants in those projects) and columns which refer to the different kinds of data collected for each case. Presenting the data in this way allows the researcher to look for patterns and differences across contrasting data sources before looking for patterns in the overall dataset.

¹⁵ This definition of mixed-method approaches is drawn from *USAID (2013)*. The triangulation protocol was developed by *Farmer et al. (2006)* for different types of qualitative data, although it has been used to triangulate qualitative and quantitative data. For a discussion of different approaches to triangulation, including protocols and mixed methods matrixes, see *O' Cathain et al. (2010)* who explore their applicability to medical evaluations.

IDENTIFYING AND USING CONTROL GROUPS ¹⁶

KEY POINTS

There are ethical and practical challenges in identifying control groups as it would not be ethical to deny an individual who is at risk of becoming involved, or who is already involved, in violent extremism access to an intervention to evaluate its effectiveness. To do so would also potentially raise national security concerns.

Several promising approaches for identifying and protecting control groups have been used in other fields and could be adapted to evaluate P/CVE interventions.

Several authors have discussed the need for control groups in evaluations of P/CVE interventions. However, these studies provide little detail about how this could feasibly be done. For example, one study argues that research designs using control groups are transferrable from the evaluation of gang interventions, but it does not identify ways of addressing the additional barriers to the use of control groups in P/CVE, which include:

Agencies may not be willing to have convicted terrorists or at-risk individuals allocated to a wait list to serve as a control group, or be unwilling to consent to them having some alternative support option (or no support) compared to an experimental group.

(Cherney & Belton, 2019b)

- National security issues that stem from the decision to deny an intervention to an individual for the purposes of testing its effectiveness.
- Ethical issues relating to consent that emerge when taking part in a P/CVE intervention, when doing so is a condition of sentencing or licence requirements.
- Practical issues associated with the challenges of a) identifying a matched sample in the context of the relatively low sample size of people eligible for some interventions; and b) the individualised nature of the radicalisation process and the difficulties controlling for the variety of different external factors that might influence individual radicalisation pathways.

The above challenges are less pressing for primary P/CVE interventions targeted at broader populations.

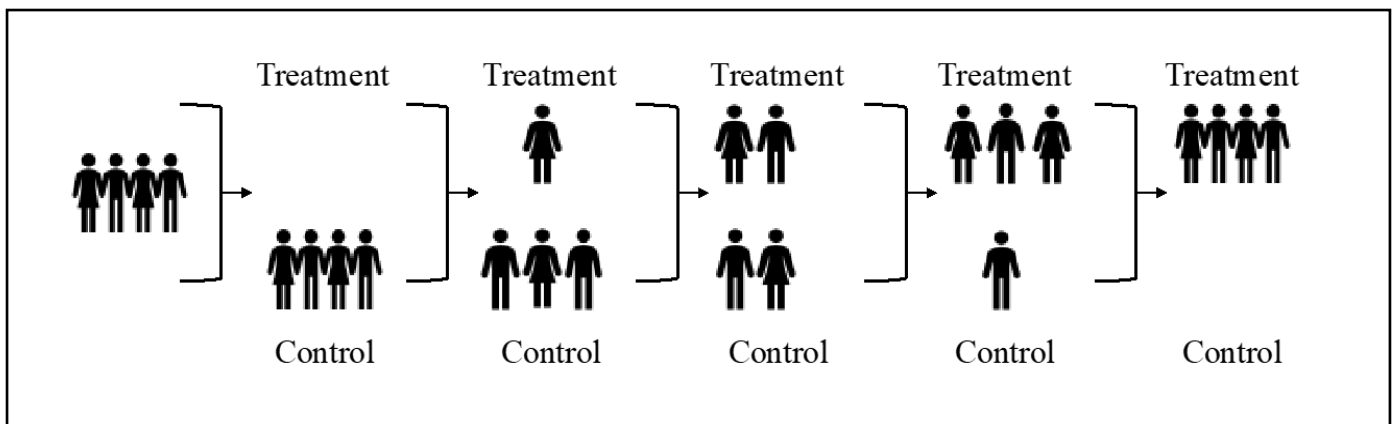


Diagram of Stepped Wedge Research Design (Based on Li et al. 2018)

¹⁶ The argument for the transferability of research designs using control groups is made by *Davies et al. (2017)*. Other authors who have called for control groups include *Khalil & Zeuthen (2016)* who also present a hypothetical research design that involves (a) identifying ‘at risk’ individuals; (b) applying ‘treatments’ to approximately half of these; and (c) assessing changes in their attitudes or behaviours as compared with the control group. The practical challenges that evaluators face in using P/CVE control groups are particularly well-discussed by *Koehler (2017)* and *Cherney (2020)*. For a discussion of the ethical challenges relating to consent, see *Hofman & Sutherland (2018)*. For specific challenges relating to secondary and tertiary interventions, see *Cherney & Belton (2019b)* and *Kutner (2016)*. The latter outlines a proposal for conducting an evaluation of Saudi Arabia’s deradicalisation programme and concludes “recidivism risks are too high to support a control or waitlist group”. Several studies have used control groups to test primary P/CVE interventions in educational settings including *Parker & Lindekilde (2020)*; *Theriault et al. (2017)* and *Amjad & Wood (2009)*. Parker and Lindekilde’s study is the most robust of these studies and is based on a large quantitative survey of young people (n=2,156).

RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES

Evaluating Programmes to Prevent and Counter Extremism

Several studies have successfully used control groups to evaluate the impact of such interventions, including a handful of experimental designs. The ethical and practical issues are more acute for secondary and tertiary P/CVE interventions. Given the security risks involved, it may not be feasible to use experimental designs with ‘no-treatment’ control groups to evaluate these interventions.

The use of RCTs in medical trials faces similar ethical issues around whether it is justifiable to deny potentially effective treatment to a control group to measure its efficacy. One approach used to overcome this challenge is the ‘stepped wedge’ or ‘dynamic wait list’ design. In contrast to a traditional ‘wait list’ design, in which half the sample receives treatment immediately before the other half receives the treatment at a pre-defined point in time, the dynamic design involves:

- All participants are initially allocated to the control group.
- Randomly allocated clusters of the control group are added to the treatment group sequentially, which can be compared to the control group, even as the control sample shrinks over time.
- By the end of the evaluation, all participants are in the treatment group (see diagram below).

This approach has been used to evaluate potentially time-sensitive interventions such as cancer treatment and suicide prevention programmes.¹⁷ It has been argued that this type of dynamic design is more appropriate than traditional wait list designs when ‘logistical or political considerations may require giving the treatment to the controls before the desired follow-up time has elapsed’. However, a limitation of this design is that the average wait time for treatment across the sample may be increased compared to traditional wait list designs, as it involves multiple stages of treatment as opposed to two. This means it may not be suitable for evaluating interventions with high-risk populations.

Medical studies have outlined how it might be possible

to use control groups in evaluations involving the most vulnerable populations, including those who have tried to commit suicide (see box below).

CASE STUDY: LEARNING FROM SUICIDE PREVENTION RESEARCH

It is unethical to deny potentially life-saving interventions to those at risk of suicide which makes it impossible to use ‘no-treatment’ control groups. However, several authors have outlined how it might be possible to offer suicide prevention treatments to control and treatment groups in an ethical way.

For example, independent teams of clinicians might be asked to evaluate whether the control condition meets the required standard of care, or research teams might offer ‘non-specific treatment conditions that are likely equal to or better to [the treatment as usual]’ to the control group. This type of approach could theoretically be used to compare different approaches to delivering specific parts of an intervention, such as alternative methodologies for providing counselling support, or for comparing different orders of treatments or different providers.

This literature also highlights how safeguards can be built into the research process to minimise the risk of the control group coming to harm. For example, researchers conducting a medical trial to compare the effectiveness of different drugs for treating bipolar disorder outlined a series of safeguards that they employed to protect their participants (all of whom had been diagnosed with bipolar disorder, were in acute stages of bipolar disorder associated with high suicide risk, and had previously attempted suicide):

- During the initial stage of the study, participants received one of the study drugs, alongside any antidepressant or antipsychotic medication required to address the acute stage they were in. They were switched to another antidepressant or antipsychotic if the original one did not work.

¹⁷ For a discussion of the stepped wedge design as it relates to medical interventions, see *Li et al. (2018)*; *Hemming et al. (2015)*. For its application to cancer treatment, see *Caminiti et al. (2017)*, and for suicide interventions, see *Hendricks Brown et al. (2006)*. For a discussion of the benefits of dynamic designs, see *Murray et al. (2010)*. The discussion of suicide prevention research, and the specific safeguards that have been used to protect treatment and control groups during studies, is drawn from *Oquendo et al. (2004)*, while *Cwick & Walkup (2008)* discuss the ethics of offering ‘non-specific treatment conditions that are likely equal to or better to [the treatment as usual]’.

- The antidepressant or antipsychotic drugs were tapered off over a set period, culminating in the participants only taking the study drug. At this point, the hypotheses could be tested.
- However, during this period, clinicians were given the flexibility to issue 'rescue drugs' to complement the study medication in the event of a relapse.
- If a participant demonstrated any of the items from the Scale for Suicide Ideation (e.g. writing a suicide note), then clinicians intervened with treatment.
- Psychosocial interventions, such as sessions with families to educate them about bipolar disorder, were offered, to provide an extra level of protection to the participant.
- Ongoing monitoring of all participants was carried out to ensure they were not at risk.
- The researchers periodically met with an independent data safety board to identify and remedy any risks emerging in the research as soon as was feasibly possible.

other fields that could be used to evaluate P/CVE interventions.

- Methods such as the switching-replications design could overcome some of the ethical issues relating to selecting control groups and could provide robust evidence of an intervention's impacts on attitudes and/or behaviour by comparing pre- and post-intervention data.

Quasi-experimental designs are well-suited to evaluating interventions where RCTs are impractical and/or unethical. They are typically easier and less costly to set up than experimental designs. In quasi-experimental evaluations, the evaluator uses a control group, but they have no say over who is assigned to the control and treatment samples. This approach often uses two different pre-existing groups, which ideally will be matched so they are as similar as possible. Once the intervention or treatment has been administered, the evaluator then compares the experimental group with the control group.

International examples of this approach include:

- A quasi-experimental study of 13 Sri Lankan terrorist rehabilitation centres compared the outcomes of a 'full-treatment group' and a 'minimal treatment group'. The former group was made up of 490 detainees in 12 centres who had received the full intervention, which included spiritual rehabilitation; sports and recreational rehabilitation; social, cultural and familial rehabilitation; and community rehabilitation. The latter was made up of 111 detainees in one centre where the full treatment programme had yet to be completed prior to their fieldwork.
- In an earlier study on the same intervention, the authors recognised that this non-matched control group was 'not exactly ideal', but argued it constituted a close enough approximation that

Learning from this approach, flexibility could be built into P/CVE interventions alongside constant monitoring and assessment to ensure that additional interventions are provided to individuals where needed, and to guarantee that they are not denied treatment that could have a positive effect on them.

USING QUASI-EXPERIMENTAL RESEARCH DESIGNS¹⁸

KEY POINTS

- Quasi-experimental designs are largely absent from evaluations of P/CVE interventions. Although quasi-experimental methods present challenges, there are a number of useful research designs that have been used to evaluate interventions in

¹⁸ The benefits and limitations of different evaluation designs are established in the literature (*Glasgow, 2013*), including those on P/CVE (*Feddes & Gallucci, 2015; Schuurman & Bakker, 2016; Baruch et al., 2018*) and other relevant contexts (*Sharma, 2016; Bellasio et al., 2018*). *Davies et al. (2017)* found that only two per cent of P/CVE interventions used quasi-experimental designs and provides an overview of how methods used for assessing gang interventions are much more rigorous than those in the P/CVE field. The same study also notes that 39 per cent of assessments of gang-related programmes employ more than one evaluation method. *Späth (2016)* recommends exploring whether comparative data can be drawn from data labs and identifying live control groups. For a discussion of quasi-experimental designs used to evaluate international examples of P/CVE interventions, see *Webber et al. (2018)* and *Kruglanski et al. (2014)* for their study of Sri Lanka; and *Aldrich (2014)* for a discussion of an intervention in Mali. *Braddock (2020)* discusses the utility of the switching-replications design for evaluating P/CVE interventions. For a discussion of opportunistic research, see *Watson et al. (2020)*.

RESPONDING TO THE MOST PRESSING EVALUATION CHALLENGES

Evaluating Programmes to Prevent and Counter Extremism

made it possible to draw meaningful comparisons based on the fact that the full and minimal-treatment group were well-matched on several relevant measures recorded in a baseline survey.

- A quasi-experimental design was used to evaluate the impact of USAID P/CVE activity in Mali. The evaluators surveyed 200 residents of two similar villages, one of which had access to this programming and one of which did not. The survey found that those with access to the programming demonstrated ‘measurably altered civic behaviour’ when compared to the control group, even after controlling for socio-demographic, political, and economic factors.

Several studies have called for the greater use of quasi-experimental research designs for evaluating P/CVE evaluations. A review of existing P/CVE evaluations found that only two per cent used quasi-experimental designs. By contrast, quasi-experimental evaluation approaches are increasingly seen in the fields of development economics, street crime, and gender-based violence, and account for almost half of all evaluations of gang-related interventions. Several of these designs are potentially applicable for assessing P/CVE programmes. Reviewing their use in other fields provides a useful learning opportunity for how to collect robust comparative data in challenging contexts. For example:

- During evaluations of community-based gang interventions, comparative data might be drawn from centralised ‘data labs’, or designated commissioners might be tasked with identifying potential live control groups that evaluators can use as a baseline to measure impacts against.
- Almost 40 per cent of assessments of gang-related programmes employ more than one evaluation method. Most are conducted by external evaluators, usually independent researchers from universities or government departments. This is in contrast to the majority of P/CVE evaluations where internal self-evaluation using a single method remains the norm.

Quasi-experimental designs face some of the same ethical challenges as experimental designs. Researchers must guarantee that all participants in the evaluation are exposed to the treatment so that they are all

Limitations to quasi-experimental designs

- Results can provide insights into change, but establishing causation can be challenging because of the risk that external factors have influenced the results.
- Quasi-experimental designs implemented after a programme has started may be limited by a lack of baseline data against which change can be measured.
- A comparison group that has been poorly matched to the intervention group will affect the strength of the findings.
- Finding suitable matches for programmes working with extremist offenders is difficult.

afforded the potential benefits of the intervention. One potential way to overcome this ethical challenge is to use a switching-replications design, in which the initial control and treatment groups are switched during the evaluation process (see case study below).

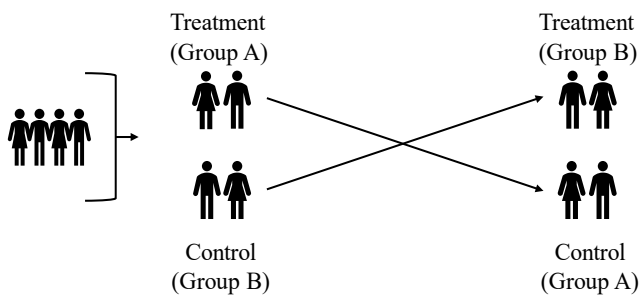
CASE STUDY: ESTABLISHING A SWITCHING-REPLICATIONS DESIGN

Braddock (2020) advocates for the use of non-equivalent switching replication designs to evaluate P/CVE interventions. This design is similar to the wait-list designs discussed earlier, except that a) participants are not randomly assigned to conditions and b) each condition serves as both treatment and control, as individuals who start in the control group end up in the treatment group, and vice versa. To illustrate, Braddock offers a hypothetical example for evaluating a P/CVE counselling programme:

- The researcher would measure all participants’ beliefs and attitudes before half of the participants (Group A) receive counselling.
- Following this first stage, the researcher measures all participants’ beliefs and compares the beliefs of Group A to the participants who have not yet received the intervention (Group B).

- Group B would then receive counselling, and Group A would not.
- The researcher would again measure and compare each group's attitudes

*Diagram of Switching-Replications Design
(Based on Li et al., 2018)*



Depending on the context and type of participants, there may be ethical and security issues in delaying the delivery of the intervention in this way. However, there

is a growing literature on 'opportunistic evaluation', where evaluators 'make use of naturally arising opportunities to study the effects and costs of those programmes and how they work'. This might include:

- Where P/CVE interventions are delivered to individuals on an ad hoc basis, and where there is some lead-in time between an individual being offered the intervention and that intervention starting, there may be naturally occurring delays that could be used to develop either a wait-list or a switching-replications design, which would avoid having to actively deny or delay treatment.
- Where there are variations in the availability of support at different times, and/or in different regions, such as when a provider only works in a specific region, or their workload necessitates a delay in working with a new case, there could be opportunities for opportunistic evaluation.

ASSESSING THE EVIDENCE BASE

The evidence base on how to evaluate P/CVE programmes is limited, although there is a significant amount of literature which has scoped the issues facing the field. This work has drawn attention to the limited quantity and quality of assessments, and the common challenges that programmes share. A growing number of studies propose using tools or methods from other fields and have considered their potential for addressing these issues. However, there is little evidence of these tools being validated in relation to real P/CVE evaluations. Where they have, studies have usually focused on single case studies, notably in Australia and Sri Lanka, which makes it difficult to generalise about their utility.

While lessons from comparable fields, such as gang-related or medical evaluations, can be applied to P/CVE programmes, further research is needed to understand where and when insights from other fields successfully translate to P/CVE evaluation. Their potential applicability needs to be confirmed with

empirical evidence as significant adaptation is likely to be required. The solutions proposed above should therefore be treated as examples of good practice rather than what has been proven to work.

Priority should be given to testing and validating the applicability of promising developments in other fields, such as the growing use of switching-replications designs in medical and educational evaluations. There are also important lessons to be learned from medical research on how to ethically use control groups in a way that limits the risk of participants coming to harm. Programme designers and evaluators should also consider how best to build adaptability and flexibility into evaluation programmes so that possibilities for opportunistic research can be exploited, although at this stage this may be impractical as it is difficult to design this level of flexibility into interventions that are already being delivered.

READ MORE

- Baruch, B., Ling, B., Warnes, R. and Hofman, J. (2018). Evaluation in an emerging field: Developing a measurement framework for the field of counter-violent-extremism. *Evaluation*, 24:4, 475–495.
- Braddock, K. (2020). *Experimentation & quasi-experimentation in countering violent extremism: Directions of future inquiry*. Washington DC: RESOLVE Network.
- Cherney, A. and Belton, E. (2019a). Assessing intervention outcomes targeting radicalised offenders: Testing the pro integration model of extremist disengagement as an evaluation tool. *Dynamics of Asymmetric Conflict* [Early Access Article].
- Cherney, A. and Belton, E. (2019b). Evaluating Case-Managed Approaches to Counter Radicalization and Violent Extremism: An Example of the Proactive Integrated Support Model (PRISM) Intervention. *Studies in Conflict & Terrorism* [Early Access Article].
- Cherney, A. (2020). Evaluating interventions to disengage extremist offenders: a study of the proactive integrated support model (PRISM). *Behavioral Sciences of Terrorism and Political Aggression*, 12:1, 17–36.
- Davies, M., Warnes, R. and Hofman, J. (2017). *Exploring the transferability and applicability of gang evaluation methodologies to counter-violent radicalisation*. Santa Monica, California: RAND Corporation.
- Ris, L. and Ernstorfer, A. (2017). *Borrowing a wheel: Applying existing design, monitoring, and evaluation strategies to emerging programming approaches to prevent and counter violent extremism*. Briefing Paper, Peacebuilding Evaluation Consortium.

For more information on CREST
and other CREST resources, visit
www.crestresearch.ac.uk



CREST

CENTRE FOR RESEARCH AND
EVIDENCE ON SECURITY THREATS

20-002-01