RAND EUROPE

Counterterrorism evaluation

Taking stock and looking ahead

Research report

Jacopo Bellasio, Joanna Hofman, Antonia Ward, Fook Nederveen, Anna Knack, Arya Sofia Meranto, Stijn Hoorens

For more information on this publication, visit www.rand.org/t/RR2628

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

RAND® is a registered trademark.

© 2018; Wetenschappelijk Onderzoek- en Documentatiecentrum

RAND Europe is a not-for-profit research organisation that helps to improve policy and decision making through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the sponsor.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

www.randeurope.org

Preface

In 2010, a study commissioned by the Research and Documentation Centre (Wetenschappelijk Onderzoek- en Documentatiecentrum, WODC) in the Netherlands first aimed to assess evaluation practice and culture in the fields of counterterrorism (CT) and preventing and countering violent extremism (PCVE).[1] The study found that evaluation of CT and PCVE strategies, policies and programmes was still in its infancy and that evaluation practices and approaches were not sufficiently focused on testing the results and underlying assumptions of CT and PCVE programmes. Seven years on from this initial study, in light of the continuous developments and investments occurring in the fields of CT and PCVE, there was a growing necessity for a renewed analysis of how strategies, policies and initiatives in these fields are evaluated and what advances could be made.

Mindful of this, in December 2017 the WODC commissioned RAND Europe to conduct a study aimed at investigating how evaluations of CT and PCVE policies in the Netherlands and abroad have been designed and conducted over the last five years, and what practical lessons can be drawn regarding such evaluations (WODC Project Number 2865). This report discusses the activities, results and findings of this study and presents recommendations for future work in this area. It is aimed at a specialist audience of academics, practitioners and policy-makers with an intimate understanding of evaluation, particularly in the context of CT and PCVE.

RAND Europe is an independent, not-for-profit policy research organisation that helps to improve policy and decisionmaking through research and analysis. This report has been peer-reviewed in accordance with RAND's quality assurance standards.

For more information about RAND Europe or this document, please contact Stijn Hoorens (hoorens@rand.org).

RAND Europe
Rue de la Loi 82, Bte 3
1040 Brussels
Belgium
Tel: +32 (2) 669 2400

RAND Europe
Westbrook Centre, Milton Road
Cambridge CB4 1YG
United Kingdom
Tel: +44 1223 353 329

[1] Nelen et al. (2010).

Table of contents

Preface i
Figures v
Tables vii
Executive summary ix
Samenvatting xvii
Acknowledgments xxv
Abbreviations xxvii
1. Introduction 1
1.1. Study rationale 1
1.2. Scope, objectives and research questions 2
1.3. Document purpose and outline 3
2. Methodology 5
2.1. Production of the CT and PCVE evaluations inventory 5
2.2. Development of an analytical framework 10
2.3. Analysis of evaluations and reporting 10
2.4. Limitations 12
3. Understanding the study context and its key definitions 15
3.1. Understanding the context 15
3.2. The definitional challenge 19
3.3. Conclusions 23
4. Building an analytical framework 25
4.1. Evaluation methodological characteristics 25
4.2. Evaluation quality characteristics 30
4.3. Initiative characteristics 33
4.4. Lessons learned 35
5. Analysing CT and PCVE evaluations 39
5.1. The study inventory of CT and PCVE evaluations 39
5.2. Characteristics of CT and PCVE initiatives in the inventory 45

5.3. The study inventory's evaluations 48
5.4. Discussion of findings in the light of previous reviews and recommendations 54
6. Identifying issues and learning lessons from CT and PCVE evaluations 59
6.1. An overview of issues identified and lessons learned from the study's inventory 59
6.2. Inherent complexities of the fields of CT and PCVE 60
6.3. Challenges associated with measuring real-world phenomena 62
6.4. Challenges associated with evaluation design 63
6.5. Practical difficulties of conducting evaluations 65
6.6. Drawbacks and benefits of specific methods 67
6.7. Discussion and recommendations 68
7. Overall conclusions and recommendations 73
7.1. Answering research questions 73
7.2. Recommendations 78
7.3. Concluding observations 82
References 85
Annex A. CT and PCVE evaluations inventory 95
Annex B. Methodology 99
B.1. Task 1 methodology and approach 99
B.2. Task 2 methodology and approach 116
B.3. Task 3 methodology and approach 117
B.4. Limitations 120
Annex C. Stakeholder elicitation details and tools 123
C.1. Interviewees list 123
C.2. Stakeholder interview protocol 124
C.3. Stakeholder interview consent form 127
Annex D. Analytical framework 129

Figures

Figure 2.3. Inventory production overview 8
Figure 5.1. Overview of the inventory production process 40
Figure B.1. Overview of study approach 99
Figure B.2. Overview of Task 1 methodology and approach 100
Figure B.3. Inventory production overview 110
Figure B.4. Overview of the inventory production process 116
Figure B.5. Overview of Task 2 methodology and approach 117
Figure B.6. Overview of Task 3 methodology and approach 118

Tables

Table 1.1. Research questions 3
Table 3.1. Definitions of key project concepts guiding the operationalisation of the study's inventory inclusion criteria 24
Table 4.1. Characteristics of formative and summative evaluation 26
Table 4.2. Approaches to assuring quality of evaluations 31
Table 4.3. Lessons learned categories questions 36
Table 4.4. Lessons learned validation questions 36
Table 4.5. Lessons learned validity scoring 37
Table 5.1. Study inventory analysis overview 41
Table 5.2. Overview of CT and PCVE initiatives evaluated by type and goals (n=50) 45
Table 5.3. Overview of CT and PCVE initiatives evaluated by target ideology, goals and level of intervention (n=50) 48
Table 5.4. Overview of CT and PCVE initiatives evaluated by implementing actor, goals and level of intervention (n=50) 48
Table 5.5. Overview of the purpose and approach of CT and PCVE initiatives' evaluations (n=48) 51
Table 5.6. Overview of the purpose, design and methods of CT and PCVE initiatives' evaluations (n=48) 51
Table 5.7. Overview of data for proxies used to characterise the quality of CT and PCVE initiatives' evaluations (n=48) 54
Table B.1. Academic databases searched 102
Table B.2. Search strings employed on the JSTOR, Scopus, EPPI-Centre, Web of Science and PubMed databases 103
Table B.3. Search strings employed on the SSRN academic database 104
Table B.4. Search strings employed on Dutch-focused academic databases 104
Table B.5. Overview of academic journals reviewed 105
Table B.6. Overview of grey literature repositories reviewed 106
Table B.7. Overview of search strings for identifying adjacent fields publications 107
Table B.8. Overview of results by academic database and search language employed 111
Table B.9. Overview of results of online databases consulted after intermediate review stages 112

Table B.10. Overview of academic journals and grey literature repositories reviewed and associated results 113
Table B.11. Overview of stakeholders and experts solicitation response rate and publications received . 115
Table C.1. Task 1 and 2 interviewees list..... 123
Table D.1. Analytical framework table structure..... 129
Table D.2. Analytical framework: *Section 1 – Evaluation methodological characteristics* 130
Table D.3. Analytical framework: *Section 2 – Evaluation quality characteristics* 133
Table D.4. Analytical framework: *Section 3 – Initiative characteristics* 135

Executive summary

Study context and rationale

Recent years have seen an uptick in terrorist and violent extremist incidents occurring across Europe. European countries, including the Netherlands, face a wide threat spectrum and the volume of terrorism- and violent extremism-related phenomena and crimes has also increased.

In response, European countries have made significant investments in strategies, policies and programmes designed to prevent and counter terrorism, violent extremism and associated phenomena. Holistic policy responses, such as a national counterterrorism strategy, have been designed and implemented with a view to both respond to terrorist threats and attacks, and increase societal and individual resilience to the lure of extremist ideologies.

Not least because of the dynamism and complexity of the phenomena involved, little is known as regards the effectiveness, relevance and impact of counterterrorism (CT) and preventing and countering violent extremism (PCVE) policies and programmes. Recent research suggests also that despite the volume of CT and PCVE initiatives established in recent years, the evidence base underpinning these remains limited and evaluation practice and investments are underdeveloped compared to the overall fields of CT and PCVE.[2]

Research scope and objectives

In 2010, a study commissioned by the WODC aimed to assess evaluation practice and culture in the fields of CT and PCVE.[3] The study found that evaluation of CT and PCVE strategies, policies and programmes was still in its infancy.

The WODC commissioned RAND Europe in December 2017 to conduct the current study, investigating how evaluations of CT and PCVE policies in the Netherlands and abroad have been designed and conducted over the last five years. Furthermore, the study investigates what practical lessons can be drawn regarding such evaluations and what actions and measures could be taken in the short and medium terms to mitigate any existing shortcomings.

[2] Lum et al. (2006); Nelen et al. (2010); Feddes & Gallucci (2015); Gielen (2017); Marret et al. (2017).

[3] Nelen et al. (2010).

Study approach

To achieve its main goal, the study revolved around three interconnected research tasks.

The study opened with the **production of a CT and PCVE evaluations inventory**. This entailed the development of an inventory of evaluations of CT and PCVE strategies, policies and interventions conducted in the Netherlands and abroad since 1 January 2013. This task entailed: (i) the undertaking of a targeted literature review, aimed at defining key project concepts and finalising the search strategy for building the study inventory; (ii) the undertaking of systematic and targeted literature searches on academic databases and grey literature repositories, as well as by engaging experts and stakeholders from the fields at hand (i.e. CT, PCVE and evaluation); (iii) stakeholder elicitation and validation activities conducted through semi-structured interviews and remote, written consultations; and (iv) the production of the study inventory through a multi-step and multi-strand review process.

In parallel to this, the study team undertook the **development of an analytical framework** to be used for assessing evaluations collected in the study's evaluations inventory. The analytical framework was based on a consultation of experts and a targeted literature review.

Lastly, the study team conducted an **analysis of the evaluations inventory and reported on findings, results and recommendations stemming from overall study activities**. This entailed the analysis of the CT and PCVE evaluations inventory produced at the start of the study through the lenses of the analytical framework previously prepared. Emerging findings and recommendations were subsequently validated and refined through peer review and expert consultations.

Answering the study's research questions

What is meant by counterterrorism policy in the Netherlands and other countries? And what different types of CT and PVCE measures, policies and interventions can be distinguished in the Netherlands and abroad?

The literature reviewed and experts consulted indicate that the scope, purpose and activities characterising CT have evolved in recent years. This is both a response to changes in the threat landscape and a result of a growing understanding of terrorism and violent extremism. In the aftermath of the 9/11 attacks, CT policy in many Western countries focused on coercive measures, relying on the role and activities of law enforcement and security agencies to prevent and tackle terrorist attacks.

In the last decade, however, there has been a growing recognition of the importance of adopting broader, holistic approaches to preventing and tackling terrorist activities. This has led to the development of initiatives designed to build resilience to violent extremist ideologies, and to address what we understand to be the root causes of violent radicalisation leading to terrorism. Such initiatives are broadly encompassed under the label of PCVE and are now considered an integral part of European CT policies and approaches.

CT and PCVE are highly interconnected and overlapping policy areas and distinguishing between the two is not always possible. For the purpose of this study, a high-level characterisation of the two work strands was adopted:

- PCVE was taken to encapsulate initiatives that address drivers of violence and extremism; that build resilience or immunity to extremist ideologies; and that deter or disrupt recruitment and

mobilisation by extremist groups, including by supporting the reintegration of (former) extremists and offenders.

- CT was taken to encapsulate initiatives designed to deter, disrupt or isolate groups that use terror, including by training and equipping the security apparatus; by increasing the capacity to prepare, prevent, protect or respond to terrorist incidents; and by interdicting or prosecuting through law enforcement activities.

The study developed a taxonomy for analysing CT and PCVE initiatives, characterising them according to the type of measure, their goal, target ideology, level of intervention, type of activities, group of focus, implementer, geographic scope and foreseen duration. It was found that no single CT and PCVE blueprint appears to exist. Terrorism and violent extremism are complex fields, and a wide array of approaches and programmes have been designed and implemented in recent years, ranging from broad communications campaigns for marginalised communities to targeted military measures for terrorists living in conflict zones.

Similarly, it was found that in recent years Dutch authorities have designed and implemented a wide array of CT and PCVE legislation, policies and measures. In particular, the current Dutch CT strategy, designed to cover the 2016 to 2020 period, recognises the need for a comprehensive approach, and emphasises. The strategy identifies five activities and areas for intervention: 1) *procure* (i.e. in a timely way gather and asses intelligence about (potential) threats to Dutch national security and national interests abroad); 2) *prevent* (i.e. prevent and disrupt extremism and foil terrorist attacks before they occur); 3) *protect* (i.e. protect people, property and vital processes from extremist and terrorist threats, both in the physical and virtual domains); 4) *prepare* (i.e. prepare optimally for extremist and terrorist violence and its consequences); and 5) *pursue* (i.e. enforce the law in the face of extremism and terrorism). Furthermore, and coherently with the framework laid out in the national CT strategy, it was found that a wide range of PCVE programmes have been developed and launched in the Netherlands in recent years, including initiatives that are both prison- and community-based.

What evaluations of counterterrorism and preventing and countering violent extremism policies have been conducted over the last five years in the Netherlands and abroad?

A pivotal aspect of this study entailed the preparation of an inventory of CT and PCVE evaluations conducted in the Netherlands and abroad since January 2013. While it is recognised that the process of compiling such an inventory was to an extent arbitrary, the study team adopted an approach designed to reduce ambiguity about the selection of sources and reducing the risk of overlooking or excluding any relevant publications.

The study inventory comprises of 48 publications (38 in English, 6 in Dutch and 4 in German). On the surface, the number of publications identified since 2013 by this study seems to suggest that CT and PCVE evaluation has gained traction in recent years and, given the inclusion criteria employed, that greater emphasis is being placed on collecting primary data for the purpose of evaluation. This appears to be consistent with previous observation made by Feddes & Gallucci (2015) and Gielen (2017), who describes this field as having 'taken flight' between 2016 and 2017.

What can be said about counterterrorism and preventing and countering violent extremism evaluations characteristics? What are the differences and similarities between the identified evaluations of counterterrorism policies?

The majority of publications included in the study inventory focus on initiatives undertaken at the level of programmes or interventions, whereas fewer publications focus on higher-level policies and strategies.

In terms of initiative goals, those with a PCVE-specific or -relevant nature dominate (45 out of 50 initiatives evaluated). In contrast, only five evaluated initiatives were CT-specific and only an additional 17 had a clear CT-relevance. The initiatives evaluated were equally distributed across different levels of intervention (at individual level, at the level of groups or organisations, or at the level of entire social systems, such as countries or cities).

The small majority of initiatives evaluated by publications in the study's inventory focus exclusively on religious extremism, whereas approximately 17 initiatives evaluated target different types of extremism. Only 6 initiatives reviewed targeted political extremism.

With regard to actors responsible for implementation, a small majority of initiatives was implemented by governmental actors, whereas non-governmental actors were responsible for 8 out of 50 initiatives and public-private partnerships for 15 initiatives respectively. Overall, it was observed that certain target groups, namely communities and individuals vulnerable or at risk, were predominantly targeted by the sample of initiatives evaluated.

Looking at evaluation characteristics, it was found that evaluations included in the study's inventory comprise primarily process evaluations as well as impact- or outcome-focused evaluations. No economic evaluations were encountered in the study inventory. The majority of evaluations reviewed had no clear evaluation approach, although a number of theory-driven, realist and participatory evaluations were identified.

From an evaluation design perspective, the majority of publications included in the study's inventory focus on attributing additionality (i.e. evaluations that look at what happened after an initiative has been implemented, taking a single measurement, and thus unable to provide evidence as to whether change has occurred due to the lack of a pre-initiative baseline measurement and counterfactual evidence). Such designs are used consistently for evaluating initiatives across different levels of intervention and for both CT- and PCVE-specific and -relevant work. Furthermore, a number of quasi-experimental and longitudinal designs were also adopted among impact or outcome evaluations.

From the point of view of methods, the majority of impact or outcome evaluations use quantitative methods alone or quantitative methods as part of a mixed approach. Conversely, most process evaluations rely solely on qualitative methods. Overall, methods employed typically include combinations of desk research, interviews, focus groups and case studies for the purpose of qualitative data collection, and surveys for quantitative data. As for methods for analysis of qualitative data, many publications left these unstated, others used "thematic analysis." Descriptive and advanced statistical methods were mostly employed for the analysis of quantitative data.

Looking at the timing of evaluations, the study's inventory comprises a majority of interim evaluations and a limited number of ex post evaluations and embedded (ongoing) evaluations. No ex ante evaluations were recorded in the study's inventory.

Overall, when compared with results from previous reviews, the findings emerging from this study suggest that a growing volume of CT and PCVE evaluations is being undertaken and that the majority of these rely on primary data from multiple sources, perspectives and methods. While this finding is encouraging, there appear to be limits to the extent to which evaluation practice has advanced and grown evenly across all areas of CT and PCVE work. Furthermore, significant gaps and shortcomings continue to mar a number of evaluations. For example, some evaluations are characterised by designs that undermine their purported focus on and ability to come to robust conclusions about an initiative's impact. Other evaluations rely solely on secondary data and did not have access to stakeholders and beneficiaries during the undertaking of their work due to the sensitivity of such initiatives.

What does the evaluation literature say about quality criteria for evaluations? To what extent do the identified counterterrorism and preventing and countering violent extremism evaluations meet these quality criteria?

This study stresses that evaluation quality is a fluid concept. This is due to the fact that: (i) evaluation quality can be perceived differently by evaluators, stakeholders and decision-makers involved in an evaluative undertaking; and (ii) the concept of evaluation quality has been subject to change over time as standards for what constitutes or is considered good evaluation evolve and progress.

Different guidelines and principles exist for assessing evaluation quality. They usually pertain to: (i) the technical quality of the information produced; (ii) the quality of the process used to obtain the information discussed; and (iii) the usefulness of the information produced. However, an assessment of how well the articles included in this review meet the criteria established in the literature was beyond the scope of this study. Such an assessment would require a comprehensive meta-evaluation of these sources.

Instead, this study has aimed to capture information on quality assessments already performed on the evaluations included in the inventory, if any and if reported. In addition, we identified whether some minimal expectations for a promising evaluation have been met. These minimal expectations include (but are not limited to) capturing multiple perspectives (to limit bias), using grids, rubrics, scores or a set of indicators (to enhance transparency and objectivity), and providing recommendations (to increase the probability that evaluation results are used).

The majority of publications included in the study's inventory were subject to blind peer review, although in several instances it was not possible to determine which, if any, mechanism had been employed for review. The majority of evaluations also used multiple data sources or methods and considered different perspectives to draw conclusions, although almost a fifth relied on individual data sources. Furthermore, the vast majority of initiatives evaluated did not entail the use of grids, rubrics or scores for monitoring implementation work throughout its undertaking.

As regards formulating recommendations, it was found that the majority of evaluations did this, but that in no instance were the intended task-owners and timeframes for implementation indicated.

What practical lessons can be drawn on the basis of the existing evaluations regarding the evaluation of counterterrorism and preventing and countering violent extremism policy?

The lessons and reflections by the authors of the manuscripts have been thematically analysed. These can be clustered into five overarching themes which are touched upon briefly in the paragraphs below. A detailed discussion of these issues is available in the study report.

Firstly, the studies reviewed emphasised some of the inherent **complexities of the fields of CT and PCVE** and the impact that these have on evaluation practice. These include challenges stemming from lack of clear and shared definitions for key CT and PCVE concepts, security concerns, complex stakeholder landscapes, a limited understanding of initiatives to be evaluated.

Secondly, the studies reviewed **discussed a number of challenges associated with measuring real-world phenomena connected to CT and PCVE**. In particular, authors lamented how the lack of outcome measures, outcome metrics, opportunities to measure long-term initiative effects, instruments accounting for social norms and expectations, and measures tracking exposure to interventions hampered their ability to design and conduct robust evaluations.

Thirdly, the studies reviewed highlighted a number of lessons learned as regards **evaluation design**. These include difficulties encountered with evaluating initiatives lacking a theory of change, challenges in isolating initiatives' effects and attributing these to a programme, constraints faced when establishing randomised experiments and quasi-experiments, and adopting a longitudinal study approach.

Fourthly, the studies reviewed discussed **practical difficulties encountered when implementing evaluations**. These include resource constraints, difficulty accessing stakeholders as well as information about interventions and their effects, and difficulties with sample size.

Lastly, some of the studies reviewed discussed the **drawbacks and benefits of specific methods** employed during an evaluation. These include lessons about the constraints of model-based investigations, the constraints of survey instruments, or the importance of triangulation of results.

Recommendations

The recommendations discussed throughout the report include a mix of more readily implementable measures, which could be adopted over the short and medium term, as well as more ambitious recommendations for the field of CT and PCVE evaluation, both in the Netherlands and abroad. The recommendations formulated are aimed at stakeholders and organisations engaged in CT and PCVE evaluations through their design, implementation and commissioning.

With a view to sustaining and expanding CT and PCVE evaluation efforts, agencies funding initiatives and research in these fields should **continue to invest in the evaluation of planned and existing initiatives in these policy areas**. Furthermore, **minimum quality and robustness requirements should be expected from future evaluations**, including the use of empirical data, multiple methodologies and adequate stakeholder engagement methods. As part of this push, commissioners and evaluators should **encourage greater transparency and clarity about evaluation methodologies and approaches used to allow better judgement of the evidence strength and learning**.

In parallel to this, known shortcomings and gaps in CT and PCVE evaluation practice should be addressed. Greater **efforts should be made to ensure that wherever possible (quasi-)experimental designs are employed within CT and PCVE evaluation**. Researchers' access to key data and stakeholders, even if potentially sensitive or confidential, should also be improved, including by:

- Designing and implementing approaches for providing researchers with access to sensitive information on CT and PCVE initiatives and beneficiaries;
- Facilitating the establishment of mechanisms for data sharing among researchers.; and

- Collecting regular baseline measurements and identifying proxy measures and alternative indicators to mitigate the impact of data gaps.

In addition to evaluations, lessons extracted from the literature indicate that **further research is needed on the dynamics, drivers and factors governing the phenomena of radicalisation, violent extremism and terrorism.** As part of these efforts, **mapping and stocktaking exercises akin to the present study should be undertaken regularly and with a comparable methodology**. Such **future exercises should also be provided with the resources and means to gain access to CT and PCVE initiatives' evaluators and beneficiaries**, with a view to providing a more in-depth and robust assessment of evaluations analysed. **Future research efforts akin to this study should also consider focusing on conducting a more in-depth comparative analysis of existing reviews** and build on results of previous research efforts. Emphasis within future research should also be placed on **developing new evaluation designs, frameworks and approaches for conducting evaluation in the CT and PCVE policy areas** to broaden and strengthen the theoretical underpinnings of evaluations conducted in these fields.

Lastly, CT and PCVE evaluators may benefit from adopting approaches aimed at disentangling complexities of this field and pursuing an approach to knowledge development focusing on the progressive accumulation of knowledge and insights. To this end, **evaluators and commissioning agencies should consider embedding in their approach to CT and PCVE evaluation a 'co-production' ethos**. The purpose of moving towards co-production is to contribute to generating new forms and approaches for initiatives to be implemented, as well as to generating values beyond economic value (e.g. by fostering the development of new working partnerships and the exchange of ideas and knowledge).

Samenvatting

Beleidscontext en aanleiding voor het onderzoek

In de afgelopen jaren heeft zich een stijging voorgedaan in het aantal terroristische en gewelddadig extremistische incidenten in Europa. Het palet aan zulke dreigingen in Nederland en in andere Europese landen is breed en het aantal misdaden en andere aan terrorisme gerelateerde verschijnselen is ook toegenomen.

Als reactie op deze ontwikkelingen hebben Europese landen aanzienlijk geïnvesteerd in maatregelen en programma's gericht op het bestrijden en voorkomen van terrorisme, gewelddadig extremisme en de daarmee samenhangende verschijnselen. Bovendien hebben veel landen nationale contraterrorismestrategieën aangenomen om daadkrachtiger te kunnen optreden bij terroristische dreiging en aanslagen, en om de maatschappelijke en individuele weerbaarheid te vergroten tegen de aantrekkingskracht van extremistische ideologieën. Zulke maatregelen worden in de literatuur aangeduid met de afkortingen CT (*counterterrorism*) en PCVE (*Preventing and Countering Violent Extremism*).

Er is echter weinig bekend over de effectiviteit, relevantie en impact van CT- en PCVE-beleid. Dit heeft onder andere te maken met de dynamiek en complexiteit van de materie. Ondanks het groeiende volume aan CT- en PCVE- maatregelen in de afgelopen jaren, geven recente studies aan dat er vooralsnog weinig geïnvesteerd wordt in evaluaties op het gebied van CT en PCVE en dat het wetenschappelijk bewijs voor de effectiviteit van veel CT- en PCVE-maatregelen nog altijd beperkt is.[4]

Onderzoeksafbakening en doelstellingen

In opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) werd in 2010 een onderzoek uitgevoerd naar de praktijk en cultuur van evaluatie op het gebied van terrorismebestrijding.[5] Hieruit kwam naar voren dat de evaluatie van verschillende CT-strategieën, -beleid en –programma's nog in de kinderschoenen stond.

In december 2017 kreeg RAND Europe de opdracht van het WODC om onderzoek te doen naar evaluaties van CT- en PCVE-beleid in Nederland en het buitenland. Hierbij is onder meer gekeken naar de lessen die uit evaluaties van de afgelopen vijf jaar kunnen worden getrokken en naar de mogelijkheden om eventuele tekortkomingen in de toekomst op te vangen.

[4] Lum et al. (2006); Nelen et al. (2010); Feddes & Gallucci (2015); Gielen (2017); Marret et al. (2017).

[5] Nelen et al. (2010).

Onderzoeksopzet

Om het doel van deze studie te realiseren, is het onderzoek opgedeeld in drie onderdelen.

Ten eerste **zijn de beschikbare evaluaties op het gebied van CT en PCVE geïnventariseerd**. De resulterende bibliografie bevat evaluaties uit Nederland én andere landen sinds 1 januari 2013. Hiervoor hebben we de volgende onderzoeksactiviteiten uitgevoerd: 1) een gerichte literatuurstudie om sleutelbegrippen af te bakenen en de zoekstrategie te bepalen; 2) een literatuuronderzoek door middel van het systematisch en gericht doorzoeken van academische databanken en registers met grijze literatuur, in combinatie met het raadplegen van belanghebbenden bij CT- en PCVE-beleid en deskundigen op dit terrein; 3) een verdieping en validatie aan de hand van semigestructureerde interviews met en het schriftelijk raadplegen van experts uit de CT- en PCVE-evaluatiepraktijk; en 4) het beoordelen van de relevantie van de evaluaties voor onze inventarisatie aan de hand van een aantal criteria.

Daarnaast hebben we **een analytisch kader opgesteld** waarmee de geïnventariseerde evaluaties konden worden geanalyseerd. Het analytisch kader is in een aantal iteratieve stappen ontwikkeld. Een eerste versie van het raamwerk was gebaseerd op de gerichte literatuurstudie en interne workshops met experts met kennis van de evaluatiepraktijk op het gebied van CT en PCVE. Daarna is het raamwerk verder verfijnd op basis van feedback van belanghebbenden bij CT- en PCVE-beleid en deskundigen op het terrein van de evaluatie ervan.

Ten slotte zijn de geïnventariseerde **CT- en PCVE-evaluaties geanalyseerd en zijn de bevindingen en aanbevelingen gerapporteerd**. Hierbij is gebruikgemaakt van het analytisch kader. De eerste bevindingen en aanbevelingen die uit de analyse naar voren zijn gekomen, zijn gevalideerd en verfijnd door een aantal interne en externe workshops en door overleg met de wetenschappelijke begeleidingscommissie en *peer reviewers*.

Beantwoording van de onderzoeksvragen

Wat wordt er in Nederland en andere landen verstaan onder contraterrorismebeleid? En welke verschillende typen CT- en PCVE-maatregelen, -beleid en -interventies zijn in binnen- en buitenland te onderscheiden?

De verschillende typen maatregelen binnen het veld van terrorismebestrijding zijn enigszins aan verandering onderhevig geweest de afgelopen jaren. Dit is een gevolg van ontwikkelingen in het dreigingsbeeld, maar het vloeit ook voort uit een verbeterd inzicht in de aard en oorzaken van terrorisme en gewelddadig extremisme.

In de nasleep van de aanslagen op 11 september 2001 hebben westerse landen drastische maatregelen genomen ter bestrijding van terrorisme. Repressieve maatregelen vormden aanvankelijk de kern van het CT-beleid. De heersende gedachte was dat terrorisme moest worden bestreden door de rol van handhavings- en veiligheidsdiensten te verstevigen. De laatste jaren wordt de nadruk geleidelijk meer gelegd op het voeren van een brede, holistische benadering van contraterrorisme. Er is een verschuiving opgetreden richting het versterken van de weerbaarheid tegen gewelddadig extremistische ideologieën en het aanpakken van gewelddadig extremisme aan de wortel. Zulke initiatieven vallen onder het label PCVE en worden inmiddels als wezenlijk onderdeel gezien van contraterrorismebeleid in Europa.

CT en PCVE zijn sterk met elkaar verbonden en er is veel overlap tussen de twee beleidsterreinen. Om die reden is het niet altijd mogelijk onderscheid te maken tussen CT en PCVE. In deze studie zijn de kernbegrippen als volgt gedefinieerd:

- Onder PCVE worden alle initiatieven verstaan die gericht zijn op: 1) de aanpak van de oorzaken van geweld en extremisme; 2) het versterken van de weerbaarheid tegen extremistische ideologieën; en 3) het ontmoedigen en ontregelen van werving en mobilisatie. Dit laatste omvat ook de ondersteuning van de re-integratie van (voormalig) extremisten en delinquenten.
- Onder CT worden alle initiatieven verstaan die gericht zijn op het ontmoedigen, ontregelen of isoleren van groepen die terreur als wapen hanteren. Hierbij kan gedacht worden aan: 1) het trainen en uitrusten van veiligheidsdiensten; 2) het optimaliseren van voorbereiding op, voorkomen van, bescherming tegen en reageren op terroristische dreigingen, geweld en de gevolgen daarvan; en 3) rechtshandhaving en vervolging.

In dit onderzoek is een taxonomie gehanteerd, waarbij de CT- en PCVE-maatregelen worden geclassificeerd naar type, doel, beoogde ideologie, niveau waarop de interventie is doorgevoerd, type activiteiten, doelgroep, uitvoerder, toepassingsgebied en verwachte duur. Hieruit is duidelijk geworden dat er geen blauwdruk bestaat voor CT- en PCVE-maatregelen. Terrorisme en gewelddadig extremisme zijn complexe domeinen en in de afgelopen jaren is een waaier aan maatregelen genomen en programma's geïmplementeerd. Deze varieerden van bewustwordingscampagnes gericht op gemarginaliseerde groepen tot gerichte militaire ingrepen tegen terroristen in conflictgebieden.

Ook in Nederland is in de afgelopen jaren een breed scala aan CT- en PCVE-beleidsmaatregelen en -programma's en aan CT- en PCVE gerelateerde wet- en regelgeving geïmplementeerd. Bovendien onderschrijft de huidige Nationale Contraterrorismestrategie 2016-2020 het belang van een brede benadering. De strategie definieert vijf interventiegebieden: 1) *verwerven* – het tijdig zicht krijgen op en duiden van (potentiële) dreigingen in of tegen Nederland en de Nederlandse belangen in het buitenland; 2) *voorkomen* – het voorkomen en verstoren van extremisme en terrorisme en het verijdelen van aanslagen; 3) *verdedigen* – het beschermen van personen, objecten en vitale processen tegen extremistische en terroristische dreigingen, zowel fysiek als online; 4) *voorbereiden* – het optimaal voorbereid zijn op extremistisch en terroristisch geweld en de gevolgen daarvan; en 5) *vervolgen* – handhaving van de democratische rechtsstaat tegen extremisme en terrorisme. Daarbij is er in lijn met de Nationale Contraterrorismestrategie een breed spectrum aan PCVE-programma's geïntroduceerd, onder andere gericht op gedetineerden en andere risicogroepen.

Welke evaluaties van CT- en PCVE-beleid zijn er de afgelopen vijf jaar uitgevoerd in binnen- en buitenland?

Centraal in dit onderzoek staat de inventarisatie van reeds uitgevoerde evaluaties van CT- en PCVE-beleid. Om de kans te verkleinen dat belangrijke bronnen over het hoofd werden gezien, hebben we naast een systematische zoekopdracht in de beschikbare databanken voor academische literatuur, ook enkele doelgerichte zoekstrategieën toegepast. Niettemin erkennen we dat het gebruikte selectieproces voor deze inventarisatie enigszins arbitrair is. Omwille van de transparantie en reproduceerbaarheid zijn de zoekstrategieën en de bijbehorende selectiecriteria uitgebreid in het rapport beschreven.

De hierboven beschreven inventarisatie leverde 48 publicaties op (38 in het Engels, 6 in het Nederlands en 4 in het Duits) met in totaal 50 evaluaties (twee van de publicaties rapporteerden over twee evaluaties tegelijk). Op basis van het aantal geïnventariseerde publicaties sinds 2013 lijkt het veld van CT- en PCVE-evaluaties zich verder te hebben ontwikkeld sinds eerdere inventarisaties. Met het oog op de gehanteerde selectiecriteria blijkt ook dat er in de evaluaties meer nadruk wordt gelegd op primaire dataverzameling dan voorheen. Dit lijkt aan te sluiten bij eerdere observaties van Feddes en Gallucci (2015) en Gielen (2017), die concludeerden dat deze gebieden recentelijk een vlucht hebben genomen.

Wat kan worden geconcludeerd over de eigenschappen van deze CT- en PCVE-evaluaties? Welke verschillen en overeenkomsten zijn te benoemen ten aanzien van de geïnventariseerde evaluaties?

De meerderheid van de geïnventariseerde publicaties beschouwt initiatieven gericht op individuen, groepen of hun sociale omgeving: op het niveau van programma's of interventies. Een minderheid richt zich op een hoger aggregatieniveau van beleid of strategieën. Verreweg de meeste geëvalueerde initiatieven (45 van de 50) zijn specifiek gericht op PCVE. Daartegenover staat dat slechts 5 geëvalueerde initiatieven specifiek op CT gericht zijn en dat 17 geëvalueerde initiatieven een CT-component hebben. De niveaus van de geëvalueerde interventies blijken redelijk evenwichtig verdeeld over individuen (micro), groepen of organisaties (meso) en het hele sociale systeem van bijvoorbeeld een stad of land (macro).

Een kleine meerderheid van initiatieven is uitsluitend gericht op religieus extremisme, terwijl 17 van de geëvalueerde initiatieven zich concentreren op verschillende soorten extremisme. Slechts 6 van de 50 initiatieven zijn uitsluitend toegespitst op politiek extremisme.

De meeste evaluaties zijn uitgevoerd door individuen of groepen binnen een overheidsinstantie. Evaluatoren buiten de overheid en publiek-private partnerschappen zijn verantwoordelijk voor respectievelijk 8 en 15 van de 50 evaluaties.

Met betrekking tot de eigenschappen en doelstellingen van de geïnventariseerde evaluaties betreft het voornamelijk proces- en effectevaluaties. We hebben geen economische evaluaties aangetroffen. In de meerderheid van de publicaties wordt er geen of geen duidelijk theoretisch kader geschetst voor de evaluatie. Maar in een aantal gevallen wordt door de auteurs een "*theory-driven*" aanpak, een zogenaamde "realistische evaluatie" of een "participatieve evaluatie" beschreven.

Vanwege beperkingen in het evaluatieontwerp, kunnen in de meeste publicaties geen robuuste uitspraken gedaan worden over de causale relatie tussen de interventies en hun effecten. Deze evaluaties moeten zich beperken tot uitspraken over "additionaliteit". Dit betreft evaluaties die zich richten op wat er gebeurd is nadat een initiatief is geïmplementeerd op basis van een enkele meting. Door het ontbreken van een nulmeting en een *counterfactual* (zoals een controlegroep) kan echter niet worden vastgesteld of de opgetreden verandering heeft plaatsgevonden als gevolg van de interventie. Dergelijke evaluaties worden binnen zowel het CT- als het PCVE-domein en op alle niveaus toegepast, min of meer gelijkmatig verdeeld over het micro-, meso- en macro-niveau. Tevens hebben we onder de effectevaluaties een aantal quasi-experimentele en longitudinale designs aangetroffen.

Vanuit methodologisch perspectief worden in het grootste deel van de effectevaluaties kwantitatieve methoden toegepast: of uitsluitend kwantitatieve methoden, of als onderdeel van een benadering waarin verschillende methodes worden gecombineerd ("*mixed method*"). In procesevaluaties worden daarentegen voornamelijk kwalitatieve methoden toegepast. Over het algemeen worden combinaties van

literatuuronderzoek, interviews, focusgroepen en case studies gebruikt voor het verzamelen van kwalitatieve data en worden enquêtes gebruikt voor kwantitatieve data. De methoden voor de analyse van kwalitatieve gegevens blijven in de meeste gevallen onduidelijk, maar in sommige gevallen wordt een benadering gebruikt waarbij per thema de bevindingen worden geanalyseerd ("thematische analyse"). Voor de analyse van kwantitatieve gegevens worden veelal beschrijvende en geavanceerde statistische methoden toegepast.

Wat betreft de timing van de evaluaties vinden we met name tussentijdse evaluaties. Er is maar een beperkt aantal *ex post* en zogenaamde "*embedded*" of lopende evaluaties opgenomen, waarbij de evaluatie een geïntegreerd onderdeel is van de implementatie en uitvoering van de maatregel. Zogenaamde *ex ante* evaluaties, waarbij de te verwachten effecten worden ingeschat alvorens de maatregel daadwerkelijk wordt geïmplementeerd, worden zelfs helemaal niet aangetroffen.

Wanneer deze resultaten naast de uitkomsten van vergelijkbare studies uit het verleden worden gelegd, kunnen we voorzichtig concluderen dat een groeiend volume aan CT- en PCVE-evaluaties gebruikmaakt van primaire data uit verschillende bronnen en van verschillende perspectieven en methoden. Hoewel dit een bemoedigende constatering is, lijkt de progressie binnen de evaluatiepraktijk zich niet in alle deelgebieden van CT en PCVE voor te doen. Bovendien blijft een aantal evaluaties belangrijke beperkingen en tekortkomingen vertonen. Hiertoe behoren bijvoorbeeld evaluaties waarbij het niet mogelijk is om robuuste conclusies te trekken over de impact van een maatregel, bijvoorbeeld omdat de steekproef niet groot genoeg is. In andere gevallen zorgt de gevoeligheid van dergelijke interventies ervoor dat evaluaties uitsluitend gebaseerd zijn op secundaire data en er geen sprake is van betrokkenheid van belanghebbenden en begunstigden gedurende het hele proces.

Wat zegt de evaluatieliteratuur over (methodologische) kwaliteitscriteria voor evaluaties? In hoeverre voldoen de geïnventariseerde evaluaties aan deze kwaliteitscriteria?

Deze studie onderstreept dat de kwaliteit van evaluaties een rekbaar begrip is. Enerzijds komt dit doordat opvattingen over evaluatiekwaliteit uiteen kunnen lopen. Anderzijds zijn ideeën over wat een evaluatie "goed" maakt en welke normen gevolgd dienen te worden aan verandering onderhevig.

Er bestaan verschillende richtlijnen voor het beoordelen van de kwaliteit van een evaluatie. Doorgaans hebben deze betrekking op drie zaken: 1) de technische kwaliteit van de informatie; 2) de kwaliteit van het proces dat gevolgd is om deze informatie te verkrijgen en; 3) de bruikbaarheid van de uitkomsten. Maar binnen de context van dit onderzoek hebben we niet kunnen nagaan in hoeverre de geïnventariseerde evaluaties voldoen aan de kwaliteitscriteria uit de literatuur. Een uitvoerige meta-evaluatie van deze bronnen valt buiten de afbakening van de voorliggende studie.

Binnen dit onderzoek is vastgelegd welke mechanismen voor kwaliteitswaarborging zijn gehanteerd, voor zo ver dit door de auteurs is gerapporteerd. Bovendien is vastgesteld of er aan enkele minimumeisen voor een zinvolle evaluatie is voldaan. Hiertoe behoren onder andere het in acht nemen van verschillende perspectieven (om vooringenomenheid te beperken), het toepassen van een evaluatiekader of scoringssysteem, het gebruik van indicatoren (om transparantie en objectiviteit te vergroten) en het doen van aanbevelingen (om de kans te vergroten dat evaluatieresultaten gebruikt worden).

Het merendeel van de geïnventariseerde publicaties is getoetst op kwaliteit door middel van peerreview, waarbij anonieme experts het manuscript hebben beoordeeld. In enkele gevallen kan niet worden

achterhaald of, en zo ja, hoe het manuscript door onafhankelijke experts is beoordeeld. De meerderheid van de evaluaties maakt gebruik van meerdere bronnen en methoden en neemt de perspectieven van verschillende betrokkenen en belanghebbenden in beschouwing. Echter, een vijfde van de geïnventariseerde evaluaties baseert zich op slechts één bron. In de overgrote meerderheid van evaluaties worden geen evaluatiekader, scoringssysteem of indicatoren toegepast die de voortgang en de effecten van de maatregelen en hun implementatie bijhouden.

Ten slotte worden in de meerderheid van de geïnventariseerde evaluaties aanbevelingen gedaan. Hierbij ontbreekt echter in alle gevallen informatie over de termijn en planning wat betreft de implementatie van de aanbevelingen en over het aanwijzen van beoogde verantwoordelijken voor de implementatie.

Welke lessen kunnen op basis van de bestaande evaluaties worden getrokken ten aanzien van het evalueren van CT- en PCVE-beleid?

Ook de reflecties van auteurs van de geïnventariseerde evaluaties en de lessen die door hen worden getrokken, zijn geanalyseerd. Deze kunnen worden gevat in vijf overkoepelende thema's die hieronder kort uiteen worden gezet.

Ten eerste benadrukken auteurs het **complexe karakter van CT en PCVE** als onderzoeksgebieden en de gevolgen die dit heeft voor het uitvoeren van evaluaties. Hierbij kan men denken aan het ontbreken van een brede consensus over definities voor kernbegrippen die horen bij CT en PCVE en de uitdagingen die dit met zich meebrengt voor evaluaties. Daarnaast blijkt het in evaluaties moeilijk om veiligheidsbelangen af te wegen tegen andere belangen. Bovendien is het landschap van belanghebbenden complex en bestaat er een beperkte kennis over de maatregelen die geëvalueerd worden.

Ten tweede zijn er **verschillende aspecten van de dagelijkse praktijk van CT en PCVE die de evaluatie van zulke maatregelen compliceren**. Hierbij wordt onder andere het gebrek aan criteria voor het meten van de effectiviteit van de maatregelen op de lange termijn genoemd. Daarnaast is het moeilijk gebleken om sociale normen en verwachtingen meetbaar te maken. Zo is het bijvoorbeeld mogelijk dat deelnemers aan een programma, naast de interventie zelf, ook worden beïnvloed door veranderingen in hun sociale omgeving. Ten slotte noemen de auteurs het gebrek aan robuuste methoden om te meten in hoeverre doelgroepen zijn blootgesteld aan interventies, zoals de mate waarin zij iets hebben meegekregen van een bewustwordingscampagne.

Ten derde wordt een aantal lessen getrokken met betrekking tot **het evaluatieontwerp**. Hieronder vallen moeilijkheden bij het evalueren van maatregelen zonder dat daarbij gebruik wordt gemaakt van een "*theory of change*" of interventielogica, waarin de logische oorzaak-gevolgketen van beoogde effecten uiteen wordt gezet. Daarnaast hebben veel evaluaties moeite met het isoleren van de effecten van beleid, een programma of ander type interventie en het garanderen dat deze alleen kunnen worden toegedicht aan de interventie (attributie). Ook het opzetten en uitvoeren van gerandomiseerde (quasi-)experimenten (zoals bijvoorbeeld *randomised controlled trials*) en longitudinale onderzoeksopzetten blijken moeilijk te verenigen met de beleidspraktijk van CT en PCVE.

Ten vierde worden **praktische problemen ervaren bij de uitvoering van evaluaties**. Hiertoe kunnen worden gerekend: een gebrek aan financiële middelen, tijd en expertise; beperkte betrokkenheid van of toegang tot verschillende belanghebbenden; onvoldoende beschikbare informatie over de interventies en hun effecten; en moeilijkheden omtrent steekproeven.

Ten slotte worden de **voor- en nadelen van specifieke methoden** aangestipt in enkele onderzochte publicaties. In het bijzonder gaat het hier over de beperkingen van bijvoorbeeld onderzoeken op basis van wiskundige of economische modellen, de beperkingen van enquête-instrumenten, of het belang van triangulatie van gegevens. Zo zijn de resultaten van kwantitatieve methoden in sterke mate afhankelijk van de beschikbaarheid en kwaliteit van invoerdata, hetgeen vaak te wensen overlaat in het veld van CT en PCVE. Zo worden er voorbeelden genoemd van vergelijkbare studies naar hetzelfde fenomeen die tot tegenovergestelde conclusies hebben geleid.

Aanbevelingen

Onze aanbevelingen vormen een mix van eenvoudig te implementeren maatregelen op korte en middellange termijn en ambitieuzere langetermijnaanbevelingen voor de evaluatie van CT- en PCVE-beleid in Nederland en daarbuiten. De aanbevelingen richten zich op belanghebbenden en organisaties die betrokken zijn bij de evaluatie van CT- en PCVE-maatregelen, als evaluator of als opdrachtgever van de evaluatie.

Het verdient de aanbeveling dat organisaties die initiatieven en onderzoek op het gebied van CT en PCVE financieren en/of implementeren, **blijven investeren in het (laten) evalueren van CT- en PCVE-maatregelen**. Als men serieus geïnteresseerd is in de effectiviteit van beleid, dan vergt dat tijd, expertise, vaardigheden, mankracht en dus financiële middelen. Verder zouden **toekomstige evaluaties moeten voldoen aan een bepaalde minimale standaard wat betreft de kwaliteit en robuustheid**. Hierbij kan gedacht worden aan: het verzamelen en analyseren van empirische data over de implementatie en de effecten van de maatregel in kwestie, het gebruik van verschillende methodologieën en het op een adequate manier betrekken van belanghebbenden bij het onderzoek. Om deze kwaliteitsslag te kunnen maken, moeten evaluatoren ook **meer transparantie en duidelijkheid bieden in het rapporteren over de evaluatiemethoden en -benaderingen die gebruikt worden** en daarvoor door hun opdrachtgevers de gelegenheid krijgen. **Met meer transparantie over data en methoden kan een beter oordeel gegeven worden over de robuustheid van het wetenschappelijk bewijs en kunnen lessen worden getrokken over hoe de evaluatie te verbeteren.**

Tegelijkertijd valt er voor CT en PCVE een inhaalslag te maken ten opzichte van onderzoeksgebieden met meer evaluatie-ervaring. **Waar mogelijk moeten (quasi-)experimentele evaluatieontwerpen worden aangemoedigd**. Daarnaast moeten belangrijke informatie en gegevens beter beschikbaar komen voor evaluatieonderzoek, ook als deze informatie mogelijk gevoelig of vertrouwelijk is. Deze situatie kan worden verbeterd door:

- Het creëren van mogelijkheden voor onderzoekers om op een veilige en vertrouwelijke wijze toegang te krijgen tot gevoelige informatie over CT- en PCVE-initiatieven en begunstigden;
- Het verbeteren van de mogelijkheden voor het delen en uitwisselen van gegevens tussen onderzoekers; en
- Regelmatig nulmetingen uit te voeren en gegevens te verzamelen over de implementatie en uitkomsten van de maatregelen. Als effecten moeilijk direct meetbaar zijn, kunnen (proxy-) indicatoren mogelijk helpen de informatieachterstand te verkleinen.

Naast leerpunten ter verbetering van de evaluatiepraktijk wijst de literatuur erop dat **meer onderzoek nodig is naar de dynamiek, drijfveren en elementen van radicalisering, gewelddadig extremisme en terrorisme.** Het **inventariseren en in kaart brengen van ontwikkelingen binnen het vakgebied,** zoals in dit onderzoek is gedaan, is hier onderdeel van. **Het verdient de aanbeveling deze met enige regelmaat te herhalen, gebruikmakend van een vergelijkbare methodologie.** Daarnaast zouden toekomstige **inspanningen ook moeten worden voorzien van voldoende middelen om de evaluaties uitgebreid te kunnen doorgronden en te verbeteren**. Eventueel zou in **toekomstig onderzoek, vergelijkbaar met de voorliggende studie, overwogen kunnen worden een diepgaandere vergelijkende analyse uit te voeren van bestaande *reviews* en voort te bouwen op resultaten van eerdere onderzoeksinspanningen.** Ook wordt aanbevolen in toekomstig onderzoek naar de evaluatiepraktijk binnen de CT- en PCVE-beleidsterreinen een nadruk te leggen op het **ontwikkelen van nieuwe evaluatieontwerpen, -methodes en analytische kaders,** om zo de CT- en PCVE-evaluaties theoretisch beter te onderbouwen en robuuste uitspraken over causaliteit mogelijk te maken.

Ten slotte zouden CT- en PCVE-evaluatoren kunnen profiteren van onderzoeksbenaderingen gericht op algemene kennisontwikkeling binnen dit veld om de complexiteit en onzekerheden te reduceren. Om dit te bewerkstelligen kunnen **evaluatoren en opdrachtgevers overwegen om in hun benadering van CT- en PCVE-evaluatie meer aan "co-productie" te doen,** waarbij individuen van verschillende organisaties samenwerken om tot een gezamenlijk product te komen. Het doel van dergelijke co-producties is het bijdragen aan beleidsinnovatie (de ontwikkeling van nieuwe benaderingen en maatregelen ter bestrijding en preventie van terrorisme en gewelddadig extremisme), evenals het toevoegen van waarde die verder gaat dan alleen de economische waarde (zoals bijvoorbeeld nieuwe samenwerkingsverbanden en het uitwisselen van ideeën en kennis).

Acknowledgments

This project could not have been conducted without the funding provided by the WODC. Within the WODC, particular thanks go to the research manager of this study, Lars Heuts, for his active engagement, constructive feedback and facilitation of activities.

The authors are also grateful to a large number of individuals and organisations who contributed by providing information, steering and advice, donating their time to be interviewed, and commenting on draft versions of this report. We acknowledge them below in no particular order and thank them for their inputs. Any mistakes in this report remain, however, the sole responsibility of the authors.

We are thankful to the members of the Scientific Advisory Committee assembled by the WODC, who offered their expertise throughout the study, providing feedback on the methodology and research design, and commenting on draft versions of this report. The Scientific Advisory Committee comprised its chairman, Prof. Antony Pemberton (Tilburg University), as well as Dr Marnix Croes (National Coordinator for Security and Counterterrorism, NCTV), Dr Lonneke van Noije (Netherlands Institute for Social Research, SCP) and Dr Marianne van Ooyen-Houben.

We are also grateful to Dr Allard Feddes (University of Amsterdam) and Prof. Tom Ling (RAND Europe) for the input, guidance and constructive criticism they provided in their role as peer reviewers in the context of RAND Europe's Quality Assurance system.

Finally, and not least, a number of individuals contributed to the study by participating in key informant interviews or by corresponding via email. Those individuals who consented to being acknowledged are listed in Appendix C. Some preferred to remain anonymous.

Abbreviations

CT	Counterterrorism
FTF	Foreign Terrorist Fighter
EU	European Union
MSMS	Maryland Scientific Methods Scale
PCVE	Preventing and Countering Violent Extremism
RAN	Radicalisation Awareness Network
RCT	Randomised Controlled Trial
RTE	Real-Time Evaluation
SAC	Scientific Advisory Committee
SAI	State Audit Institution
SSRN	Social Science Research Network
TER	Terrorists, Extremists and Radicals
UK	United Kingdom
US	United States
WODC	Research and Documentation Centre (Wetenschappelijk Onderzoek- en Documentatiecentrum)

1. Introduction

Recent years have seen an uptick in terrorist and violent extremist incidents occurring across Europe. European countries, including the Netherlands, face a wide threat spectrum that spans from complex attacks organised by highly skilled transnational organisations to improvised but effective actions carried out by so-called 'lone-wolves'.[6]

The volume of terrorism- and violent extremism-related phenomena and crimes has also increased. Terrorist and extremist propaganda has taken advantage of modern technologies and especially social media to reach a wider audience and maximise its recruitment efforts. As a result, thousands of European citizens and long-term residents have travelled in the last decade to conflict zones in Syria, Iraq and elsewhere to join extremist groups. Several women, young adults and children have also been involved, often only to return to their home or host countries in Europe after a span of months. In addition, the prolonged refugee and migration crisis as well as terrorist attacks faced by European countries have contributed to growing societal polarisation, leading to a resurgence of both far-right- and far-left-inspired violence and agitation.[7]

In response, European countries have made significant investments in strategies, policies and programmes designed to prevent and counter terrorism, violent extremism and associated phenomena. Holistic policy responses, such as a national counterterrorism strategy,[8] have been designed and implemented with a view to both respond to terrorist threats and attacks, and increase societal and individual resilience to the lure of extremist ideologies.[9]

1.1. Study rationale

Not least because of the dynamism and complexity of the phenomena involved, little is known as regards the effectiveness, relevance and impact of counterterrorism (CT) and preventing and countering violent extremism (PCVE) policies and programmes. Recent research suggests also that despite the volume of CT and PCVE initiatives established in recent years, the evidence base underpinning these remains limited

[6] Europol (2017).

[7] Europol (2017).

[8] For example, the Dutch National Counter-terrorism Strategy (*Nationale Contraterrorismestrategie*): NCTV (2016a).

[9] Gielen (2017).

and evaluation practice and investments are underdeveloped compared to the overall fields of CT and PCVE.[10]

In 2010, a study commissioned by the WODC aimed to assess evaluation practice and culture in the fields of CT and PCVE.[11] The study found that evaluation of CT and PCVE strategies, policies and programmes was still in its infancy. Furthermore, the work found that CT and PCVE evaluations disproportionately focused on legislative measures and that the majority of studies comprised *ex post* evaluations with a focus on assessing the implementation and execution of policies, as opposed to disclosing and testing underlying assumptions and theories of change. Not least in light of the continuous developments and investments occurring in the fields of CT and PCVE, in the years since the pivotal 2010 study there has been a growing necessity for a renewed analysis of how to effectively evaluate policies and initiatives in these fields.[12]

1.2. Scope, objectives and research questions

Mindful of the gap in research described above, in December 2017 the WODC commissioned RAND Europe to conduct the current study, investigating how evaluations of CT and PCVE policies in the Netherlands and abroad have been designed and conducted over the last five years (WODC Project Number 2865). Furthermore, the study investigates what practical lessons can be drawn regarding such evaluations and what actions and measures could be taken in the short and medium terms to mitigate any existing shortcomings.

To achieve this, the study was designed to pursue a number of intermediate objectives revolving around three interdependent activities:

- Producing an inventory of existing evaluations of CT and PCVE policies and interventions conducted in the Netherlands and abroad over the last five years.
- Developing an analytical framework to assess evaluations collected in the study inventory.
- Analysing evaluations collected with a view to identifying and assessing practical lessons learned concerning the evaluation of CT and PCVE policies and programmes at the individual and aggregate level.

Table 1.1 provides an overview of the research questions that the study team focused its work on answering. The questions presented in the table were identified by the WODC in the project terms of reference.

[10] Lum et al. (2006); Nelen et al. (2010); Feddes & Gallucci (2015); Gielen (2017); Marret et al. (2017).

[11] Nelen et al. (2010).

[12] Marret et al. (2017).

Table 1.1. Research questions

	Research questions
1	What is meant by counterterrorism policy in the Netherlands and other countries? What different types of counterterrorism measures/policies/interventions can be distinguished in the Netherlands and abroad?
2	What evaluations of counterterrorism and preventing and countering violent extremism policies have been conducted over the last five years in the Netherlands and abroad?
3	What can be said about these evaluations on, among other things: • The subject of evaluation. • The type of measure/intervention/policy under evaluation. • The intervention logic. • The reason for conducting the evaluation (did any specific events or decisions precede the evaluation?). • The objective(s) of the evaluation. • The evaluation methodology. • The scope of the evaluation. • The extent to which the evaluation assesses costs and benefits of the intervention (and the level of evidence underpinning this assessment). • The extent to which the evaluation assesses the effects and side-effects of measures (and the level of evidence underpinning of this assessment). • The constraints on, limitations of, and assumptions underpinning the evaluation and the way in which they have been addressed. • The evaluation's strengths and weaknesses.
4	What are the differences and similarities between the identified evaluations of counterterrorism and preventing and countering violent extremism policies?
5a 5b	What does the evaluation literature say about quality criteria for evaluations? To what extent do the identified counterterrorism and preventing and countering violent extremism evaluations meet these quality criteria?
6	What (practical) lessons can be drawn on the basis of the existing evaluations regarding the evaluation of counterterrorism and preventing and countering violent extremism policy (by type of evaluation)?

1.3. Document purpose and outline

This document serves as the study final report; it discusses the activities, results and findings of the study team and presents recommendations for future work in the area of CT and PCVE evaluation. The report is aimed at a specialist audience of academics, practitioners and policy-makers with an intimate understanding of evaluation in the context of CT and PCVE. The document is structured as follows:

- **Chapter 2 – Methodology**. This chapter presents an overview of the study team's approach to performing the research and analysis activities underpinning this study.
- **Chapter 3 – Understanding the study context and its key definitions**. This chapter discusses the context in which the study took place and the key definitions underpinning the research.
- **Chapter 4 – Building an analytical framework**. This chapter provides an overview of the analytical framework designed to analyse CT and PCVE evaluations gathered in the study inventory.
- **Chapter 5 – Analysing CT and PCVE evaluations**. This chapter discusses the study inventory, focusing on the characteristics of both the initiatives being evaluated and their evaluations.

- **Chapter 6 – Identifying issues and learning lessons from CT and PCVE evaluation**. This chapter discusses issues identified and lessons learned about CT and PCVE evaluations as encountered in the literature included in the inventory.
- **Chapter 7 – Overall conclusions and recommendations**. This chapter provides a synthetic overview of the study's findings vis-à-vis the overarching research questions, and of recommendations and reflections formulated by researchers in light of the results observed.

In addition, this document contains four annexes: Annex A lists the publications included in the study inventory; Annex B presents a detailed overview of the study's methodology; Annex C lists the stakeholders and experts consulted, and reproduces the protocols and tools employed for this purpose; and Annex D presents an overview of study's analytical framework.

2. Methodology

This chapter provides an overview of the methodology and approach employed by the research team to achieve the objectives of study. A more detailed description of each of the steps, which would allow the approach to be replicated in the future, is given in Annex B.

2.1. Production of the CT and PCVE evaluations inventory

As a first step, in order to produce an inventory of evaluations of CT and PCVE strategies, policies and interventions conducted in the Netherlands and abroad since 1 January 2013, the study team developed and refined definitions for key concepts and underpinning phenomena. To achieve this, the study team undertook four activities: (i) the development of inclusion criteria and refinement of definitions; (ii) systematic and targeted literature searches for evaluations; (iii) stakeholder elicitation of (unpublished) evaluation manuscripts and validation; and (iv) compilation of the inventory of evaluations.

2.1.1. Develop inclusion criteria and refine definitions

The study team conducted a non-systematic review of academic and grey literature on CT, PCVE and evaluation. The purpose of this activity was twofold: (i) to refine and consolidate the study's inclusion criteria to be employed for determining whether a manuscript or publication should be included in the study inventory; and (ii) to produce operational definitions for key concepts and underpinning phenomena, with a view to guiding the review process and the determination of whether sources are within or outside the scope of the study inventory (see Section 2.1.4).

Literature reviewed was identified through a non-systematic snowballing technique (i.e. a technique whereby, based on an initial set of key sources, additional studies and resources are identified through their citations) and was limited to sources available in the English language.

2.1.2. Systematic and targeted literature searches

The study team conducted literature searches, using both systematic (i.e. replicable) and targeted research approaches, to investigate academic and grey literature repositories. The study team also engaged with expert and professional networks to solicit the submission of unpublished manuscripts. The purpose of this activity was twofold: (i) to identify publications of evaluations of CT and PCVE strategies, policies and interventions in English, Dutch, French and German published since 1 January 2013; and (ii) to identify publications from fields adjacent to CT and PCVE that may present transferable lessons learned about evaluation approaches.

Systematic searches

The study team conducted systematic searches for CT and PCVE evaluation publications through online academic journal databases. Searches focused on publications released since 1 January 2013 and written in English, Dutch, German and French. In addition to Dutch and English (as the lingua franca for scholarly communication), German and French were selected on the basis of the following criteria:

1. Feasibility and availability of language skills within the research team.
2. The language expected body of evaluation literature.
3. The anticipated number of CT and PCVE interventions available in the countries where these languages are used.

For more details on the protocol applied for the systematic searches, see Annex B.

Targeted review of academic journals and grey literature repositories

In addition to searches on generalist academic databases, the study team conducted a review of publications released since 1 January 2013 through a selected number of (i) academic journals relevant to the scope of this study; and (ii) national and international repositories of grey literature. This was done to ensure that:

1. No relevant publication released under a specialist journal was overlooked because it did not meet the search criteria used for database searches.
2. Relevant publications not indexed on academic databases but available through national and international institutions active in the fields of CT and PCVE were included in the study inventory.

For more details on the journals and grey literature repositories reviewed, see Annex B.

Solicitation of unpublished or draft manuscripts from external experts and practitioners

The study team sent direct requests via email for unpublished or forthcoming manuscripts and evaluations to a wide network of CT and PCVE policy-makers, practitioners, experts and evaluators. Each expert identified by the study team was initially contacted with a request and, in case of no response, was contacted a second and final time after two weeks. A total of 52 experts from 21 countries from across North America, Europe, Asia and Oceania were contacted. Overall, 23 responded to our solicitations, providing the details of 30 manuscripts.

Structured search of relevant publications from adjacent fields

The study team conducted targeted searches to identify publications discussing evaluation approaches and lessons learned from fields that can be considered adjacent to CT and PCVE from the perspective of evaluation challenges. The fields selected for review were those of criminology, gangs' desistance, peacebuilding and cult exit. Previous research by RAND Europe indicates that evaluations from

established academic fields can be used to inform the development of evaluation practice in relatively novel research areas.[13]

For more details on the approach to identifying relevant publications from adjacent fields, see Annex B. It should be noted that this strand of the research and analysis produced limited results in the context of this study. Where feasible and relevant, publications identified through this approach are referenced in the discussion of study findings and results.

2.1.3. Expert consultation and validation activities

In addition to the previous steps, the study team conducted a series of stakeholder and expert consultation activities. Interviews were conducted with seven academics, experts and practitioners from organisations and institutions active in CT and PCVE in the Netherlands and abroad, including Europe, North America and developing countries in Asia and the Middle East. When selecting interviewees, the study team endeavoured to engage experts active in different areas of CT, PCVE and evaluation from both the Netherlands and abroad. The primary purpose of these interviews was that of:

- Understanding how CT and PCVE can be conceptualised and what are the differences / overlap / tension lines between activities in these two fields.
- Producing an initial mapping of challenges, barriers and enablers characterising CT and PCVE evaluations, capturing any lessons learned from interviewees as regards these.
- Gathering information on any public or non-public manuscript or publication that should be included in the study inventory.

In addition to semi-structured interviews, remote consultations with the project Scientific Advisory Committee (SAC) were held to discuss emerging results and validate upcoming research activities.

2.1.4. Compiling the inventory of evaluations

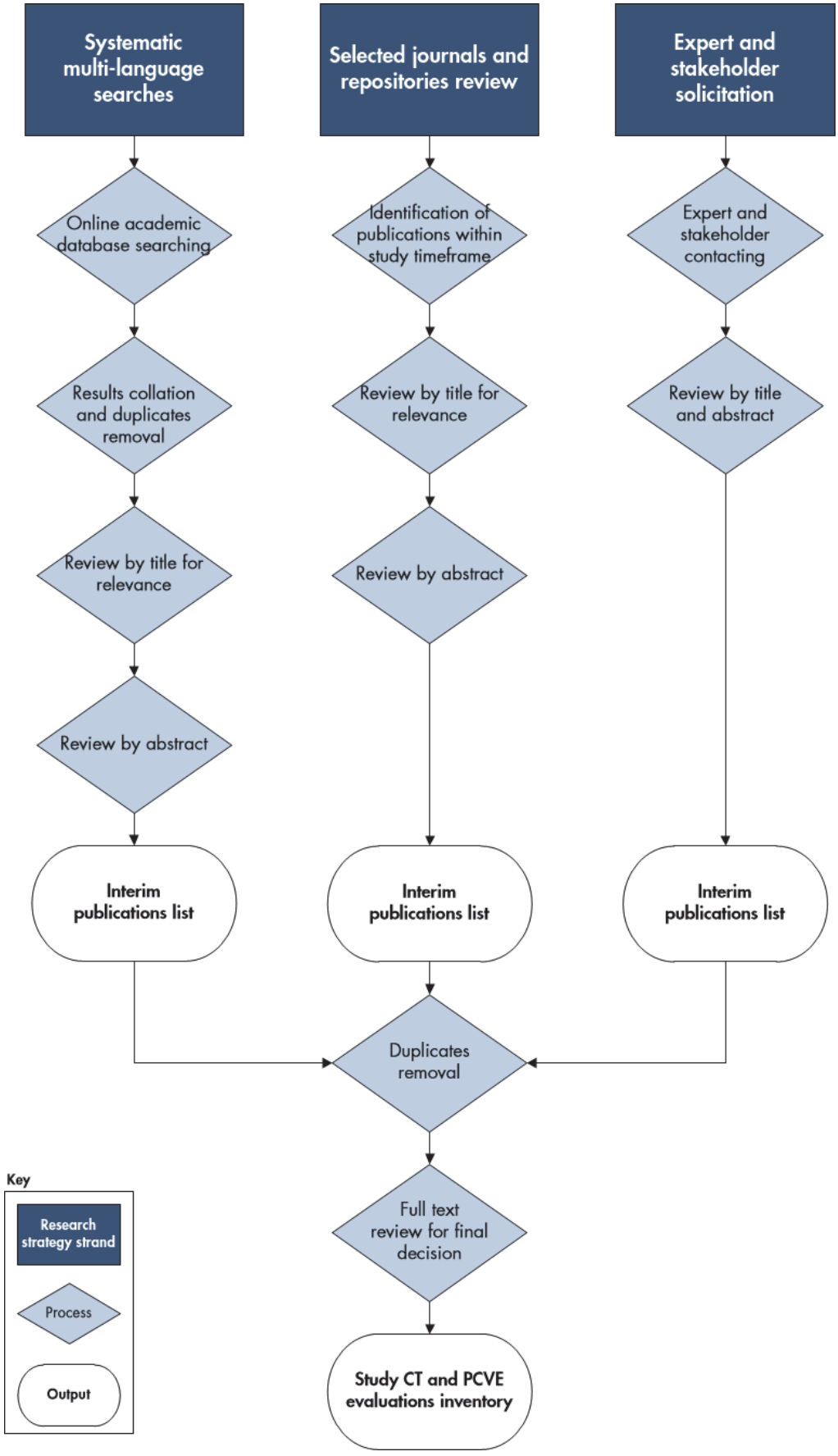
Following the systematic and targeted searches for CT and PCVE evaluation literature and the solicitation of manuscripts from experts, the study team reviewed the sources identified according to a multi-step process, with a view to finalising the study CT and PCVE evaluation inventory.

Figure 1.3 gives a schematic overview of the review process that led to the finalisation of the study's CT and PCVE inventory. The figure highlights how publications identified through the three different research strands pursued were reviewed and collated to produce the final inventory.

The following sections discuss the undertaking of the review and inventory production process. Further details about this process, including intermediate results from each of the inventory research and production phases, are available in Annex B.

[13] Davies et al. (2017).

Figure 2.3. Inventory production overview



Strand 1 – Reviewing results from systematic multi-language searches

The study team followed a series of steps to filter the search results to ensure they would meet the inclusion criteria for the inventory:

Step 0 – Collating results and removing duplicates. The bibliographic details of all publications identified through systematic searches were saved into a database that collated all results obtained from the different academic databases consulted, clustering them according to the four search languages that were employed. The study team then removed duplicate results encountered across all language clusters.

Step 1 – Review by title for relevance. In order to assess the relevance of publications identified to the scope of the study, all publications were first screened by reviewing their title.

Step 2 – Review by abstract for relevance. A second round of screening of results obtained through systematic searches was conducted by assessing the relevance of each publication's abstract according to the defined inventory inclusion criteria (see Section 2.1.1). In particular, this review identified sources that appeared to comply with the following conditions: (i) the source refers to a study in which a CT or PCVE strategy, policy, programme or intervention is evaluated; (ii) the CT or PCVE strategy, policy, programme or intervention being evaluated is underpinned by an explicit or implicit theory of change, and is evaluated using a qualitative or a quantitative evaluation approach that entails the collection and analysis of primary data to investigate a set of clearly defined evaluation questions; and (iii) the manuscript was written or published after 1 January 2013.

Strand 2 – Reviewing publications from selected journals and repositories

In parallel to the review by title and abstract of publications identified through systematic searches, a review of publications released since 1 January 2013 through the selected list of academic journals and grey literature repositories presented in Section 2.1.1.

Each publication from the journals and repositories selected for review was first screened for relevance by looking at its title. If this was deemed potentially relevant, the publication was also reviewed by abstract (consistent with the study inclusion criteria previously discussed). Publications that were deemed relevant or potentially relevant were downloaded and included in a separate, interim list of publications to be further assessed for inclusion in the study inventory. As was done for sources identified through systematic searches, publications were excluded only if it was possible to determine beyond doubt that they were not relevant to the scope of the present study.

Strand 3 – Reviewing submissions from experts and stakeholders

Lastly, as indicated in Section 2.1.1, the study team contacted a number of experts and stakeholders from the fields of CT, PCVE and evaluation to solicit unpublished or forthcoming manuscripts relevant to the scope of the study. While no unpublished manuscripts were received, a number of experts provided the details of publically available sources. Each publication received was screened for relevance by title and abstract.

Finalising the inventory

After the steps detailed above, the resulting groups of publications were collated into a single, interim inventory list of sources. At this stage, further duplicates were removed and a second screening of results

by abstract was conducted. Following this, the full text of each of the publications included in the interim inventory was reviewed to determine its relevance to the study. If the reviewer could not determine with certainty whether a publication should be included or not, a separate full-text review was conducted by a second reviewer. Consensus decisions were taken for these sources and the study team endeavoured to err on the side of inclusion.

For example, although primary data were not collected during evaluative efforts, two studies by Jordan (2014) and Carson (2017) assessing the outcomes of drone and targeted killing campaigns were included in the inventory as it was felt that their design and approach could contribute to shedding light on how the constraints around evaluating sensitive CT policies may be mitigated. Even though the specific policy approach investigated would not be implemented within the context of the national Dutch policy arena, the study team opted to include these studies with a view to identifying any lessons learned that may be transferable across or applicable to the fields of CT and PCVE. Similarly, a study by Williams et al. (2015) was included in the study inventory even though it did not refer to a single, specific PCVE intervention; instead it employed a robust evaluative approach to investigate PCVE dynamics in the context of a population exposed to different initiatives. Conversely, a study by Barkindo & Bryans (2016) on a prison-based intervention in Nigeria was excluded as this was not perceived to adopt a systematic evaluation approach, but rather to be based on non-structured observations, interactions and reflections.

More broadly, it is recognised that the inventory selection and production process as described above was to an extent arbitrary. The issues and phenomena investigated (i.e. PCVE, CT, evaluation) are complex and do not always lend themselves well to clear-cut definitions and inclusion or exclusion criteria. This was inevitably at odds with the basic premise of the study, which rests on a set of operational definitions and inclusion criteria that guided the work of reviewers.

The study team endeavoured to reduce ambiguity and the extent to which the review process may have led to the exclusion of relevant sources by adopting a transparent, multi-step and multi-strand review approach, which is outlined in more detail in Annex B. Throughout the review process sources were excluded only when it could be determined beyond doubt that they were not relevant to the scope of the work being conducted.

2.2. Development of an analytical framework

The development of the analytical framework to assess CT and PCVE evaluations was conducted in an iterative manner to allow for refinement through the use of expert consultation and validation activities.

Experts were asked to provide feedback on drafts of the analytical framework with a view to validating and finalising its content. In particular, members of the study SAC contributed through two iterative engagements to the reviewing and refining of the study analytical framework. Annex B provides further details as to the content and purpose of the approach to developing this framework.

2.3. Analysis of evaluations and reporting

The study team then analysed the CT and PCVE evaluations in the inventory through the lenses of the analytical framework in order to present overall findings, lessons learned and recommendations stemming

from the analysis in a synthesis report. To achieve this, the study team reviewed the sources in the repository, consulted experts, analysed the results and reported on the findings and conclusions.

2.3.1. Repository review

The research team analysed publications included in the study inventory presented in Annex A according to the analytical framework. To do so, the analytical framework was implemented as an analysis matrix in an electronic spreadsheet. In the analysis matrix, each column represented a data category from the framework and each row was used to capture data from an individual publication.

For each of the data categories of the analytical framework, reviewers analysing a publication were asked to complete two cells on adjacent columns. The first cell was designed to capture text or a summary of text extracted directly from the publication. The second cell was used by reviewers to provide a high-level synthesis of the value to be recorded for each data category. This was done to facilitate in subsequent phases an aggregate-level, quasi-quantitative analysis of the literature reviewed. Where possible, a list of pre-determined values to select from was included in the spreadsheet to ensure consistency. As part of this process, reviewers also extracted lessons learned concerning the evaluation as identified by the authors of the source being reviewed.

For example, with regard to evaluation methods, the spreadsheet-based implementation of the framework allowed researchers to capture information about the different data collection methods employed by extracting relevant excerpts from the publication in a dedicated cell. Researchers were asked to include all relevant details pertaining to data collection methods in this cell (e.g. number of interviews conducted, whether interviews were conducted remotely or in person, etc.). Next to the first cell, the spreadsheet provides a space for researchers to synthesise information about what data collection methods were used (e.g. interviews, focus groups, survey) in a synthetic manner, facilitating comparability across the inventory.

At the start of the extraction phase, each study team member involved with this strand of work was asked to review and extract two publications. The extractions were then reviewed by a second senior researcher who would validate the extraction and coding approaches taken, providing feedback in case any adjustment should be made. A total of four researchers conducted the review and extraction of the sources included in the study inventory.

Upon conclusion of the first review and extraction of all sources, a study team member reviewed and consolidated the coding prepared by different researchers across all sources reviewed, not least to clean data entries. Lastly, a second and final round of validation of data points extracted in relation to the evaluation approach, evaluation design and evaluation type categories for each publication was undertaken by a senior researcher.

2.3.2. Expert consultation and validation

Throughout the analysis phase and following its completion, the study team engaged with peer reviewers tasked with reviewing the study's work and outputs in the context of RAND's Quality Assurance system through a feedback and validation workshop. This introduced the reviewers to the study, provided them with an overview of its approach, activities and emerging results, and solicited their views on emerging

findings and recommendations. In addition, the SAC was consulted to discuss emerging results and steps required for the finalisation of the study's work.

2.3.3. Analysis and reporting

Upon conclusion of the repository review, the data extracted from the inventory's publications were analysed at an aggregate level. In particular, compatibly with and mindful of limitations stemming from the size of the sample available, data collected under different sections of the analytical framework were examined through descriptive statistics and cross tabulations. The analysis was first conducted by a member of the study team. Emerging results were then discussed through an internal workshop with other study team members and peer reviewers.

Two researchers conducted in parallel a thematic analysis of the lessons learned extracted from publications, coding them using a bottom-up approach. The two coding and analysis results were compared and discussed by the research team in order to come to a consensus as to the content and implications of findings stemming from the inventory reviewed.

Finally, the implications inferred by different study team members from the results of the analysis of the inventory's extractions and lessons learned helped to formulate the initial recommendations presented in the concluding chapters of this study. Initial recommendations were discussed through a series of internal meetings as well as during the internal study validation workshop with peer reviewers and during the final validation meeting with the project SAC.

2.4. Limitations

This section briefly outlines various limitations to the methodology and results of the work discussed in the report.

Systematic searches

The study team endeavoured to adopt a transparent, traceable, well-documented and repeatable process for the systematic searching and reviewing of CT and PCVE evaluations published over the last five years in English, Dutch, French and German. Search and review strategies and approaches employed were designed with a view to maximising results obtained and to ensure a sufficient degree of redundancy that would allow for as many relevant sources as possible to be identified and included in the study inventory. However, due to time and resource constraints, an element of compromise was required, in particular with regard to searches on academic databases:

1. Search strings were limited to a certain number of key terms across all languages employed to the detriment of other potentially relevant ones.
2. Only a certain number of academic databases could be investigated.
3. Search engines were set to review publications by title, abstracts and keywords only, rather than by full text.

As such, it is possible that a limited number of academic publications (which did not contain keywords used, were not indexed on the databases searched, or were not written in one of the four languages selected) may have been overlooked.

More broadly, only a limited number of non-academic repositories and websites could be reviewed in the framework of the study. As such, the study inventory building approach may be biased in favour of academic journal publications and may not be able to account for:

1. Publications and repositories belonging to institutions not selected for targeted reviews.
2. Publications released through quasi- or non-peer reviewed journals and outlets not indexed in academic databases and not selected for targeted reviews.
3. Non-academic publications written in languages other than those employed for the purpose of this study.

Expert consultation

In light of the project's timeframe and resource constraints, the study team's engagement was limited to a sample of domain experts and practitioners involved in CT, PCVE and evaluation activities.

Review process

As discussed in Section 2.1.4, it is recognised that the inventory selection and production process was to an extent arbitrary. The issues and phenomena investigated (i.e. PCVE, CT, evaluation) are complex and do not necessarily lend themselves well to clear-cut definitions and inclusion or exclusion criteria. Should this study be repeated, different research teams may come to different conclusions as regards the inclusion or exclusion of a range of sources. Bearing these limitations in mind, the authors trust the study to offer relevant and reliable insights on the status of CT and PCVE evaluations based on a review of the selected sources.

The study team endeavoured to reduce ambiguity and the extent to which the review process may have led to the exclusion of relevant sources by adopting a transparent, multi-step and multi-strand review approach. Furthermore, throughout the review process sources were excluded only when it could be determined beyond doubt that they were not relevant.

Finally, 11 sources included in the study inventory date from before January 2013 and do not comply with the study's inventory third inclusion criteria. These publications were brought to the attention of the study team by external experts and stakeholders contacted during the study's undertaking. In agreement with WODC and the SAC, the study team decided to include these publications, given that they had been highlighted from experts in the fields of CT, PCVE and evaluation. It should be noted, however, that the pre-2013 sample of publications included is neither exhaustive nor systematic, as publications from before January 2013 were not searched for and reviewed under other strands of the research protocol employed.

Limitations of the analytical framework

The analytical framework presented in Chapter 4 of this report has been designed with a view to accommodating the specific purpose, needs, requirements and constraints of the present study. In particular, the framework aims to analyse in a comprehensive and homogeneous manner a broad set of

evaluation reports pertaining to different types of initiatives, spanning from high-level strategies to grassroots programmes, and covering the entire spectra of CT and PCVE work. For the framework's overarching taxonomy and for the values envisioned under each analytical category to be designed, this entailed finding a balance between comprehensiveness and retaining a manageable and comparable level of detail that would ensure the study's capacity to offer meaningful, aggregate-level analysis of the evaluations in the inventory.

Furthermore, the analytical framework has been designed with a view to conducting the review, assessment and extraction of lessons learned from evaluation manuscripts based exclusively on data contained in them. Additional review of underpinning primary data or consultation activities with evaluation implementers, beneficiaries or practitioners was not feasible within the context of this study.

As a consequence, the analytical framework is limited in relation to the information it seeks to capture, particularly as regards the quality of the evaluation's design and conduct. This is in recognition of the fact that detailed data pertaining to evaluation design and implementation activities are unlikely to be available in short academic manuscripts and quasi-peer reviewed evaluation reports. Only a limited assessment of the quality and robustness of evaluations conducted was thus possible. The analytical framework would need to be significantly expanded in the event that it was to be used to perform an in-depth critical appraisal of CT and PCVE evaluations for which access to evaluation implementers and beneficiaries to collect primary data was possible. Further details about how the analytical framework could be expanded are discussed in Chapter 4 of this document.

Fundamental rights compliance

It was beyond the scope of this study to advise on what constitutes acceptable or suitable CT and PCVE strategy, policy or programming and to assess whether initiatives being evaluated and their evaluations complied with human, civil and any other fundamental rights. The inclusion of any particular CT or PCVE initiative in the study's inventory should not be taken as an endorsement of such practices by the study team, nor an indication that an assessment of the suitability, relevance or fundamental rights compliance of such initiatives has been undertaken in the context of this study.

Ethical considerations and human rights form a basis for sound evaluations. In the Netherlands and in a broader European context all evaluations should be designed and conducted in accordance with the rights and principles set out in the European Convention on Human Rights and the Charter of Fundamental Rights of the European Union. Compliance of CT and PCVE strategies, policies and interventions with human and civil rights should be assessed as part of an evaluation, alongside other relevant ethical issues.[14]

[14] Marret et al. (2017).

3. Understanding the study context and its key definitions

This chapter reports on the results of the initial targeted literature search. It provides an overview of the context of the study, discussing how CT policy and approaches have evolved in recent years and what implications this has for the study at hand. The chapter discusses how some of the key concepts underpinning the study research have been framed in policy and academic literature over the years and concludes by presenting a set of key definitions that were used to operationalise the study's inventory inclusion criteria.

3.1. Understanding the context

3.1.1. The rise and evolution of CT approaches

Since the 9/11 terrorist attacks in 2001 in the United States, CT has been a key policy concern for Western countries. While at the turn of the century the initial focus and response of CT activities was on threats and groups emanating from abroad, terrorist attacks such as the Madrid and London bombings of 2004 and 2005 highlighted the rising issue of home-grown terrorism and extremism.[15] These events heralded a shift in CT activities, broadening their focus to include both external and internal threats and risks. Currently, the terrorist threat faced by European countries emanates with different intensity and magnitude from a wide range of ideological, religious and political sources. In 2016 alone, eight EU member states – among them the Netherlands – reported a total of 142 failed, foiled and completed terrorist attacks that resulted in 142 victims, 379 injured and 1,002 individuals arrested for terrorist offences. While a number of terrorist plots have reportedly been attempted or foiled in the Netherlands in recent years, the last known death caused by terrorism in the country was that of movie director Theo van Gogh in 2004.[16]

In this context, governments, policy- and decision-makers have come under pressure to address terrorism and violent extremism threats, with a view to preventing terrorist attacks from occurring. Initially, CT policy in many Western countries focused on coercive approaches, relying on the role and activities of law enforcement and security agencies to prevent and tackle terrorist attacks before they occurred.[17] For example, the 2002 Patriot Act in the United States granted the government broader powers, including increased scope for use of surveillance of suspected terrorists, search warrants, and enhanced maximum

[15] European Parliament (2017).

[16] Europol (2017).

[17] Romaniuk & Chowdhury Fink (2012).

penalties for various crimes committed by terrorists.[18] Similarly, the UK introduced Prevention of Terrorism Acts in 2005 and 2006, expanding the use of search warrants and increasing maximum penalties for crimes relating to possession for terrorist purposes, offences relating to nuclear material, and the contravention of notices relating to encrypted information.[19]

More recently, however, policy-makers and CT practitioners have recognised the importance of adopting broader, holistic approaches to preventing and tackling terrorism activities. These typically seek to build resilience to violent extremist ideologies at the individual and community level, and to address the root causes of violent radicalisation leading to terrorism. Such initiatives are broadly encompassed under the label of PCVE and are now considered an integral part of European CT policies and approaches (in addition to more traditional CT measures). For example, one of the four strands of work of the *EU Counter-Terrorism Strategy* focuses on 'prevent[ing] people turning to terrorism by tackling the factors or root causes which can lead to radicalisation and recruitment, in Europe and internationally.'[20]

3.1.2. The Dutch approach to CT and PCVE

Over the course of the last two decades, Dutch authorities have designed and implemented a wide array of CT legislations, policies and protocols. In accordance with the trajectories of many other countries, the Dutch initially relied on traditional security measures, before shifting towards a more holistic approach to CT.

Similarly to the US and elsewhere in Europe,[21] in the early 2000s the Dutch CT policy response focused on (i) facilitating the criminal prosecution of all stages of terrorist activities; (ii) implementing administrative measures aimed at curtailing breeding grounds for terrorism; and (iii) developing terrorism crisis management tools and capabilities. For example, this entailed a qualitative and quantitative reform of bodies tasked with combatting terrorism and the development of improved surveillance and security systems and capabilities.[22]

Dutch CT policy was first formulated into a coherent strategy with the National Counterterrorism Strategy (Nationale Contraterrorismestrategie) for 2011–2015.[23] This focused on a comprehensive approach to CT and aimed to both prevent and repress terrorism and violent extremism activities, employing a variety of security- and society-based measures.[24] Following this, a revised CT strategy was launched by Dutch authorities for the 2016 to 2020 period (Nationale Contraterrorismestrategie 2016–2020). This adapts the 2011 scope and approach to tackle some of the criticisms and shortfalls identified

[18] US Department of Justice (2017).

[19] UK Government (2017).

[20] Council of the European Union (2005).

[21] Byman (2007); Van Dongen (2010).

[22] Demant & De Graaf (2010).

[23] NCTV (2011).

[24] Manjikian (2017).

during the evaluation of the first strategy's implementation.[25] The revised 2016–2020 strategy focuses on five areas for intervention:[26]

1. **Procure:** designed to timely gather and assess intelligence about (potential) threats to Dutch national security and national interests abroad.
2. **Prevent:** designed to prevent and disrupt extremism and to foil terrorist attacks before they occur.
3. **Protect:** designed to protect people, property and vital processes from extremist and terrorist threats (both physical and virtual).
4. **Prepare:** designed to prepare optimally for extremist and terrorist violence and its consequences.
5. **Pursue:** designed to enforce the law in the face of extremism and terrorism.

The 2016–2020 strategy recognises the need for a comprehensive, multi-level approach that is threat-based and respects the rule of the law with both fixed and flexible measures.[27]

In parallel to the development of the updated 2016–2020 strategy, a wide range of PCVE programmes was developed and launched in the Netherlands, including initiatives that are both prison- and community-based. For example, Team TER (Terrorists, Extremists and Radicals) is a programme of the Dutch Probation Service. It employs 13 internationally trained probation officers who ensure a tailor-made probation approach to disengage individuals from extremism, carried out in close cooperation with a variety of partners including the judiciary, police and municipal authorities. The Dutch National Police PCVE programme, Allies, involves engaging with police officers and key figures in the community to ensure a proactive approach after an incident occurs.

Current context

There have been no recent terrorist attacks in the Netherlands comparable in size, scope or frequency to those witnessed in other European countries, such as Belgium, France and the UK. Dutch policy-makers' concerns have been raised, however, in light of the phenomena of radicalisation and of so-called foreign terrorist fighters (FTFs).[28] According to a definition proposed by the United Nations, FTFs are 'individuals who travel to a State other than their State of residence or nationality for the purpose of the perpetration, planning or preparation of, or participation in, terrorist acts or the providing or receiving of terrorist training, including in connection with armed conflict.'[29]

In the Netherlands, as well as in Europe, the phenomenon of FTFs has been closely intertwined, although not exclusively, with that of violent jihadism.[30] Terrorist jihadist groups active across the Middle East,

[25] Manjikian (2017).

[26] NCTV (2016a).

[27] NCTV (2016b).

[28] Van Ginkel & Entenmann (2016).

[29] United Nations Security Council (2014).

[30] Van Ginkel & Entenmann (2016).

North Africa and Asia have been increasingly capable of convincing new personnel, through their networks and propaganda channels, to travel to conflict areas to join and support their activities.[31]

In response to this, in August 2014 Dutch authorities launched a Comprehensive Action Programme to Combat Jihadism.[32] This programme focuses on tackling jihadist radicalisation and FTFs through both disruptive measures, aimed at limiting the ability to travel of potential terrorist and recruits, and preventive ones, aimed at preventing the risk of and stifling the channel facilitating radicalisation.[33]

While the flow of new recruits to conflict areas has diminished over time, in recent months growing concerns have been expressed by European policy-makers and practitioners at the potential security threat posed by FTFs returning to their countries of origin and residence, including the Netherlands.[34] Figures provided by Dutch authorities in March 2018 put the number of Dutch FTFs who travelled to Iraq, Syria or other conflict areas to join terrorist organisations at around 300. Of these, to date 60 have reportedly been killed, 160 remain abroad for jihadist purposes and 50 have returned to the Netherlands.[35]

3.1.3. Evaluating CT and PCVE policies and interventions

While the number and volume of CT and PCVE policies and programmes implemented across Western countries has rapidly increased, the political imperative and speed with which they are implemented has meant that such programmes are rarely evaluated and are often designed on the basis of untested assumptions. In this policy area, robust and rigorous evaluations assessing effectiveness have not yet become the norm.[36]

Evaluation is an assessment method that allows policy-makers to investigate whether an intervention, programme or policy is actually delivering the expected results, and how much it has contributed to achieving these results. In the fields of CT and PCVE, evaluation may help policy-makers and practitioners understand: (i) if and how results are being achieved, for whom, and under what conditions; (ii) how better outcomes could be achieved; (iii) how to achieve the same outcomes for less (or obtain better value for money); and (iv) identify any changes that might be necessary to keep up with the dynamic contexts and evolving needs and objectives in these fields.

While no consensus has emerged yet on the evaluative approach to be employed in the fields of CT and PCVE, policy-makers, practitioners and stakeholders have increasingly recognised the necessity for embedding evaluation practice more thoroughly. For example, in 2011 the EU Counter-Terrorism Coordinator called for a systematic investigation into successful and unsuccessful practices, lessons learned and an analysis of why certain procedures have or have not worked in the CT context.[37] Similarly, the

[31] Weggemans et al. (2014); Van Ginkel & Entenmann (2016); European Parliament (2017).

[32] NCTV (2014).

[33] NCTV (2014).

[34] NCTV (2016b; 2018); European Parliament (2017).

[35] NCTV (2018).

[36] Gielen (2017); Feddes & Gallucci (2015).

[37] European Parliament (2017).

2015 European Parliament resolution on the prevention of radicalisation and recruitment of European citizens by terrorist organisations calls on the European Commission:[38]

> *(…) to establish as a priority an action plan to implement and evaluate the EU strategy for combating radicalisation and recruitment to terrorism, on the basis of the exchange of best practice and the pooling of skills within the European Union, the evaluation of measures undertaken in the Member States and cooperation with third countries and international organisations, on a basis of full respect for international human rights conventions and through a multi-stakeholder and multi-sectoral participative and consultative approach.*

A number of programmes and initiatives have been launched in recent years in response to these calls. For example, the European Commission established the Radicalisation Awareness Network (RAN) in 2012. This initiative brings together practitioners from across Europe to share their knowledge and experience and to develop effective programmes tackling violent radicalisation and extremism leading to terrorism.[39] Similarly, between 2014 and 2017, a consortium led by RAND Europe delivered IMPACT Europe,[40] a project investigating methods and approaches for evaluating PCVE programmes and policies.[41]

3.2. The definitional challenge

As described above, recent years have seen CT and PCVE become policy priorities across the EU and the Netherlands. This has not been without its challenges as both CT and PCVE posit complex, multi-faceted challenges that touch and span over a range of equally complex social phenomena, such as issues connected to migration, belonging, minorities and marginalisation. This is further exacerbated by the lack of an evidence-based understanding as to what constitutes effective CT and PCVE practice and the lack of a culture of evaluation within these fields.[42]

The following section explores the academic and policy debates around some of the key concepts and definitions that underpin the study research, which were briefly introduced in the previous section. The chapter concludes by presenting the set of definitions that was employed to operationalise the inclusion criteria guiding the production of the study's CT and PCVE evaluations inventory.

3.2.1. Defining terrorism, CT and PCVE

A wealth of publications has debated the issue of defining terrorism, a key concept around which both CT and PCVE revolve.[43] Although 'terrorism' has been described as one the most important words in the contemporary political vocabulary, agreement on a shared definition has proved elusive within both

[38] European Parliament (2015).

[39] European Commission (2017).

[40] For further information, see: http://impacteurope.eu/

[41] Feddes & Gallucci (2015); IMPACT Europe (2017b).

[42] Vermeulen (2014); Gielen (2017); Marret et al. (2017).

[43] Wensink et al. (2017).

academic and policy circles.[44] The complex and multidimensional nature of the terrorist phenomenon and the promiscuous use of the 'terrorist' label to describe a wide range of manifestations with a variety of goals and purposes has even led some to question the need for describing the phenomenon through a unitary definition.[45] One answer to this challenge has been to develop multiple lenses and prisms through which to look at the concept of terrorism in an effort to define it in a comprehensive and inclusive manner.[46]

In the context of the present study, while resisting the temptation of adopting a purely reductionist approach, a flexible definition of terrorism has been devised with a view to building inclusive and flexible definitions of CT and PCVE around it. In light of the growing variety of CT and PCVE measures and approaches being developed and implemented by different actors, the study's working definition of terrorism retains a sufficient degree of breadth and to accommodate as broad and diverse a set of CT and PCVE interventions' evaluations as possible in the study's inventory.[47]

The study refers to the concept of terrorism by employing the 'academic consensus definition' produced by Schmid & Jongman (1988), which defines terrorism as:[48]

> *(…) an anxiety-inspiring method of repeated violent action, employed by (semi-) clandestine individual, group, or state actors, for idiosyncratic, criminal or political reasons, whereby – in contrast to assassination – the direct targets of violence are not the main targets. The immediate human victims of violence are generally chosen randomly (targets of opportunity) or selectively (representative of symbolic targets) from a target population, and serve as message generators. Threat- and violence-based communication processes between terrorist (organisation), (imperilled) victims, and main targets are used to manipulate the main target (audience(s)), turning it into a target of terror, a target of demands, or a target of attention, depending on whether intimidation, coercion, or propaganda primarily sought.*

This definition encapsulates the one currently employed in the Dutch National CT strategy and provides greater clarity around some of the attributes, purposes and means that may be considered relevant for shaping a holistic CT and PCVE policy response.[49]

On the basis of the above definition, CT can be seen as a complex set of strategies, policies and programmes designed to take direct action against terrorists or their sponsors and supporters. CT can include both proactive and passive measures that are designed and implemented across a wide array of fields and domains, including those of politics, law, finance, communications, defence, intelligence and infrastructure. CT may engage and be implemented by a variety of actors, including government, law enforcement, the military, private sector and civil society actors, and the general public.[50]

---

[44] Weinberg et al. (2010).

[45] Schmid (2004a); Ramsay (2015).

[46] Schmid (2004b).

[47] Vermeulen (2014); Wensink et al. (2017).

[48] Schmid (1988).

[49] NCTV (2016a).

[50] Arce M. & Sandler (2005); Hofman (2002); Wensink et al. (2017).

As discussed in Section 3.1.1, recent years have seen the development of the relatively novel domain of PCVE. Broadly speaking, PCVE stems from the growing recognition of the importance of adopting holistic approaches to preventing and tackling terrorism through activities going beyond traditional CT measures and tackling root causes and conditions conducive to terrorism and political violence.[51] Furthermore, PCVE revolves around the relational concepts of radicalisation and extremism.[52] As in the case of terrorism, these are terms for which no agreed-upon definition exists and that have come under significant scrutiny due to what has been perceived as their use in policy circles to justify the mushrooming and opening of new areas of intervention for political economies of danger, governance, and security practices.[53]

Keeping the above caveats in mind, extremism can be seen as a phenomenon whereby individuals or groups who are motivated by a certain ideology distance themselves from moderate, mainstream or status quo positions, reject pluralism and, disregarding the rule of law, engage in serious criminal behaviour or take actions that undermine the democratic legal order.[54] Building on this, radicalisation can broadly be seen as a process by which an individual comes to hold extremist views and ideologies, which may also lead to supporting terrorism and other forms of violent extremism leading to terrorism.[55] Schmid (2013) emphasises that radicalisation is a process that thrives in the presence of a situation of polarisation or seeping conflict, defining it as:[56]

> *an individual or collective (group) process whereby, usually in a situation of political polarisation, normal practices of dialogue, compromise and tolerance between political actors and groups with diverging interests are abandoned by one or both sides in a conflict dyad in favour of a growing commitment to engage in confrontational tactics of conflict-waging. These can include either (i) the use of (non-violent) pressure and coercion, (ii) various forms of political violence other than terrorism or (iii) acts of violent extremism in the form of terrorism and war crimes. The process is, on the side of rebel factions, generally accompanied by an ideological socialization away from mainstream or status quo-oriented positions towards more radical or extremist positions involving a dichotomous world view and the acceptance of an alternative focal point of political mobilization outside the dominant political order as the existing system is no longer recognized as appropriate or legitimate.*

This study employs an ad hoc definition of PCVE, building on those previously produced in academic and policy circles.[57] The adopted definition sees PCVE as a complex set of strategies, policies and programmes designed to strengthen the resilience of individuals and groups to the appeal of radicalisers, extremists and terrorists by preventing individuals and groups from completing a process of radicalisation and mobilising to commit violence, and by disengaging and deradicalising individuals and groups who are

[51] Schmid (2013); Marsden (2017); RAN (2017).

[52] Vermeulen (2014); RAN (2017).

[53] Heath-Kelly (2013); Schmid (2013); RAN (2017); Marret et al. (2017).

[54] Schmid (2013); NCTV (2016a).

[55] Schmid (2013); UN (2014); Marsden (2017); RAN (2017).

[56] Schmid (2013).

[57] UN (2014); Marsden (2017); RAN (2017).

planning to commit or have already engaged in extremist and terrorist violence. As in the case of CT, PCVE can include both proactive and passive measures designed and implemented across a wide array of fields and domains, including politics, law, education, communications, public health and social work. PCVE may engage and be implemented by a variety of actors, including government, law enforcement, the military, private sector and civil society actors, and the general public.

3.2.2. Defining evaluation

Finally, it is important to clarify the concept of evaluation in the context of this study. Evaluation is understood as a systematic determination of the quality or value of something.[58] However, there is no single and commonly agreed definition of evaluation and various interpretations exist in parallel. Perhaps the most well-known and cited definition is the one developed by the OECD:[59]

> *Evaluation is the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. Evaluation also refers to the process of determining the worth or significance of an activity, policy or program.*

This definition and the evaluation criteria against which a programme should be assessed have been employed by a number of international organisations and bodies, such as the World Bank[60] and the International Labour Organisation.[61] Also in the European context, the European Commission defined evaluation as an evidence-based judgement of the extent to which an intervention has:

- Been effective and efficient
- Been relevant given the needs and its objectives
- Been coherent both internally and with other EU policy interventions
- Achieved EU added-value. [62]

Evaluation is often explained and reinterpreted at the national level, for example in guiding documents for policy-makers and government officials. The Magenta Book is such an example in the UK, and it introduces evaluation as 'an objective process of understanding how a policy or other intervention was implemented, what effects it had, for whom, how and why.'[63] Leeuw (2009) explains changes in how evaluation was viewed in the Netherlands and points to the 1990s, when the Dutch government started to establish, formalise and truly cultivate its evaluation policy.

Looking at these various definitions, a number of common themes can be identified:

[58] Scriven (1991).

[59] OECD DAC Network on Development Evaluation (2010).

[60] Independent Evaluation Group – World Bank (2007).

[61] International Labour Organization (2017).

[62] European Commission (2015).

[63] HM Treasury (2011).

1. Evaluation is a form of assessment that involves a value judgement
2. This assessment is systematic, objective and evidence-based
3. Evaluation is comprehensive as it looks across an initiative's design, implementation and results and/or effects.

For the purpose of this study we draw on the concept of evaluation established by the OECD and expand it in two dimensions. Firstly, we expect that an intervention that is evaluated should be underpinned by an explicit or implicit theory of change. Secondly, we expect that an evaluation should entail the collection and analysis of primary data to investigate a set of clearly defined evaluation questions.

The reason for introducing these modifications is twofold. Firstly, a good understanding of the intervention is critical for a successful evaluation. Evaluations often draw on or reconstruct an intervention logic or a theory of change that presents the essential elements of an intervention, help structured thinking about it and make explicit the underlying assumptions about how it is expected to work.[64] Secondly, while there are many useful PCVE studies that may cover some of these elements (e.g. value judgement, assessment of effects) not all contain sufficient information about how systematic the assessment was and what evidence supports their claims to qualify as them as evaluations.[65]

3.3. Conclusions

This chapter has provided an overview of the context in which the study took place, and has detailed some of the key challenges associated with defining the key concepts underpinning the study's research and the production of the study's CT and PCVE evaluations inventory. Table 3.1 provides a concise overview of the definitions employed throughout the study to guide the operationalisation of the inventory's inclusion criteria.

[64] Rogers et al. (2000).

[65] Van Hemert et. al. (2014).

Table 3.1. Definitions of key project concepts guiding the operationalisation of the study's inventory inclusion criteria

Concept	Definition
Terrorism	An anxiety-inspiring method of repeated violent action, employed by (semi-) clandestine individual, group or state actors, for idiosyncratic, criminal or political reasons, whereby – in contrast to assassination – the direct targets of violence are not the main targets. The immediate human victims of violence are generally chosen randomly (targets of opportunity) or selectively (representative of symbolic targets) from a target population, and serve as message generators. Threat- and violence-based communication processes between terrorist (organisation), (imperilled) victims and main targets are used to manipulate the main target (audience(s)), turning it into a target of terror, a target of demands or a target of attention, depending on whether intimidation, coercion or propaganda is primarily sought.
Counterterrorism (CT)	A complex set of strategies, policies and programmes designed to take direct action against terrorists or their sponsors and supporters. CT can include both proactive and passive measures designed and implemented across a wide array of fields and domains, including those of politics, law, finance, communications, defence, intelligence and infrastructure. CT may engage and be implemented by a variety of actors, including government, law enforcement, the military, private sector and civil society actors, and the general public.
Preventing and countering violent extremism (PCVE)	A complex set of strategies, policies and programmes designed to strengthen the resilience of individuals and groups to the appeal of radicalisers, extremists and terrorists by preventing individuals and groups from completing a process of radicalisation and mobilising to commit violence, and by disengaging and deradicalising individuals and groups who are planning to commit or have already engaged in extremist and terrorist violence. PCVE can include both proactive and passive measures designed and implemented across a wide array of fields and domains, including those of politics, law, education, communications, public health and social work. PCVE may engage and be implemented by a variety of actors, including government, law enforcement, the military, private sector and civil society actors, and the general public.
Evaluation	The systematic and objective assessment of an ongoing or completed strategy, policy, programme or intervention and of its design, implementation and results. The aim of an evaluation is to determine the relevance and fulfilment of the objectives, development efficiency, effectiveness, impact and sustainability of a strategy, policy, programme or intervention that is underpinned by an explicit or implicit theory of change. An evaluation should entail the collection and analysis of primary data to investigate a set of clearly defined evaluation questions.

4. Building an analytical framework

This chapter presents the analytical framework developed to characterise the sources gathered in this study's inventory of CT and PCVE evaluations. The chapter is structured along the main framework components:

- **Evaluation methodological characteristics** (Section 4.1). Categories capturing an evaluation's purpose, methods (approach, design, methods), scope (timing, duration, geographic scope) and evaluation criteria.
- **Evaluation quality characteristics** (Section 4.2). Categories capturing information about the quality of the evaluation. The components of this section have been selected with a view to adopting an approach to quality assessment based on transparent metrics and the information available.
- **Initiative characteristics** (Section 4.3). Categories capturing information about the strategy, policy, programme or initiative being evaluated, including its nature, purpose, type of activities, group and unit of focus, status, scope (temporal and geographic) and duration.
- **Lessons learned** (Section 4.4). An overview of the approach to capturing lessons learned.

Annex D presents the study analytical framework in greater detail, and includes key definitions and values employed by the study team when analysing sources included in the study inventory.

4.1. Evaluation methodological characteristics

4.1.1. Theoretical underpinnings

There is a rich body of literature concerned with different types and forms of evaluation. These typologies, and the categories or models they offer, depend on different aspects of evaluations, such as the purpose they serve, the way they are implemented, the audiences they aim to target, the methods they draw on, or the ways in which they are supposed to be used.[66] However, there is neither agreement on a single taxonomy, nor on the variables that such a taxonomy should take into account.

Among the most commonly recognisable typologies is the dichotomy between formative and summative evaluations introduced by Scriven (1967). Formative evaluation aims to provide, during the phases of development and implementation of a new programme, data that permit successive adaptations.

[66] See, for example: Alkin & Christie (2004); Contandriopoulos & Brousselle (2012); Frye & Hemmer (2012); Shadish et al. (1991).

Conversely, summative evaluation aims to provide a final assessment of whether the results of a programme or intervention met its stated goals.[67] Further details about this commonly used typology are briefly summarised in Table 4.1.

Table 4.1. Characteristics of formative and summative evaluation

Characteristics	Formative evaluation	Summative evaluation
Purpose	'*Need to improve*': quality assurance; improvement	'*Call to prove*': provide an overall judgement
Use	Guide for decision-making	Determining accountability for successes and failures
Target audience	Managers, staff, practitioners (insiders)	Funders, policy-makers, consumers (outsiders)
Methods	Qualitative and quantitative	Emphasis on quantitative
Focus of data collection	Clarification of goals, implementation, outcomes	Implementation issues, outcome measures
Orientation	Prospective and proactive	Retrospective and reactive

SOURCE: Based on Stufflebeam & Shinkfield (2007) and Clarke (1999).

Possibly the most comprehensive taxonomy of evaluation is offered by Stufflebeam & Shinkfield (2007), who differentiate between:

- Pseudo-evaluations, including: public relations studies, politically controlled studies, pandering evaluations, evaluation by pretext, empowerment under the guise of evaluation, and customer feedback evaluation.
- Quasi-evaluation studies: objectives-based studies, the success case method, outcome evaluation as value-added assessment, experimental and quasi-experimental studies, cost studies, connoisseurship and criticism, theory-based evaluation, and meta-analysis.
- Improvement- and accountability-oriented evaluation approaches, including: decision- and accountability-oriented studies, consumer-oriented studies, including formative and summative evaluation, amateur versus professional evaluation, intrinsic and payoff evaluation, goal-free evaluation,[68] and accreditation and certification.
- Social agenda and advocacy evaluation approaches, including: responsive or stakeholder-centered evaluation,[69] constructivist evaluation, deliberative democratic evaluation, and transformative evaluation.

[67] Scriven (1967).

[68] Scriven (1967).

[69] Stake (1983).

- Eclectic evaluation approaches, including: utilization-focused evaluation,[70] and participatory evaluation.

However, there is no commonly accepted terminology and terms such as 'approaches', 'models' or 'theories' are not always clearly defined and, more often than not, they are used interchangeably by authors. This can easily result in confusion, and it hampers the application of a single taxonomy for the purpose of this study. While the taxonomy proposed by Stufflebeam & Shinkfield (2007) is relatively comprehensive and detailed, our study required a framework that would be simpler and easier to use, and yet would capture the variables and dimensions according to which the evaluation studies included in this review were to be described. With this in mind, we developed a customised section of the framework focusing on broadly defined methodological characteristics to guide the characterisation of the sources in the inventory and the further analysis of the latter.

4.1.2. The framework – Section 1

The first section of the analytical framework comprises a set of categories capturing information about an evaluation's purpose, methods, scope and criteria. In particular, it focuses on:

- **Purpose**: the main reason for which an evaluation is undertaken or the main type of question that the evaluation aims to address. The Magenta Book (HM Treasury 2011) distinguishes three broad classes of question that evaluation might be used to answer: (i) how was the policy delivered (process evaluation)? (ii) what difference did the policy make (impact/outcome evaluation)? And (iii) did the benefits of the policy justify the costs (economic evaluation)? While these categories are not mutually exclusive (an economic evaluation study may also consider impacts and/or outcomes), we used this high-level categorisation as a starting point. When classifying studies (e.g. as 'impact' or 'process' evaluations), the study team relied on explicit statements in the manuscripts to this effect or, in the absence of these, the team relied on the wording of the aims of the study analysed to determine an evaluation's purpose. For example, whenever the aims of a study referred to results, effects and impacts, it was classified as an 'impact evaluation'. Whether a study that declares an interest in the results or impacts of an initiative adopts an appropriate design to investigate these is a separate issue that pertains to the quality and robustness of evaluation design choices made, rather than to the analysis conducted.
- **Conceptual approach**: the strategy and tools used to implement an evaluation, including an analysis of the overarching evaluation approach, design, data collection methods and analysis methods. In addition to the high-level categorisation ('purpose'), some evaluation studies adopt and follow specific conceptual approaches ('models') that are well-defined in the evaluation literature. These include – but are not limited to – theory-based evaluations,[71] realist evaluations,[72] participatory evaluations,[73] gender-sensitive evaluations[74] and so on. While not all

[70] Patton (1997; 2008).

[71] See: Weiss (1995); Chen & Rossi (1983).

[72] Pawson & Tilley (1997).

[73] Cousins & Earl (1992).

studies may use or report using these approaches, all evaluations need a design that outlines how data will be collected and analysed.

- **Evaluation design**: evaluation designs outline how the evaluation will address the question of attribution (i.e. whether the desirable changes – if achieved – have been caused by the intervention alone, or whether other factors played a role). In their handbook on the design of social science research, Shadish et al. (2002) explain there are three critical conditions that must be satisfied in order to claim causal interference: (i) the hypothesized cause must precede its anticipated effect in time; (ii) if the levels of the cause differ in some systematic way, then there must also be corresponding variation in the effect; and (iii) all other plausible explanations – other than the anticipated causal one – must be eliminated before a link can be established between the hypothetical cause and effect. The most convincing approach to an evaluation that can infer causality between the intervention and its outcomes is to conduct an experiment. Shadish et al. (2002) define an experiment as 'an empirical investigation in which the levels of a potential cause are manipulated by an outside agent functioning independently of the participants in the research, and after which the consequences for an important outcome are measured.' We follow Murnane & Willett (2010) in distinguishing between two types of experiment: randomised experiments (or simply experiments) and quasi-experiments. Randomised experiments are those 'in which units are assigned to experimental conditions by a random process, such as the toss of a fair coin.' Quasi-experiments are designs in which units are not assigned to conditions randomly; in the difference-in-differences method, for example, which compares a treatment and comparison group before and after treatment, participants are not selected randomly. However, the generated results should make it possible to make legitimate causal inferences.[75] Furthermore, for the purpose of this study, we consider so-called natural experiments to be a sub-set of quasi-experiments.[76] We consider all other designs non-experimental, including before-and-after assessments conducted without a comparison group, for example. This typology also draws on the guide to scoring methods in the Maryland Scientific Methods Scale (MSMS).[77] More specific descriptions of the evaluation design were derived from the manuscripts. When coding the study design of a given evaluation, the study team either relied on the presence of explicit statements about the design chosen or, in the absence of these, used available descriptions of methods to discern the likely design used and to categorise it in line with the study's analytical framework. In theory, the evaluation design should match the purpose of an evaluation. For example, an impact evaluation should seek designs that would allow causal

[74] Sielbeck-Bowen et al. (2002).

[75] Murnane & Willet (2010).

[76] Murnane & Willet (2010) describe natural experiments as experiments in which individuals are (randomly) assigned to potential 'treatment' and 'control' groups not by researchers' randomisation, but rather by an external agency (e.g. based on their birth date or last name), by particular geographic conditions, or by other natural conditions (e.g. a natural disaster). The challenge for the researcher is to recognise such natural experiments when they occur and to translate them into a study design.

[77] Madaleno & Waights (2014).

inferences to be claimed. However, using a research design inadequate to the purpose of the evaluation is not uncommon in evaluation practice.

- **Evaluator**: this category includes details pertaining to the status of the evaluation implementer vis-à-vis the initiative being implemented, in particular whether the evaluator is independent from the policy or programme under evaluation (external), or not (internal). Discussion of this difference features prominently in the evaluation literature.[78]
- **Timing**: where does the evaluation sit in relation to the life cycle of the initiative being evaluated? The evaluation could take place prior to its implement (*ex ante*), at a predefined interim milestone (interim), after the programme has been completed or the policy is no longer active (*ex post*), or throughout the life cycle of the initiative (embedded or ongoing). This aspect is, for instance, clearly prescribed in the evaluations of funding programmes in the EU, where studies are categorised as taking place before the implementation started, during the programme, or after it has been completed.[79]
- **Subject scope**: the Magenta Book (HM Treasury 2011) asks one to consider whether an evaluation should cover the entire intervention or if it could be limited to 'areas, impacts or processes where knowledge is most uncertain.'[80] We have distinguished three possible aspects: material (i.e. related to the scope of the policy being evaluated), temporal and geographical (see below).
- **Temporal scope**: the proportion of the policy or intervention life cycle that is being examined. Evaluations may cover part of a policy's life cycle, its entire life cycle, as well as multiple subsequent policies.
- **Geographical scope**: the geographical scope of the evaluation, which can cover aspects at local, regional, national or transnational level.
- **Evaluation criteria**: the standards against which the initiative being assessed is evaluated, which could include aspects such as relevance, effectiveness, efficiency, impact, sustainability, etc. The OECD provides an elaborate and often-cited list of possible evaluation criteria.[81]
- **Evaluation questions**: evaluations are often carried out to provide answers to a series of questions. Has the initiative succeeded in achieving its original objectives? Or is the programme offering value-for-money? This category captures the extent to which such evaluation questions are clearly stated. Numerous standard textbooks present a variety of ways and examples of how evaluation questions can be formulated.[82]

Table D.2 in Annex D of this report provides a detailed overview of the *Evaluation methodological characteristics* section of the analytical framework developed for this study.

[78] See: Conley-Tyler (2005); Scriven (1997).

[79] EC (2015).

[80] HM Treasury (2011).

[81] OECD DAC Network on Development Evaluation (2010).

[82] See: OECD DAC Network on Development Evaluation (2010); HM Treasury (2011).

4.1.3. Limitations

As discussed in Section 2.2, the overall design of the analytical framework has been tailored to match the needs, requirements and resources of the current study. The categories have been selected to allow relevant information and metrics to be captured from literature sources. In order to conduct a more rigorous assessment of evaluation quality, for example by appraising the robustness of the data collection or analysis methods, it would be necessary to access the underlying primary data and to interview the evaluators and potentially stakeholders, which was beyond the scope of this study. In this case a number of the categories and associated values should be revised and expanded, possibly drawing on some established taxonomies of evaluation studies presented above. In particular, with a view to collecting primary data, the framework could be expanded to record a more detailed description of the evaluation being assessed and information on the quality of its implementation (e.g. capturing details about the selection and deployment of specific methods used in the evaluation).

It should also be noted that not all characteristics chosen for our framework are consistently used (or explicitly reported) by evaluators. It is not uncommon to see studies where evaluation criteria or questions are not spelled out. This is not unique to CT and PCVE evaluations, but it begs a wider question: in the absence of these parameters, considered important in the evaluation literature, can these studies be considered to be evaluations? It is likely that they would be defined as 'pseudo-evaluations' or other categories instead, if more stringent exclusion criteria were to be used. Given the nascence of the CT and PCVE evaluation fields, we have allowed some flexibility in including studies for the review, although the absence of some methodological characteristics has implications for the analysis of their quality.

4.2. Evaluation quality characteristics

4.2.1. Theoretical underpinnings

The concept of quality in evaluation is not set in stone: it is different for evaluators, stakeholders and decision-makers. Quality may also evolve over time: what once was considered good enough might not necessarily meet current standards.[83] The quality of evaluation is therefore a fluid concept. Moreover, the theory and practice of assuring quality in evaluation and methods of doing it are very diverse. Schwartz & Mayne (2005) distinguish a wide variety of structural, formative, summative and systemic approaches adopted to assure the quality of an evaluation, each of which can be implemented through different means and in pursuit of different purposes. Table 4.2 provides an overview of the different quality assurance approaches identified, the objectives they pursue, and options for their implementation.

[83] Stake & Schwandt (2006).

Table 4.2. Approaches to assuring quality of evaluations

Approaches used	Evaluations
Structural approaches: setting guidelines and standards	Professional evaluation society standards
	Professional practice guidance (e.g. textbooks, academic writing)
	Organizational/governmental guidance and standards
	Training and capacity development
Formative approaches: real-time assessments of individual reports	Advisory committees
	Internal quality control procedures
Summative approaches: ***ex post*** assessments of individual reports	Independent assessments, such as State Audit Institution (SAI) audits
	Semi-independent assessments
	Self-assessments
Systemic approaches: assessments of systems and procedures for producing evaluations	SAI audits

SOURCE: Based on Schwartz & Mayne (2005).

According to Schwartz & Mayne (2005), structural approaches draw on guidelines and principles (varying from a few such standards to nearly 30, often grouped in fewer broader categories). Most common are the three types of standards that relate to:

1. **The technical quality of the information produced**. This includes the following aspects: i) a well-defined scope, which means that the objectives of the information, the purposes to be served and the range of coverage should be clearly set out; ii) accurate data, implying that the data collected should be valid and reliable; iii) sound analysis, based on a robust methodology; and iv) substantiated (or evidence-based), impartial and objective findings and conclusions.
2. **The quality of the process used to obtain the information.** This type of standard primarily focuses on stakeholder involvement, assessing the extent to which the concerns and interest of those being evaluated and those affected by the evaluation have been explicitly addressed.
3. **The usefulness of the information produced**. This includes the following aspects: i) timeliness (the information is produced at a time when it can make a difference); ii) the 'right' scope (the information produced is relevant to the issues of the day); and iii) clarity (the information is understandable by the intended audience).

Summative approaches assess reports after their completion and, typically, publication. Schwartz & Mayne (2005) conclude that meta-evaluations (i.e. studies that evaluate evaluations or in other words studies that determine the quality and/or value of an evaluation) are still rather uncommon. Those meta-

evaluations that exist range from mandatory external reviews to SAC reviews, and sometimes reviews conducted by academics or by state audit institutions, and they offer learning for future evaluation work. The attempt to assess the quality of evaluations presented in this study falls neatly under the summative umbrella.

Davidson (2004) lists five criteria according to which meta-evaluations should be assessed:

> *validity, utility, conduct, credibility and costs. In other words, evaluations should produce valid and justifiable conclusions; be useful to the client and other relevant audiences; be conducted in an ethical, legal, professional, and otherwise appropriate manner; be credible to relevant audiences; and be as economical, quick, and unobtrusive as possible.*

The obvious question is how this can be done. Using checklists is quite common and highly recommended in the evaluation literature.[84] However, even the most helpful checklists require some degree of subjectivity from the person assessing an evaluation. Like any evaluation, meta-evaluation requires value judgement and avoiding it seems impossible. Additional difficulties lie in the fact that each evaluation is formed by its specific objective(s) and limitations under which it operates and that any methodological designs imply certain trade-offs. In order to limit the subjectivity of our assessments, we developed a customised section of the framework that captures quality characteristics in the most factual way possible.

4.2.2. The framework – Section 2

The second section of the analytical framework comprises a set of categories capturing information about evaluation quality. The criteria in this section have been designed with a view to adopting, to the extent possible, an objective approach to quality assessment, based on transparent metrics available in publications selected for the inventory.

In particular, this section of the analytical framework focuses on:

- **Quality assessment**: records whether and which quality assessment method, such as peer review or self-assessment, was used to ensure the quality of the evaluation.
- **Perspective**: records whether evidence and data used in the evaluation come from multiple sources and stakeholder groups.
- **Performance assessment**: records whether a framework codified in a grid, rubric, score or set of indicators was used to assess levels of performance during the initiative implementation.
- **Lessons learned**: records whether the evaluation identifies key conclusions, lessons learned or observations in regard to the evaluation design and undertaking.
- **Recommendations**: records whether the evaluation formulates recommendations for improvement of the initiative being implemented.
- **Unintended effects**: records whether the evaluation takes into account or reviews possible positive or negative unintended side effects of the initiative being evaluated.

[84] See: Scriven (2015); Scriven (2011); Stufflebeam (2001).

Table D.3 in Annex D provides a detailed overview of the *Evaluation quality characteristics* section of the project analytical framework.

4.2.3. Limitations

As discussed in Section 2.2, the overall design of the analytical framework has been tailored to match the needs, requirements and resources of the current study. With regard to this section, in the event that the framework was to be employed to perform an in-depth critical appraisal of a CT or PCVE evaluation, including through the collection of primary data from evaluation implementers and beneficiaries, a number of the categories and associated values could be revised and expanded in line with well-established frameworks in the evaluation literature. For example, this could be done with a view to allowing for the inclusion of criteria focusing on assessing the validity, quality and extent to which the evaluation approach and its implementation are systematic in nature.

Hierarchies of evidence have developed as a method through which the level of evidence underlying interventions can be assessed, with each 'level' implying greater confidence about the effectiveness of the intervention in relation to a particular outcome. These hierarchies range from expert opinions and case studies, to increasingly rigorous methods of data collection and verification, such as Randomised Controlled Trials (RCTs), systematic reviews and meta-analyses. Sometimes, they may also take into account the quality of each study design (e.g. assessing sample sizes), and the significance of the findings (e.g. examining the effect size).[85]

The Maryland Scientific Methods Scale (MSMS) is one example of such hierarchies of evidence and puts RCTs at the top of the 5-point scale.[86] However, experimental designs have their shortcomings and in some contexts are not achievable or desirable.[87] Without entering this debate, we note that with data collected in Section 1 and Section 2 of the study analytical framework, the study could cluster and classify evaluations reviewed according to the MSMS.

4.3. Initiative characteristics

4.3.1. Theoretical underpinnings

As discussed in Chapter 3, recent years have seen an increase in the volume and typologies of programmes and interventions designed to prevent and counter terrorism and violent extremism more broadly.[88] As emphasised by Koehler (2017), no one-size-fits-all approach to CT and PCVE can or should exist. As a result, a wide array of approaches and programmes have been designed and implemented to account for and respond to context-specific needs and requirements in the areas of CT and PCVE. From the perspective of the present study, the variety and breadth of the CT and PCVE spectrum constitutes a challenge that the study's analytical framework must be able to tackle, accounting for the different

[85] See: Kleemans et al. (2007); Alkin & Christie (2004).

[86] Sherman et al. (1997); Farrington et al. (2002).

[87] Campbell & Stanley (1966).

[88] Romaniuk & Chowdhury Fink (2012); Marsden (2017).

structural characteristics, intended and unintended effects, target groups, methods, activities and actors characterising CT and PCVE initiatives.[89]

At a basic level, the study framework must be able to account for differences in PCVE and CT work and in the goals and purposes pursued by such programmes. To this end, building on the definition provided in Chapter 3 of this document, the study and its analytical framework take PCVE to encapsulate initiatives that address drivers of violence and extremism; that build resilience and immunity to extremist ideologies; and that deter and disrupt recruitment and mobilisation by extremist groups, including by supporting the reintegration of (former) extremists and offenders.[90] The study will take CT to cover all those initiatives designed to deter, disrupt and isolate groups that use terror, including by training and equipping the security apparatus; by increasing the capacity to prepare, prevent, protect and respond to terrorist incidents; and by interdicting and prosecuting through law enforcement activities.[91]

In addition to characterising initiatives evaluated according to their goal and purpose, a framework for analysing CT and PCVE work should be able to account for other key features determining the nature of such initiatives. Programmes with similar purposes may target different groups or operate at different levels of interventions within society (e.g. macro, meso, micro) and from a geographic perspective (e.g. transnational, national, regional, local). In turn, these characteristics are likely to reflect and influence the types of programmes and activities envisioned under a particular initiative and on its planned temporal scope.[92]

Keeping the above considerations and requirements in mind, the following section presents a customised framework, focusing on initiatives' characteristics, to guide the characterisation of CT and PCVE programmes evaluated in the sources included in the study's inventory.

4.3.2. The framework – Section 3

The third section of the analytical framework comprises a set of categories capturing of information about the strategy, policy, programme or intervention being evaluated. In particular, it focuses on the following characteristics of a CT or PCVE initiative:

- **Type:** the type of initiative being evaluated, distinguishing between strategies, policies and programmes and interventions.
- **Goal**: the main CT- or PCVE-specific goals of the initiatives being evaluated, for example protecting people or property from extremist and terrorist threats (protect) or altering and removing the commitment of targets to an extremist ideology or belief (deradicalise).
- **Ideology**: the type of extremist ideology being addressed by the initiative being evaluated, for example religious or political extremism.

[89] Bjorgo & Horgan (2009); Koehler (2017); Ris & Ernstofer (2017).

[90] Koehler (2017); Ris & Ernstofer (2017).

[91] Lum et al. (2006); Ris & Ernstofer (2017).

[92] Koehler (2017); Marsden (2017).

- **Level of intervention**: the level at which the initiative being evaluated aim to operates, for example at the individual level (micro), at the level of groups or organisations (meso), or at the level of entire social systems, such as nations or cities (macro).
- **Type of activities**: the main types of activities conducted or planned under the strategy, policy or programme being evaluated, such as education, capacity building, policing, etc.
- **Group of focus**: the primary groups of focus of the initiative being evaluated, for example radicalised subjects, offenders, prisoners, etc.
- **Implementers**: the type of organisation implementing the strategy, policy or programme being evaluated, such as an international organisation, a governmental actor, a non-governmental actor or a public-private partnership.
- **Status** (i.e. ongoing or completed) of the initiative being evaluated.
- **Geographic scope** of the initiative being evaluated, for example local, regional, national or transnational.
- **Foreseen duration** of the initiative being evaluated.
- **Length of implementation** of the initiative being evaluated at the time of the evaluation.

Table D.4 in Annex D provides a detailed overview of the *Initiative characteristics* section of the project analytical framework.

4.3.3. Limitations

As discussed in Section 2.2, the overall design of the analytical framework has been tailored to match the needs, requirements and resources of the current study. With regard to this section, in the event that the framework was to be employed to perform an in-depth critical appraisal of a CT or PCVE initiative and of its evaluation, including through the collection of primary data from the initiative's implementers and beneficiaries, a number of the categories and associated values could be revised and expanded in line with frameworks established in the literature. For example, a number of the categories and associated values could be revised and expanded, with a view to collecting primary data and recording a more detailed description of the initiative being assessed and of information on the quality and results of its implementation.

4.4. Lessons learned

A pivotal element of the study entailed the capturing and classifying of lessons learned from evaluation reports and manuscripts. It should be noted that the study did not focus on capturing lessons learned about the design, operation and results achieved by individual CT and PCVE strategies, policies and programmes being evaluated. Instead, the study team collected lessons learned about the design and conduct of the evaluations of such programmes.

At the level of individual evaluations, the study recorded 'lessons learned' identified as such by the authors of the publication reviewed. Lessons were then categorised using a high-level pre-determined list of categories as well as through an open-text data point allowing a more granular level of detail to be captured.

Table 4.3 lists the pre-determined high-level categories that were used and gives examples of free-text, granular inputs from previously extracted publications.

Table 4.3. Lessons learned categories questions

High-level lessons categories
• Evaluation goals • Evaluation design • Evaluation timeframe • Evaluation method • Evaluation management • Evaluation reference data • Evaluation outcomes • Facilitators to evaluating the intervention • Inhibitors to evaluating the intervention
Examples of categories extracted from publication text
• Control group • Attribution problem • Sample size • Randomisation method • Resource constraints • Security concerns

Lastly, to determine whether the lessons learned presented in the literature were of general validity, each lesson identified was reviewed against a set of specific criteria, as presented in Table 4.4.[93]

Table 4.4. Lessons learned validation questions

Lessons learned review questions	
Is the study clear about what was meant by a 'lesson' and 'learned'?	Yes / No
Is the study clear about by 'whom' was the lesson learned?	Yes / No
Is the study clear about the 'evidence' supporting the lesson?	Yes / No
Is the study clear about under what 'conditions' the lesson applies?	Yes / No
Is the lesson 'specific' enough to guide practice?	Yes / No

[93] This is adapted from Patton (2001).

If the answer to one of the questions presented in Table 4.4 was positive, a score of 1 was assigned to the lesson. If the answer was negative, a score of 0 was assigned. The overall score of a lesson determines its overall level of validity, as indicated in Table 4.5.

Table 4.5. Lessons learned validity scoring

Lessons learned review questions	
Low validity	0 – 1
Medium validity	2 – 3
High validity	4 – 5

Lessons learned were then also analysed at aggregate level, providing a high-level overview of recurring themes, barriers and opportunities to the design and conduct of quality evaluations in the fields of CT and PCVE.

5. Analysing CT and PCVE evaluations

This chapter presents the inventory of CT and PCVE evaluations produced by this study and discusses findings stemming from an aggregate-level analysis of its sources. The chapter opens by providing an overview of the inventory and of the sources included in it. The chapter then discusses the characteristics of the initiatives evaluated, before focusing on the characteristics of their respective evaluations.

5.1. The study inventory of CT and PCVE evaluations

The inventory of CT and PCVE evaluations compiled by this study was developed via a three-pronged, systematic approach, followed by an iterative and multi-strand review process. The approach, as described in Chapter 2 and Annex B, was designed and implemented with a view to reducing ambiguity about the selection of sources and reducing the risk of overlooking or excluding any relevant sources. Nonetheless, it is recognised that the process of compiling the inventory was to an extent arbitrary. The issues and phenomena investigated (i.e. PCVE, CT, evaluation) are complex and do not lend themselves well to clear-cut definitions and inclusion criteria. Should this study be repeated, different research teams may come to different conclusions as regards the inclusion or exclusion of specific sources.

Figure 5.1 provides a schematic overview of the process followed for compiling the inventory and shows the number of sources included after each of the different research strands and review phases. For an overview of the review and analysis process employed, please refer to Sections 2.3 and B.3. Overall, the final study inventory includes 48 publications, of which 38 are in English, 6 in Dutch and 4 in German. The majority of sources included (37) were published in the originally envisioned timeframe (i.e. after 1 January 2013). The remaining sources (11), dating from before January 2013, were flagged as relevant by external experts and stakeholders contacted by the study team. We decided to include these publications in the inventory and analyse them because they had been highlighted by experts in the fields of CT, PCVE and evaluation. It should be noted, however, that the pre-2013 sample of publications is not meant to be exhaustive nor systematic, as publications from before January 2013 were not searched under other strands of the research protocol employed. The findings presented below include data from pre-2013 manuscripts unless explicitly mentioned.

Figure 5.1. Overview of the inventory production process

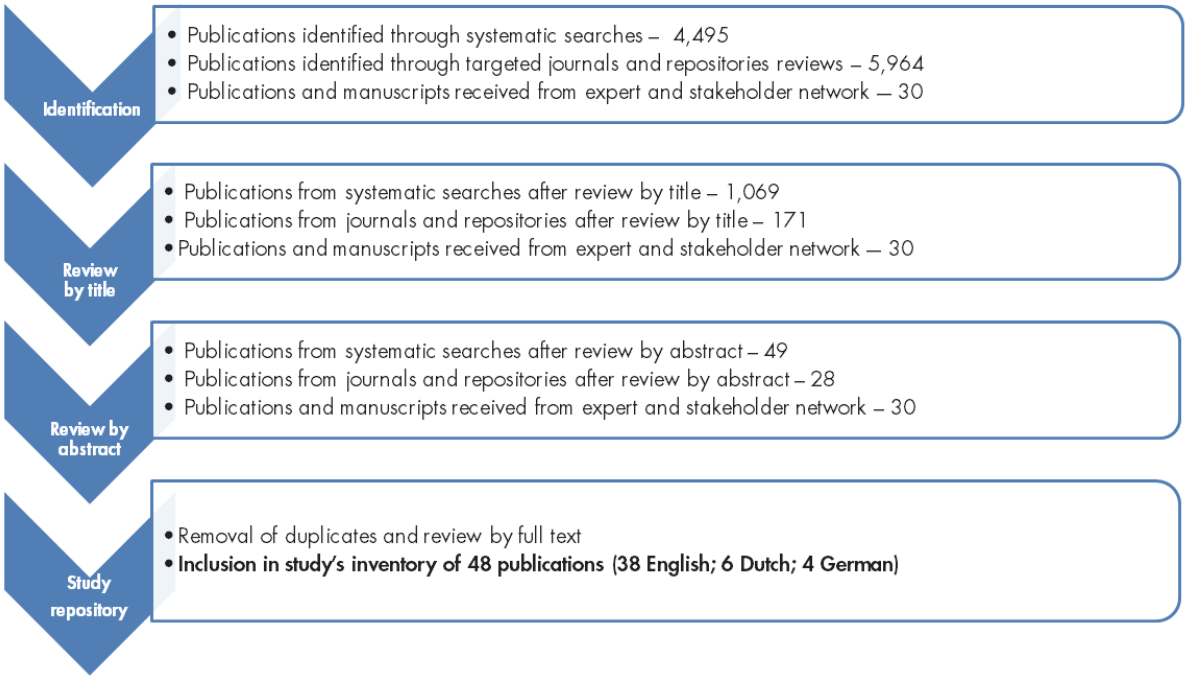


Table 5.1 provides an overview of the analysis and coding prepared by the study team for each source included in the study repository. It should be noted that this table includes only a synthesis of the extraction and coding produced for each manuscript reviewed. The following sections of this chapter discuss the study's inventory at an aggregate level, focusing first on the characteristics of the initiatives being evaluated, and then on those of their evaluations. Where possible, and mindful of the limited sample size, the discussion highlights any noteworthy association between various initiatives and evaluation characteristics.

Table 5.1. Study inventory analysis overview

ID	Author(s)	Year	Evaluation purpose	Evaluation design	Methods	Evaluator	Intervention type	Intervention group	Target ideology data	Intervention geographic scope	Intervention foreseen duration
1	Aldrich	2012	Outcome	Quasi-experimental design	Quantitative	Internal	Programme or intervention	PCVE	Any	Transnational	3+ years
2	Aldrich	2014	Outcome	Quasi-experimental design	Quantitative	External	Programme or intervention	PCVE	Religious extremism	Local	3+ years
3	Aly et al.	2014	Outcome	Before-and-after	Quantitative	Internal	Programme or intervention	PCVE	Religious extremism	National	3+ years
4	Argomaniz & Vidal-Diez	2014	Outcome	Quasi-experimental design	Quantitative	External	Policy	CT	Political extremism	National	3+ years
5	Boyd-MacMillan	2016	Outcome	Additionality	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	National	3+ years
6	Broadbent	2013	Process; Outcome	Additionality	Mixed methods	Unclear	Programme or intervention	PCVE	Religious extremism	National	<6 months
7	Carson	2017	Outcome	Quasi-experimental design	Quantitative	External	Policy	CT	Religious extremism	Transnational	3+ years
8	Choudhury & Fenwick	2011	Outcome	Additionality	Qualitative	External	Policy	Mixed	Any	National	3+ years
9	Demant & De Graaf	2010	Outcome	Additionality	Qualitative	External	Strategy	Mixed	Political extremism	National	3+ years
9	Demant & De Graaf	2010	Outcome	Additionality	Qualitative	External	Strategy	Mixed	Religious extremism	National	3+ years
10	Dunn et al.	2015	Outcome	Additionality	Quantitative	External	Programme or intervention	PCVE	Religious extremism	Regional	3+ years
11	Feddes et al.	2015	Outcome	Before-and-after	Quantitative	External	Programme or intervention	PCVE	Religious extremism	Local	<6 months
12	Finn et al.	2016	Process; Outcome	Additionality	Qualitative	External	Policy	PCVE	Religious extremism	National	3+ years
13	Van Gestel & De Poot	2014	Process; Outcome	Additionality	Qualitative	External	Policy	CT	Any	National	3+ years
14	Haider & Martinez	2014	Process; Outcome	Time series	Quantitative	External	Programme or intervention	PCVE	Any	National	3+ years

ID	Author(s)	Year	Evaluation purpose	Evaluation design	Methods	Evaluator	Intervention type	Intervention group	Target ideology data	Intervention geographic scope	Intervention foreseen duration
15	Hirschfield et al.	2012	Process	Additionality	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	National	3+ years
16	Johns et al.	2014	Outcome	Additionality	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	Regional	6–12 months
17	Jordan	2014	Outcome	Additionality	Quantitative	External	Programme or intervention	CT	Religious extremism	National	3+ years
18	Khalil & Zeuthen	2014a	Process; Outcome	Additionality	Qualitative	External	Programme or intervention	PCVE	Religious extremism	National	1-3 years
19	Khalil & Zeuthen	2014b	Process; Outcome	Additionality	Qualitative	External	Programme or intervention	PCVE	Religious extremism	National	1-3 years
20	Krafchik & Ryszkowska	2011	Process	Additionality	Mixed methods	External	Programme or intervention	PCVE	Any	Local	3+ years
21	Kundnani	2009	Process; Outcome	Additionality	Qualitative	External	Strategy	Mixed	Religious extremism	National	3+ years
22	Kurz et al.	2016	Outcome	Quasi-experimental design	Mixed methods	Internal	Programme or intervention	PCVE	Political extremism	Regional	3+ years
23	Lakhani	2011	Process	Additionality	Qualitative	External	Strategy	Mixed	Religious extremism	National	3+ years
24	Lamb	2012	Process	Additionality	Qualitative	External	Strategy	Mixed	Any	National	3+ years
25	Liht & Savage	2013	Outcome	Before-and-after	Mixed methods	External	Programme or Intervention	PCVE	Religious extremism	National	<6 months
26	McNeal	2014	Process	Additionality	Qualitative	External	Policy	CT	Religious extremism	Transnational	3+ years
27	Minor	2016	Process; Outcome	Additionality	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	Local	1–3 years
28	Mitts	2017	Impact	Quasi-experimental design	Quantitative	External	Policy	PCVE	Religious extremism	National	3+ years
29	Moffet & Sgro	2016	Process	Additionality	Mixed methods	Internal	Programme or intervention	PCVE	Any	Transnational	3+ years
30	Möller & Neuscheler	2016	Process; Outcome	Additionality	Qualitative	External	Programme or intervention	PCVE	Political extremism	Local	NA

ID	Author(s)	Year	Evaluation purpose	Evaluation design	Methods	Evaluator	Intervention type	Intervention group	Target ideology data	Intervention geographic scope	Intervention foreseen duration
31	Möller & Neuscheler	2018	Process; Outcome	Additionality	Qualitative	External	Programme or intervention	PCVE	Religious extremism	Local	
32	Noordegraaf et al.	2016	Process	Additionality	Qualitative	External	Strategy	Mixed	Any	National	3+ years
33	Noordegraaf et al.	2017	Process	Unclear	Unclear	External	Strategy	Mixed	Any	National	3+ years
34	O'Toole et al.	2012	Process; Outcome	Additionality	Qualitative	External	Strategy	Mixed	Religious extremism	National	3+ years
35	Octavia & Wahyumi	2015	Outcome	Before-and-after	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	Regional	1–3 years
36	Palloks & Steil	2015	Process; Outcome	Additionality	Qualitative	External	Programme or intervention	PCVE	Political extremism	Local	NA
37	Roodnat & Eijkman	2016	Outcome	Additionality	Qualitative	External	Policy	PCVE	Any	National	3+ years
38	Schuurman & Bakker	2016	Process; Impact	Additionality	Qualitative	External	Programme or intervention	Mixed	Religious extremism	National	3+ years
39	Sheikh et al.	2012	Process; Outcome	Additionality	Mixed methods	External	Programme or intervention	PCVE	Religious extremism	Regional	1–3 years
40	Veldhuis et al.	2011	Process	Unclear	Mixed methods	External	Programme or intervention	Mixed	Any	National	3+ years
41	Veldhuis	2015	Process	Unclear	Mixed methods	External	Programme or intervention	Mixed	Any	National	3+ years
42	Vermeulen	2014	Outcome	Additionality	Qualitative	External	Programme or intervention	Mixed	Religious extremism	Local	3+ years
42	Vermeulen	2014	Outcome	Additionality	Qualitative	External	Programme or intervention	Mixed	Any	Local	3+ years
42	Vermeulen	2014	Outcome	Additionality	Qualitative	External	Programme or intervention	Mixed	Religious extremism	Local	3+ years
43	Webber et al.	2018	Outcome	Quasi-experimental design	Quantitative	External	Programme or intervention	PCVE	Political extremism	National	3+ years
44	Webster et al.	2017	Process	Additionality	Qualitative	External	Programme or intervention	Mixed	Any	National	1–3 years

ID	Author(s)	Year	Evaluation purpose	Evaluation design	Methods	Evaluator	Intervention type	Intervention group	Target ideology data	Intervention geographic scope	Intervention foreseen duration
45	Williams et al.	2015	Process	Additionality	Mixed methods	External	N/A	N/A	N/A	N/A	N/A
46	Williams et al.	2016	Outcome	Quasi-experimental design	Quantitative	External	Programme or intervention	PCVE	Religious extremism	National	1–3 years
47	Wilner & Rigato	2017	Process	Additionality	Quantitative	Internal	Programme or intervention	PCVE	Any	Transnational	<6 months
48	Winter et al.	2014	Process; Outcome	Additionality	Qualitative	External	Strategy	Mixed	Any	National	3+ years

5.2. Characteristics of CT and PCVE initiatives in the inventory

A review of the publications included in the study's inventory resulted in the coding of information about 50 initiatives spanning the entire CT and PCVE spectrum. The discrepancy in the number of initiatives coded and manuscripts reviewed is due to two publications[94] that present an evaluation of multiple initiatives with a level of detail sufficient to codify different initiatives separately and according to the study's analytical framework. In other instances, although the evaluation manuscript reviewed focuses on multiple interventions, it was not possible to code each CT or PCVE initiative at a sufficient level of detail to generate separate data entries.[95] In these cases, multiple initiatives evaluated were coded at an aggregate level on the basis of their collective traits and characteristics.

The following paragraphs discuss findings pertaining to the characteristics of initiatives being evaluated in the study's inventory publications, focusing on the characteristics captured through the study's analytical framework (i.e. initiative type, goal, type of activities, target ideology, implementers, level of intervention and group of focus).

Initiative type

The majority of evaluations in the study inventory (33 out of 50) focus on initiatives undertaken at the level of programmes or interventions, whereas fewer publications focus on higher-level policies (8) and strategies (9). Table 5.2 provides an overview of initiatives evaluated in the study inventory, focusing on their type and goals.

Table 5.2. Overview of CT and PCVE initiatives evaluated by type and goals (n=50)

Initiative type	CT-specific	CT- and PCVE-relevant	PCVE-specific	TOTAL
Strategy	-	9	-	9
Policy	4	1	3	8
Programme	1	7	25	33
TOTAL	**5**	**17**	**28**	**50**

As for initiatives evaluated at programme and intervention level, it is worth noting that the majority of these had a PCVE-specific nature (25) or pursued PCVE-specific goals as part of their activities (7), whereas only one CT-specific programme-level evaluation was identified. The preponderance of PCVE-specific or PCVE-related (i.e. pursuing at least one PCVE-specific goal or entailing at least one PCVE-specific activity) initiatives is not limited to evaluations at programme and intervention level. Taking the study inventory as a whole, 28 initiatives evaluated were PCVE-specific and 17 PCVE-relevant. In contrast, only five initiatives were CT-specific, and an additional 17 had a clear CT-relevance.

[94] Demant & De Graaf (2010); Vermeulen (2014).

[95] See, for example: Argomaniz & Vidal-Diez (2015).

Initiative goal

The analytical framework distinguished seven categories of objectives for CT initiatives: procure, prevent, protect, prepare, repress/prosecute, disengage and deradicalise. Preventative goals characterised the majority of interventions included in the study inventory: 29 initiatives pursued a purely preventative goal, which corresponds to three-quarters of PCVE-specific interventions. More broadly, a majority of initiatives, 42 in total, pursued preventative goals alongside other CT- and PCVE-specific aims. It is also worth noting that only five PCVE-specific and six PCVE-relevant initiatives pursued goals pertaining to disengagement and deradicalisation at the individual or group levels.

A multiplicity of factors may explain the domination of PCVE initiatives within the study's evaluations inventory. Firstly, programmes and practices that can be characterised as aligning with traditional CT work are often shrouded in secrecy and remain largely inaccessible to scholarly inquiry.[96] Secondly, both qualitative and quantitative traditional evaluation methods are faced with significant obstacles and constraints when investigating CT.[97] For example, the relative low frequency with which relevant incidents in this field occur,[98] compared to the field of criminology for instance, complicates detection, measurement and attribution of results observed (e.g. lack of terrorist attacks) to a particular initiative. However, this does not provide a complete explanation of the differences between the fields of CT/PCVE and criminology evaluation.[99]

Type of activities

Considering the type of activities implemented, 39 initiatives entailed either educational or communications or capacity-building (including mentoring and counselling) activities. Initiatives that entailed exclusively one or a combination of activities from the areas of education, communications and capacity building account for about half (26) of all initiatives and hail almost exclusively (25) from among PCVE-specific programmes. As for CT-specific initiatives (5), these comprise three military initiatives, a policing programme, and a multi-year series of measures combining legal, policing, intelligence and capacity-building work. As for initiatives that involved a clear policing role or presence, these account for nearly a third of initiatives (16) in the inventory. It is worth noting, however, that in several of such initiatives, the role and purpose of law enforcement appears to be in support or in the undertaking of communications and community engagement activities.

Initiative target ideology

A small majority of publications (28 out of 50) in the inventory involved evaluations of initiatives that focus exclusively on religious extremism and almost entirely on jihadist and radical Islamist-inspired ideologies. The majority of these (18) comprised PCVE-specific preventative policies and programmes with a focus on education, capacity-building and communications activities. About a third of the initiatives (17) targeted different types of extremism. Among these, seven were policy and strategic-level

[96] Lum et al. (2006); McNeal (2014).

[97] Noordegraaf et al. (2017).

[98] Reding et al. (2014).

[99] Lum et al. (2006).

initiatives designed to provide an overarching canvass and framework for lower-level CT and PCVE activities. Only six initiatives exclusively targeted political extremism, and two of these were implemented prior to the turn of the century.

Actors responsible for implementation

With regard to actors responsible for implementation, a small majority (27 out of 50) of initiatives was implemented by governmental actors, whereas non-governmental actors were responsible for 8 initiatives and public-private partnerships for 15 initiatives. Governmental actors were found to be exclusively responsible for implementing CT-specific initiatives, whereas non-governmental actors (8) and public-private partnerships (10) were predominantly responsible for the implementation of PCVE-specific efforts.

Level of intervention

The initiatives captured in the study's inventory appear to be fairly equally distributed across different levels of intervention: at the individual level (micro), at the level of groups or organisations (meso), and at the level of entire social systems, such as countries or cities (macro). Some 15 initiatives took place at the macro-level, 19 at the meso level, and 16 at the micro level. The initiatives implemented at the meso level primarily targeted communities perceived to be at risk of radicalisation and extremist rhetoric. Of these, half focused primarily on individuals at-risk or vulnerable to radicalisation (8), while most of the others focused on offenders and prisoners (7). Macro-level initiatives targeted a mix of vulnerable communities and groups judged to be at risk (8 out of 15).

Initiative group of focus

Finally, as regards initiatives' groups of focus, it is worth noting in the present study's inventory the limited presence of initiatives targeting offenders (8 out of 50), radicalised subjects (2), prisoners (4), and first-line workers (1). Initiatives targeting offenders were found to be primarily pursuing CT-specific goals and to entail the implementation of repressive and punishing measures. As for initiatives focusing on radicalised subjects, prisoners and first-line workers, these revolved around risk-management frameworks and approaches, as well as capacity-building interventions.

Table 5.3 provides an overview of the initiatives evaluated in the study inventory, focusing on their target ideology, goals and level of intervention. Table 5.4 focuses on the initiatives' implementing actors, goals and levels of intervention.

Table 5.3. Overview of CT and PCVE initiatives evaluated by target ideology, goals and level of intervention (n=50)

		CT-specific			CT- and PCVE-relevant			PCV-specific			
Level of intervention		Macro	Meso	Micro	Macro	Meso	Micro	Macro	Meso	Micro	**TOTAL**
Target ideology	Any	1	-	-	4	4	1	3	2	1	**16**
	Religious	-	-	3	3	3	1	2	10	6	**28**
	Political	1	-	-	1	-	-	-	-	4	**6**
	TOTAL	**2**	**-**	**3**	**8**	**7**	**2**	**5**	**12**	**11**	**50**

Table 5.4. Overview of CT and PCVE initiatives evaluated by implementing actor, goals and level of intervention (n=50)

		CT-specific			CT- and PCVE-relevant			PCV-specific			
Level of intervention		Macro	Meso	Micro	Macro	Meso	Micro	Macro	Meso	Micro	**TOTAL**
Implementer	Governmental	2	-	3	7	3	2	-	5	5	**27**
	Non-governmental	-	-	-	-	-	-	1	4	3	**8**
	Public-private partnership	-	-	-	1	4	-	4	3	3	**15**
	TOTAL	**2**	**-**	**3**	**8**	**7**	**2**	**5**	**12**	**11**	**50**

5.3. The study inventory's evaluations

A review of the publications included in the study's inventory resulted in the coding of data pertaining to 48 evaluations of 50 CT and PCVE initiatives. This section discusses findings relating to the methodological and quality characteristics of evaluations included in the study's inventory through the prism of the study's analytical framework. Where possible, and mindful of the limited sample size, the discussion highlights any noteworthy association between various evaluation characteristics and between initiative and evaluation characteristics.

5.3.1. Evaluation methodological characteristics

Evaluation purpose

Looking at the purpose pursued by evaluations analysed, the study's inventory comprises primarily process evaluations and impact- or outcome-focused evaluations, the characteristics of which are discussed below. Most of the studies in the inventory involved impact or outcome evaluations (35 out of 48). In the majority of these instances (25) no clear evaluation approach could be discerned from information

included in the publications. Impact or outcome evaluations appear to be used more frequently for evaluating initiatives at higher levels (i.e. strategies and policies) across both CT and PCVE. Among process evaluations in the inventory, it does not seem to be the case that any particular initiative's type, goal or activity is disproportionally represented.

Aside from impact or outcome evaluations, a majority of evaluations (28 out of 48) included in the study's inventory focused on process, either in isolation (13) or as part of a broader process and impact or outcome evaluation (15).

Evaluation approach and methods

In the majority of studies included in the inventory (33 out of 48), no clear evaluation approach could be discerned from the information provided. Of the 15 other studies, 10 were characterised by authors as theory-driven evaluations, three were based on a 'realist evaluation' approach, one was described as a participatory evaluation and one was described as taking a transboundary evaluative approach.

Among the impact or outcome evaluations identified above, a majority used quantitative methods alone (12 out of 35) or quantitative methods as part of a mixed approach (8). Some 15 evaluations used exclusively qualitative methods. Methods employed for impact or outcome evaluations typically include combinations of desk research, interviews, focus groups and case studies for the purpose of qualitative data collection, whereas in-person and remote surveys tended to be employed by those using quantitative methods. As for analytical methods, thematic analysis was used for qualitative methodologies and descriptive statistics as well as advanced statistics (e.g. bivariate analysis, propensity score matching or hazard models) for evaluations using quantitative methods.

Similarly to impact evaluations, for the majority of process evaluations included in the study's inventory (17 out of 28) no clear approach could be identified on the basis of the information available. A number of other evaluative efforts focusing on process were found to be theory-driven (6), some used a 'realist evaluation' approach (3), and one could be characterised as 'participatory evaluation'. With regard to evaluation design, the process evaluations reviewed focused on additionality. Unsurprisingly, most process evaluations relied solely on qualitative methods (16 out of 28), a higher share than for impact or outcome evaluations. Only two evaluations employed quantitative methods, and nine relied on a mixed approach. Methods used for process evaluations broadly align with those employed for impact or outcome evaluations. In particular, almost all of the process evaluations included in the study inventory used combinations of desk research, interviews and focus groups for the purpose of qualitative data collection, whereas in-person and remote surveys were employed by those using quantitative methods. As for data analysis methods, almost a third of the process evaluation manuscripts (9) do not indicate what data analysis method was employed. This is particularly common among evaluations using qualitative methods only. As for the remaining process evaluations, thematic analysis and descriptive statistics are the most frequently cited analysis methods for evaluations using qualitative and quantitative methods respectively.

Evaluation design

As regards evaluation design, the majority of publications reviewed focused on additionality (22 out of 48). Additionality denotes evaluations that look at what happened after an initiative has been implemented, taking a single measurement (equivalent to level 1 in the MSMS). These evaluations may

acknowledge that the initiative evaluated may not be responsible for the entirety of the changes observed, but nonetheless fail to provide sound evidence as to whether change has occurred (due to the lack of a pre-initiative baseline measurement) and fail to provide counterfactual evidence (as no control or comparison group is established). This design is used consistently for evaluating initiatives across the strategic (7), policy (5) and programme (19) levels for both CT- and PCVE-specific and -relevant work.

In eight evaluations a quasi-experimental design was adopted. These designs included four approaches using difference-in-differences (level 3 in the MSMS), two using propensity score matching (level 3 in the MSMS) and two using interrupted time series (level 2 in the MSMS). Of the quasi-experimental designs, most were used to evaluate CT or PCVE policy- or programme-level work conducted as part of larger government rehabilitation or international development initiatives conducted in developing countries. All of the quasi-experimental designs included in the study's inventory used quantitative data collection and analysis methods.

Finally, five evaluations used longitudinal designs (before-and-after or time series, but without a comparison group – level 2 in the MSMS). In all instances, these designs were used to evaluate PCVE-specific interventions focusing on capacity-building activities designed to increase individual and group-level resilience to violent extremist rhetoric or activities.

Evaluator

The vast majority of evaluations reviewed, a total of 42, were conducted by external evaluators. These are distributed equally across evaluations pursuing different purposes and adopting different designs. The study team did not observe any obvious differences between the characteristics of those evaluations that were conducted by external evaluators and those that were not.

Evaluation timing

The study's inventory comprises interim evaluations (30 out of 48), *ex post* evaluations (6) and embedded (ongoing) evaluations (11). Different evaluation timings appear to be evenly distributed across different initiative types (i.e. strategies, policies, programmes) and areas of work (i.e. PCVE- and CT-specific initiatives). Furthermore, no significant correlation between evaluation timing and design could be inferred given the limited sample size and that additionality and quasi-experimental evaluation designs have been employed consistently at different timings of evaluation. Finally, no *ex ante* evaluations were recorded in the study's inventory. While one might expect process evaluations to be associated with early phases of the implementation of an initiative, and impact/outcome evaluations with the later stages of delivery, the broad categorisation of the evaluation timing employed (before, during and after) lacked the granularity to capture these possible differences.

Table 5.5 provides an overview of the evaluations included in the study inventory, focusing on their purpose and approach; Table 5.6 focuses on their design and methods.

Table 5.5. Overview of the purpose and approach of CT and PCVE initiatives' evaluations (n=48)

		Evaluation approach					
		Missing	Participatory evaluation	Realist evaluation	Theory-driven evaluation	Transboundary evaluation approach	**TOTAL**
Evaluation type	Impact / outcome	16	-	-	4	-	**20**
	Process	8	-	-	4	1	**13**
	Process and impact / outcome	9	1	3	2	-	**15**
	TOTAL	**33**	**1**	**3**	**10**	**1**	**48**

Table 5.6. Overview of the purpose, design and methods of CT and PCVE initiatives' evaluations (n=48)

		Evaluation methods				
Evaluation purpose	**Evaluation design**	Mixed methods	Qualitative	Quantitative	Unclear	**TOTAL**
Impact / outcome	Additionality	2	4	2	-	**8**
	Before-and-after	2	-	2	-	**4**
	Quasi-experimental	1	-	7	-	**8**
Process	Additionality	4	5	1	-	**10**
	No data	2		-	1	**3**
Process and impact / outcome	Additionality	3	11	-	-	**e**
	Time series	-	-	1	-	**1**
	TOTAL	**14**	**20**	**13**	**1**	**48**

5.3.2. Evaluation quality characteristics

As discussed in Chapter 4, the review of evaluations focused to the extent possible on capturing several proxy metrics that offer an aggregate-level indication of evaluation quality. Data concerning lessons learned identified through a review of inventory sources are presented in Chapter 6.

Given the evolving nature of the concept of quality in evaluation and the limitations of document review as the chosen method of assessment, this report refrains from identifying commendable practices in the reviewed CT and PCVE evaluations and from developing a prescriptive recipe for good CT and PCVE evaluations. A judgement of how well the articles included in this review meet the criteria established in the literature could be more appropriately made as a result of a systematic meta-evaluation of each source. This, according to the study's definition of evaluation, would require the collection and analysis of primary data to capture different perspectives on a particular evaluation against a set of criteria (e.g. validity, credibility, utility, costs-effectiveness, etc.). Since this was not possible, this study aimed to capture quality assessments already performed, if any and if reported. In addition, what could be established is whether some minimal expectations for a promising evaluation have been met. These minimal expectations include (but are not limited to) capturing multiple perspectives (to limit bias), using grids, rubrics, scores or a set of indicators (to enhance transparency and objectivity), and providing recommendations (to increase the probability that evaluation results are used).

Evaluation quality assessment

As discussed in Section 5.3.1, most studies included in the study's inventory (42 out of 48) were conducted by external evaluators. Only in five instances were PCVE-specific programme-level interventions evaluated by the practitioners themselves. Furthermore, the majority of publications (27 out of 48) were subject to blind peer-reviewing by at least two reviewers, and in five evaluations an independent or academic review panel assessed the quality and outputs of evaluations being conducted. For the remainder of the manuscripts (16 out of 48) it was not possible to determine which, if any, mechanism had been used for reviewing evaluations.

Evaluation perspectives

The study team endeavoured to identify for each evaluation reviewed whether this is based on a multiplicity of perspective and data sources. The majority of evaluations (30 out of 48) used multiple data sources or methods and different perspectives were taken into account. But 18 evaluations relied on individual data sources or methods. In part, challenges in ensuring that evaluations relied on multiple data sources and perspectives may have been exacerbated by the fact that the vast majority of evaluations reviewed (37) did not report the use of grids, rubrics or scores throughout the implementation of initiatives to keep track of work being performed and its quality.

Evaluation recommendations

The study team looked at formulated evaluation recommendations and the degree of detail associated with these (e.g. identification of a clear stakeholder tasked with implementation, identification of a clear timeframe for implementation, etc.) as a proxy indicator of evaluation quality. A majority of evaluations

(29) did present recommendations for policy-makers and stakeholders, but these were never formulated in a way that clearly identified the intended task-owners and timeframes for implementation.

Unintended effects

Half of the evaluations reviewed (24) investigated and reported on the potential unintended effects of initiatives. It is worth noting that evaluations of CT-specific (5 out of 5) and CT-relevant (10 out of 14) initiatives seem to place greater emphasis on investigating unintended effects than evaluations of PCVE-specific initiatives. In contrast, less than half of the evaluations of PCVE-specific initiatives (9 out of 19) considered the potential unintended consequences of an intervention. However, the small numbers of evaluations in each group calls for caution in generalising these findings.

Table 5.7 provides an overview of the evaluations included in the study inventory, focusing on quality assessment mechanisms employed, the use of multiple perspectives, the presence of recommendations and the consideration of unintended effects.

Table 5.7. Overview of data for proxies used to characterise the quality of CT and PCVE initiatives' evaluations (n=48)

		Evaluation purpose			
		Impact / outcome	Process	Process and impact / outcome	**TOTAL**
Quality review mechanism employed?	Blind peer review	15	6	6	**27**
	Independent scientific steering group / panel	-	3	2	**5**
	Unclear	5	4	7	**16**
Multiple perspectives considered?	No	10	4	4	**18**
	Yes	10	9	11	**30**
Lessons learned discussed?	No	6	5	7	**18**
	Yes	14	8	8	**30**
Recommendations formulated?	No	8	8	3	**19**
	Yes	12	5	12	**29**
Unintended effects considered?	No	10	6	8	**24**
	Yes	10	7	7	**24**

5.4. Discussion of findings in the light of previous reviews and recommendations

This chapter has presented findings stemming from an aggregate-level analysis of publications included in the study's CT and PCVE evaluations inventory through the lenses of the study's analytical framework. While the inclusion criteria, analytical framework and key definitions used were tailored to the specific requirements of this research endeavour, a comparison of findings discussed so far with the results of other reviews of CT and PCVE evaluation literature conducted in recent years may help shed some light on any overarching trends in these fields.

Given the diverse theoretical underpinnings and inclusion criteria employed, any comparisons discussed here should be taken only as tentative and further research should be conducted to validate them. In particular, in the future, **mapping and stocktaking exercises akin to the present study should be undertaken regularly and with a comparable methodology** with a view to assessing developments in CT and PCVE evaluation, collecting lessons learned and promising practices, identifying existing gaps and shortcomings and formulating recommendations for tackling these. As discussed in Chapter 4, such **future exercises should also be provided with the resources and means to gain access to CT and PCVE initiatives' evaluators and beneficiaries**, with a view to providing a more in-depth and robust assessment of evaluations analysed. Furthermore, **future efforts akin to this study should also consider focusing on conducting a more in-depth comparative analysis of existing reviews**, assessing how different factors and inclusion criteria were defined by different studies and what implications this held for studies' findings.[100]

Our study identified more evaluations than other similar reviews in the field of CT have done in the past. Therefore, although the reviews did not use the same search protocol, it seems that, at least on the surface, CT and PCVE evaluation has gained traction in recent years. In addition, given the inclusion criteria employed by the current study, it appears that greater emphasis is being placed on collecting primary data for the purpose of evaluating CT or PCVE initiatives. Lum et al. (2006), in the first systematic review of evidence underpinning CT efforts, found that out of 20,000 studies concerned with terrorism, only seven discussed moderately rigorous evaluations and that little scientific knowledge existed as regards the effectiveness of CT work. A similar conclusion was reached by Christmann (2012), whose review of literature focused on PCVE initiatives; he found that very few studies contained empirical data or systematic analysis, even though great emphasis had been placed in previous years on investigating violent extremism and radicalisation issues. More recently, Feddes & Galluci (2015) concluded that hardly any evidence-based PCVE intervention existed. Furthermore, they found that most of the PCVE evaluation manuscripts published between 1990 and July 2014 relied on anecdotal evidence, and that only 16 out of 135 publications reviewed used primary quantitative or qualitative data. As for the present study, 33 out of the 37 manuscripts included in the inventory and published after 2013 rely on primary data and all publications employ quantitative or qualitative data in their evaluations.

The increase in volume of empirically based evaluations published in the last five years seems to indicate that greater attention is being placed on the research gaps and lack of evidence marring CT and PCVE work. This appears to be consistent with the observation made by Feddes & Gallucci (2015) that 88 per cent of manuscripts included in their review were published after 2008. A trend towards a greater volume of evaluation in the fields of CT and PCVE has also been noted by Gielen (2017) who, in her realist review of PCVE evaluations, describes this field has having 'taken flight' between 2016 and 2017, during which half of the manuscripts reviewed in her work had been published. Similarly, 17 out of the 37 post-2013 manuscripts reviewed in this work were published from 2016 onwards.

As emphasised by Gielen (2017), however, while the overall increase in empirically based evaluations published is a positive sign, this trend should be further investigated and dissected, assessing the focus,

[100] Romaniuk & Chowdhury Fink (2012); Dawson et al. (2014).

quality and overall robustness of evaluations undertaken in recent years. For example, as regards this study's finding, while it appears that the overall volume of evaluations produced in recent years has grown, caveats may apply concerning the extent to which these fill or try to address long-standing research gaps in the evidence base underpinning CT and PCVE work.[101] In particular, it appears that PCVE-specific programmes entailing educational, communications and capacity-building activities are over represented in the study's inventory, to the detriment of CT-specific initiatives, as well as disengagement- and deradicalisation-focused PCVE work. While an increasing focus on designing and implementing preventative initiatives and interventions is expected and noted in CT and PCVE literature,[102] the range of such programmes available in the study's inventory seems to suggest that the increase in CT and PCVE evaluations may be closely tied to the greater evaluability of such initiatives,[103] rather than to a more distributed evaluation culture change and advancement within CT and PCVE more broadly.

Significant gaps continue to persist in the quality and robustness of evaluation design and methods employed in a sizeable proportion of the inventory assessed. For example, and most notably, in most instances evaluations were characterised by an additionality-focused design, despite attempting to investigate with a focus on the impact/outcomes of the evaluated interventions. As such, the evaluations fall short of observing changes and attributing these with a sufficient degree of scientific confidence to the CT or PCVE initiative being evaluated.

A limitation observed with regard to evaluation purposes stems from the scarcity of evaluations that investigate the economic aspects of the evaluated interventions. While costs of interventions might be a contentious issue and measuring such costs is certainly difficult, the lack of evidence for improving cost-effectiveness or cost-efficiency of CT and PCVE initiatives is startling when compared to other fields. **Commissioners and evaluators should more systematically focus on the efficiency of evaluated interventions and, in time and with improvements in impact evaluation designs, funders, commissioners and implementers themselves should be interested in economic evaluations, cost-benefit and cost-effectiveness analyses** to determine what CT and PCVE policies and programmes offer best value for money. With a view to sustaining and expanding CT and PCVE evaluation efforts, agencies funding initiatives and research in these fields should **continue to invest in the evaluation of planned and existing initiatives in these policy areas**.

As emphasised by Lum et al. (2006), **minimum quality and robustness requirements should also be expected from future evaluations**, including the use of empirical data, multiple methodologies and adequate stakeholder engagement methods. As part of this push, commissioners and evaluators should **encourage greater transparency and clarity about evaluation methodologies and approaches used to allow better judgement of the evidence strength and learning.** Despite some improvements in the quality of CT and PCVE evaluations shown in this report, a substantial proportion of studies does not describe the adopted methodology. In the absence of other quality control mechanisms, such a description is necessary to make a judgement on whether the design was suitable and used appropriate methods, if the

[101] Ris & Ernstofer (2017).

[102] Romaniuk & Chowdhury Fink (2012).

[103] Davies (2013).

execution of the methods was effective, if appropriate data have been collected or used, and, ultimately, how strong is the evidence supporting study conclusions and recommendations. Providing more explicit definitions of the key concepts and including data collection tools such as survey questionnaires and topic guides in evaluation reports would not only allow such judgements to be made, but also facilitate learning and building on knowledge and experience gained in the past.

Finally, the review conducted on this study's CT and PCVE evaluations inventory highlighted a preponderance of impact or outcome evaluations, a significant number of which were flawed by a design and focus not suited to the drawing of scientifically robust conclusions about initiatives' outcomes and results. In a field such as that of CT and PCVE evaluation, which is characterised by normatively and politically charged concepts, external pressures, as well as time and resource constraints, **agencies commissioning evaluations should consider adopting evaluative approaches aimed at disaggregating and, to the extent possible, isolating and addressing individually the characteristics making CT and PCVE evaluation challenging**. This could facilitate a transfer to CT and PCVE of known approaches and strategies that have produced results in other fields marred by some of the same challenges and characteristics. This would likely entail a shift from an approach aimed at identifying definitive responses concerning the impact and results of individual CT and PCVE initiatives, towards a more gradual approach focusing on the progressive accumulation of knowledge produced through narrower evaluation approaches.

For instance, evaluation commissioners may **consider encouraging the undertaking of rigorous real-time evaluations (RTEs)** that have been used, inter alia, in the humanitarian context.[104] RTEs are not instruments contributing to investigating an initiative's impacts. Rather, they entail a participatory evaluative approach and are intended to provide immediate feedback during fieldwork with a view to (i) promoting learning; (ii) improving decision-making occurring within an initiative's implementation organisation; and (iii) responding to the needs of an initiative's managing and implementing stakeholders.[105] RTEs can be seen as a learning-oriented approach that is normally employed in contexts where those implementing a particular initiative are also interested in learning and responding in real time to evidence emerging from the evaluation, adapting their approach. To strengthen the approach, RTEs can be designed with built-in pauses and reflection points designed to take stock of results emerging from the evidence collected and informing the continuation and potential adjustment of an initiative's implementation.[106] To this end, with a view to maximizing the benefits of such undertakings, the requirements and the design of RTEs could also be established through real-time interactions between evaluators and beneficiaries (i.e. initiatives' implementers).

[104] Polastro at al. (2011); Grünewald & Binder (2010).

[105] Patton (2008); Ling (2012).

[106] Ling (2012).

6. Identifying issues and learning lessons from CT and PCVE evaluations

This chapter discusses the issues emerging and lessons learned derived from the study's inventory of CT and PCVE evaluations. Alongside the coding of initiatives and evaluation characteristics, as discussed in Chapter 5, the review of publications also focused on capturing lessons learned and recurring issues concerning the design and conduct of evaluations of CT and PCVE initiatives.

6.1. An overview of issues identified and lessons learned from the study's inventory

Throughout the study, the review and analysis of CT and PCVE evaluation manuscripts focused on identifying issues marring these fields and extracting lessons learned that had been identified and discussed directly by the authors of the manuscripts. As such, the study team adopted a broad definition of lessons learned with a view to capturing discussions of limitations and lessons learned *a priori* by evaluators. Nonetheless, lessons discussed in this chapter should be taken as illustrative examples of some of the challenges and lessons stemming from CT and PCVE evaluations, rather than as a comprehensive overview of issues and developments in these fields. For example, as a result of the approach taken, the vast majority of issues and lessons identified and discussed below pertain to individual evaluations, rather than to the overall field. Moreover, it should be noted that many of the issues and lessons, although stemming from a review of CT and PCVE evaluations, are not unique to these fields, but are in fact recognised in wider evaluation literature. Furthermore, the study did not focus on capturing issues and lessons learned about individual CT and PCVE initiatives, their design, implementation or results achieved. In other words, this chapter does not make any recommendations in terms of how to counter terrorism and radicalisation most effectively.

Using the approach discussed above and in Section 4.4, the study team identified 61 challenges and lessons learned across 30 of the publications reviewed. The remaining 18 manuscripts did not identify any significant challenges or lessons learned in the text. The study team attempted classification of the issues, challenges and lessons learned according to a range of high-level categories. Most were found to pertain to evaluation design (28 out of 61) and evaluation methods (25 out 61). Others were concerned with

evaluation reference data (16), inhibitors to evaluating the intervention (12), evaluation goals (8), evaluation outcomes (7), evaluation timeframe (6) and facilitators to evaluating the intervention (2).[107]

With regard to the validity of lessons extracted, for approximately half of the lessons learned (31 out of 61) the reviewed study was clear about what was meant by a 'lesson' and 'learned'. In over half of lessons learned (39 out of 61), it was clearly indicated by whom the lesson was learned. Similarly, most studies were clear about the evidence that underpinned these lessons (39 out of 61). Approximately half (30 out of 61) of the studies clearly indicated under which conditions the lessons would apply. Finally, in slightly under half of the lessons learned (26 out of 31), the lessons were found to be specific enough to guide practice.

Following the review of the publications in the inventory, the study team identified recurring themes in the literature and clustered findings about issues identified and lessons learned according to five overarching themes (discussed in the sections that follow):

- Inherent complexities of the field of CT and PCVE (Section 6.2)
- Challenges associated with measuring real-world phenomena (Section 6.3)
- Challenges associated with the evaluation design (Section 6.4)
- Practical difficulties of conducting evaluations (Section 6.5)
- Drawbacks and benefits of specific methods (Section 6.6).

This chapter concludes by discussing the implications of issues and lessons learned from the literature as well as recommendations stemming from challenges characterising the fields of CT and PCVE evaluation (Section 6.7)

6.2. Inherent complexities of the fields of CT and PCVE

Several authors have recognised that evaluations can struggle to account for the complexity of the phenomena investigated. Minor (2016), for instance, concludes that social research methods struggle with sensitive and complex topics, such as identity, religion and cultural affiliation. Reflecting on the evaluation of the Netherlands' national counterterrorism strategy 2011–2015, Noordegraaf et al. (2017) argue that the dynamics and uncertainty of the fields of terrorism and counterterrorism inhibit identification of cause and (unintended) effects. In an analysis of the effectiveness of targeted killings, Carson (2017) cites the complex nature of many recent conflicts as a potential threat to the validity the model used. In the context of this complexity, Argomaniz & Vidal-Diez (2015) and Noordegraaf et al. (2017) therefore emphasise the importance of subject matter expertise in evaluation. Terrorism is only a relatively young field of research, they argue, and therefore understanding of the root causes of terrorism and the ways it can be combatted is still limited. This hampers the interpretation of existing evaluation criteria in the context of CT and PCVE, for instance.

Specific aspects of the subject matter that complicate the evaluation have also been raised, such as: the target group, the nature of the intervention(s), the institutional complexity, and its security aspects.

[107] These categories were not mutually exclusive: some lessons applied to multiple categories.

Target group. The issue of selecting, labelling and having access to CT and PCVE initiatives' target groups is one frequently discussed in the existing literature.[108] In the study's inventory, Mitts (2017) argues that empirically evaluating PCVE interventions is an inherently challenging task, since they focus on the behaviour of a very small minority in society and they almost always selectively target specific communities. Webster et al. (2017) also warn that the heterogeneous nature of al-Qa'ida extremists, in terms of their religious and political beliefs, inhibits the extent to which evaluation findings can be extrapolated to other contexts.

Stakeholders. Countering the terrorism threat requires extensive collaboration between services including the police, intelligence services, education, social and health services. This is because the terrorism threat is associated with a range of risk factors. Collaboration may be national, international or local. This poses challenges not only to the design and implementation of interventions in this domain, but also to the evaluation of initiatives. Noordergraaf et al. (2017) argue that the complex and overlapping networks of stakeholders in CT policy, such as professionals in the social and security domain, international, national and local governments, NGO's and private sector organisations, make it more difficult to decide on which stakeholders should be engaged in the evaluation and by what method that engagement should take place, especially bearing in mind the often limited resources available for evaluation.

Security aspects. Security aspects were raised in a few manuscripts as a hurdle in the evaluation of CT policies or interventions. Aldrich (2014) cites security concerns as a limitation to experimental design. Similarly, Veldhuis et al. (2014) highlight the notion of fear-based institutionalised barriers. In their evaluation of deradicalisation efforts in Dutch prisons, they argue that limited access to prisons was a barrier to developing a knowledge base that is rooted in evidence and realistic appraisals.

Interventions. When drawing lessons about the evaluation and highlighting some of its challenges, one author referred to the nature of CT or PCVE interventions. In his evaluation of approaches to tackle violent extremism in three major European cities, Vermeulen (2014) describes how a local-level counterterrorism practice quickly devolves into a complicated, multiplex discussion about immigration, belonging, citizenship, Islam and the position of Muslim communities in Western cities. Above all, he argued that better micro-level understanding of the implementation of such policies in cities, city districts and neighbourhoods is needed. The evolving nature of PCVE interventions in response to the ever-changing and dynamic environment in this area, as well as a lack of clear definition of objectives, or of how success will be measured, pose challenges to evaluation that are recognised among evaluators.[109]

Concepts and definitions. A number of authors suggested that evaluations struggle to account for the complexity of the phenomena investigated due to the lack of clarity of overarching CT and PCVE concepts and our limited understanding of the drivers of violence.[110] This represents a practical challenge for evaluations in many phases. From the moment of defining the objectives of an intervention and scoping its evaluation accordingly, to developing data collection tools that use terminology that may be perceived as charged, contested and open to various interpretations (see Section 6.6).

[108] Heath-Kelley (2012).

[109] Marret et al. (2017).

[110] Hirschfield et al. (2012); Khalil & Zeuthen (2014a; 2014b); Kurtz et al. (2010).

6.3. Challenges associated with measuring real-world phenomena

A number of sources reflected on specific aspects of the methods deployed in the evaluation and the challenges of accounting for a variety of factors and real-world dynamics contributing to the results of particular CT initiatives.

Rarity of events and lack of outcome measures. Terrorist attacks can be characterised by their relatively low frequency and their potential to have a high or very high impact. In contrast to some of its neighbouring countries, in recent years there have been no successful terrorist attacks in the Netherlands. The low incidence rate or in this case absence of observations makes it challenging to measure the effectiveness of CT measures. Indeed, Noordegraaf et al. (2017) acknowledge that measuring the ultimate outcome of CT policy, the absence of a terrorist attack, is impossible for this reason. In light of this, it is expected that for pragmatic reasons evaluations in this field must rely on heuristics based on risk factors of terrorism, generating a mismatch between what evaluations aspire to assess and what can be actually measured.

Lack of available outcome metrics. Measuring risk factors such as radicalisation or extremism is also difficult. Several authors claim that a lack of available metrics hampers evaluation considerably. For instance, Hirschfield et al. (2012) argue that extremism is a 'mind set' and 'one cannot evaluate a mind-set.' Others run into similar issues, highlighting the difficulty of measuring abstract concepts such as personal change[111] or marginalisation.[112] Wilner & Rigato (2017) acknowledge that metrics taking account of the success and/or failure of specific CT and PCVE interventions have yet to be developed. This lesson concurs with broader literature on the subject that concludes that there is little coherence in metrics used to measure the success of PCVE interventions and there are no commonly agreed metrics to measure behavioural changes.[113]

Measuring long-term effects. Several authors highlight the importance of measuring long-term effects in evaluation, but Carson (2017), Argomaniz & Vidal-Diez (2015) and Kurtz et al. (2016) also acknowledge the difficulties of doing so. In his evaluation of targeted killings, Carson (2017) argues that the long-term implications of this particular intervention may have yet to be realized. In this context he refers to the intelligence concurrently seized in incidents, such as the killing of Osama bin Laden, which has taken some time to prove its value. Similarly, employment programmes such as the one evaluated by Kurtz et al. (2016) are unlikely to take effect in the short term, and the authors call for an investment in studying their long-term effects on violent extremism. According to the authors, panel and longitudinal studies on trends in employment and the impacts of employment programmes are required to fully assess the roles of long-term employment and improved economic conditions on political violence and violent extremism. This lesson corresponds with previous studies and the experiences of other authors who point out that

[111] Johns et al. (2014).

[112] Khalil & Zeuthen (2014a; 2014b).

[113] Romaniuk (2015).

their evaluations were unable to record long-term effects.[114] This highlights a possible mismatch between what PCVE evaluations aspire to assess and what they actually manage to measure.

Accounting for social norms and expectations. Several authors highlight the limitations of their evaluation approaches in terms of taking into account the impact of social norms and expectations in data collected from beneficiaries and target stakeholder groups of CT and PCVE initiatives.[115] The biases associated with self-reporting are well documented. Some authors provide relevant examples in the context of CT evaluation. Johns et al. (2014) and Minor (2016), for instance, explain that interviewees may potentially be prone to giving socially desirable answers, 'wanting to please interviewers' and thereby potentially skewing recollections of their experiences. Similarly, Krafchik & Ryszkowska (2011) explain that some participants refrained from answering certain parts of the attitude survey, indicating that they were reluctant to provide what would appear to be the 'wrong' answer to the interviewer.

Accounting for exposure to an intervention. Aldrich (2014) describes how individuals are often exposed to multiple interventions simultaneously. It is therefore, in his view, important to account for the level of dosage or exposure to a particular initiative that individuals taking part in multiple interventions have had. He laments the challenges associated with making a precise measurement or estimate of the dosage received or exposure to a particular initiative by individuals consulted as part of an evaluation.

6.4. Challenges associated with evaluation design

Many authors of the manuscripts in the inventory acknowledge challenges in making claims about causal interference. In particular, they report how difficult it is to isolate effects and attribute them to the policy or intervention that is being evaluated.

Difficulties with claiming causality and conducting experiments and quasi-experiments. The most persuasive way to design an evaluation that successfully addresses causal questions is to conduct an experiment. Indeed, several authors in the study inventory explicitly indicate an ambition to use an experimental or quasi-experimental design. Aldrich (2012), for instance, says that '[f]uture research… should seek to carry out randomized, double-blind field experiments to ensure that we can better measure the impact of interventions and treatments.' However, the real-world complexities of the CT and PCVE context do not always lend themselves to setting up such designs. Two important aspects of experimental designs were highlighted in particular: control groups and randomisation. Citing Lub's (2013) analysis of the validity of 'theories of change' of anti-polarisation and anti-radicalisation intervention, Feddes et al. (2015) state that it is often not possible to include a control group in these studies for practical reasons. Because of this, making causal claims is difficult and potential positive or negative effects of the intervention, if any, may remain undetected. Similarly, Aldrich (2014) and Mitts (2017) highlight the difficulties of randomisation in a CT and PCVE setting.

[114] Feddes & Gallucci (2015).

[115] Aldrich (2014); Boyd-MacMillan (2016); Feddes et al. (2015); Johns et al. (2014); Khalil & Zeuthen (2014a; 2014b); Krafchik & Ryszkowska (2011); Minor (2016).

In the absence of evaluation designs that would satisfy conditions for causal interference, it is not surprising that several authors report on the challenges associated with isolating effects and attributing them to the interventions. Without experiments or quasi-experiments, it is impossible, for example, to rule out other mechanisms that could have been responsible for the effects of repressive CT measures. Carson accepts that targeted killings of high profile al-Qa'ida leaders are certainly not the only explanations for a decrease or increase in the frequency and hazard of incidents perpetrated by the global jihadist movement. In discussing her findings, she acknowledges that an increase in terrorism could also be representative of terrorists ignoring the risks associated with their act or of a lack of consideration for possible punishment. Carson recommends that future studies should narrow the mechanisms responsible for the effects found and continue to replicate them in other contexts with other controls.

It is also difficult to attribute outcomes to the deradicalisation interventions alone. While Mitts (2017) found consistent patterns in his study for large number of community engagement activities and tens of thousands of individuals, the results were inconclusive with respect to whether the reduction in pro-ISIS rhetoric was caused by deradicalization or by the suppression of political expression in these areas. In this context, Krafchik & Ryszkowska (2011) note that in these community settings multiple interventions are often simultaneously occurring, which makes it difficult to attribute outcomes to any one intervention. Feddes et al. (2015) flag another difficulty in separating intervention effects from other factors. In their evaluation, participants knew beforehand that they would receive training aimed at increasing self-confidence, which dealt with judgement of cultural differences, and in which they would learn how to deal with conflicts. It would be difficult to tell whether improved outcomes could be attributed to the actual intervention or to an expectancy or placebo effect on the part of the participants.

Even in a contained setting such as a prison, both Webster et al. (2017) and Veldhuis et al. (2011) report methodological difficulties in identifying causality. Webster et al. (2017) explain that a number of factors may also affect offender outcomes in a pilot project with so-called Structured Risk Guidance (SRG) for Extremist Offenders, developed by Her Majesty's Prison and Probation Service (HMPPS), such as changing beliefs and ideologies as a result of exposure to imams and intervention programmes in the prison, which may also have played a part in influencing the reported outcomes.

Challenges of adopting a longitudinal study approach. Several authors highlighted problems with conducting multiple data collection moments to assess before and after effects and the long-term implications of complex CT and PCVE endeavours, thus limiting the robustness of evaluation findings. This issue can also be considered in the context of the limited collection and availability of baseline data for purposes relevant to PCVE and CT. Calls for more systematic, continuous or regular baseline data collection are long-standing in the literature on CT and PCVE evaluation.[116] In terms of data collection, authors noted problems with access to routine monitoring data and limitations to datasets that are available,[117] as well as difficulties in establishing a baseline and tracking progress over time.[118]

[116] Aldrich (2012, 2014); Hirschfield et al. (2012); Vermeulen (2014).

[117] Lobley et al. (2001).

[118] Hirschfield et al. (2012).

The lack of a theory of change. The challenges associated with establishing causality and attributing outcomes to an intervention may be due to limitations in the methodological design such as, for example, the absence of a control group or randomisation, as explained above. But manuscripts in the inventory also highlight that the evaluation of interventions can often be hampered by the absence of a theory of change; in addition they stress the importance of identifying the mechanisms behind policies and spelling out often implicit assumptions and hypotheses about how these are expected to work. Veldhuis et al. (2011), for instance, note that little theoretical work is available that provides insight into the nature and dynamics of prisoner radicalization. The authors argue that prior to testing interventions empirically, future research to develop theoretical models on the dynamics of violent extremism in relation to the correctional system (e.g. both during and after imprisonment) is needed. Winter et al. (2015) propose an approach to addressing this challenge by using Pawson & Tilley's (1997) 'Context-Mechanism-Outcome' model to analyse two interventions and show how this model, in combination with two theoretical frameworks, can contribute to the identification of mechanisms behind policies and an understanding of the policies' functioning, side effects and outcomes. Pawson & Tilley's model involves the development of evaluative frameworks designed to help researchers understand what works, for whom, and under what circumstances. This approach enables evaluators to investigate which aspects of an intervention are effective or ineffective, and what contextual factors and enablers affected the results achieved.

6.5. Practical difficulties of conducting evaluations

In addition to the research and design challenges discussed above, the literature reviewed during the study highlights a series of practical challenges and obstacles that negatively affect CT and PCVE evaluation implementation.

Resource constraints. Several authors highlight in their discussions resource-related constraints and challenges that negatively affected the design, implementation and quality of evaluations conducted. Boyd-MacMillan (2016), Johns et al. (2014), Wilner et al. (2017) and Minor (2016) focus in particular on time constraints. More specifically, two different examples of time constraints and their adverse effects are discussed: firstly, the difficulty of conducting evaluative activities at the most appropriate or suitable time, thus missing opportunities to collect relevant or optimal data; and second the lack of sufficient time to conduct evaluation work appropriately.

Johns et al. (2014) emphasise the difficulty encountered in identifying and obtaining access to the beneficiaries of an initiative at the most suitable and significant times from an evaluative perspective, including in particular before an initiative is established and launched with a view to conducting baseline measurements facilitating the subsequent attribution of results observed to an initiative. Boyd-MacMillan (2016), Wilner et al. (2017) and Minor (2016) focus on the lack of sufficient time being allocated to evaluation efforts, undermining the planning and conduct of rigorous measurements and analysis of data.

Financial constraints are also highlighted as a factor hampering the design and conduct of robust evaluations. For example, Aldrich (2014) emphasises that the volume of survey participants engaged during an evaluation was meagre, if compared with comparable research efforts in the same region funded by other US federal actors, and that this was a function of limited financial resources. Similarly, Boyd-MacMillan (2016) and Wilner et al. (2017) respectively emphasise how the lack of adequate funding for

conducting and sponsoring data collection and analysis resulted in a smaller sample size and data pool than originally envisioned, and in the overall lack of an empirical evaluative approach being implemented. Overall, the scarcity of funding for empirical research endeavours was noted by several authors, who found it to be obstructing the formation of a CT and PCVE knowledge base rooted in evidence and robust empirical appraisals.[119]

Difficulty accessing information about interventions and effects. The CT field is marked by secrecy, between institutions and the public but also within 'need to know' circles of experts.[120] This makes it challenging for evaluators to access the information and data necessary to measure the effectiveness of specific interventions or policies (see Section 6.2). Sheikh et al. (2012), for instance, flag that accessing information about the intervention can be difficult, which has meant that there is insufficient data to make firm conclusions about the impact and benefits of some elements of the programme.

Access to data. A number of authors also lamented the difficulty of obtaining access to key stakeholders and beneficiaries for the purpose of data collection and findings validation (see Section 6.2). This ultimately results in a weaker evidence base underpinning evaluation findings and results. For example, Webster et al. (2017) report being unable to access the target of a programme due to concerns with how engagement with researchers may have affected their individual disengagement and deradicalisation trajectory, depriving the study of the beneficiaries' perspective and input. In the context of a large, national-level initiative, Sheikh et al. (2012) discuss practical challenges stemming from being unable to identify relevant beneficiaries who had taken part in a particular programme and who should have been consulted through interviews. In other instances, challenges involving access to data stem from social norms and participant bias (see Section 6.3). For example, Krafchik & Ryszkowska (2011) and Minor (2016) report concerns about the veracity of data collected from beneficiaries and interviewees who may be reluctant to share views or feedback that they believe may be perceived as 'wrong' by interviewers.

Difficulties with sample size. Several authors acknowledge that there are challenges stemming from the limited size and scope of the samples available and accessible during the data collection phase. These constraints are often found to undermine the robustness and extent of general conclusions at the end of an evaluation. For example, Krafchik & Ryszkowska (2011) emphasise how the limited number of beneficiaries accessible for interviews and attitude surveys made it challenging for any general conclusion to be drawn about the initiative evaluated. It should be noted that a sample size is only one consideration in making inferences from a sample to a wider population. Other factors, such as sampling method and response rate, are important too. Feddes et al. (2015) found that the small and homogeneous sample of beneficiaries engaged in their evaluation implied that no analysis of the effects of context and age on overall outcomes could be assessed. Similarly, Sheikh et al. (2012) lament a lack of diversity in the perspectives of the sample populations engaged during their evaluation as these tended to belong to homogeneous stakeholder groups – which echoes the point made earlier about the need to capture a multitude of views and perspectives (Section 5.3.2). As with the other practical evaluation challenges discussed above, sample size and scope constraints are found to be the product of a multiplicity of factors,

[119] Aldrich (2014); Boyd-MacMillan (2016); Wilner et al. (2017); Kurtz et al. (2016).

[120] Reding et al. (2014).

including lack of sufficient funding,[121] compressed evaluation timelines,[122] and lack of access to key stakeholders due to ethical and sensitivity issues.[123]

6.6. Drawbacks and benefits of specific methods

The final set of lessons learned gleaned from CT and PCVE evaluation authors pertains to the inherent drawbacks and benefits of the evaluation methods employed in their work. While the challenges and benefits discussed below and in the inventory sources are not unique to the fields of CT and PCVE, and they are well documented in social research literature, their significance from a research perspective compels us to discuss them.

Constraints of model-based investigations. A number of sources included in the study inventory rely on a model-based investigation and analysis for researching complex CT and PCVE phenomena. While model-based approaches enable authors to investigate complex policy issues within the research and resource constraints discussed above, model-based investigation is (i) inherently limited as regards its ability to account for complex dynamics and a range of factors and drivers that researchers may be unaware of; and (ii) dependent on the availability and quality of data available to inform the model devised for investigation. To this end, Williams et al. (2015) note that their study's model and resulting findings should be tested by means of experiments as factors unaccounted for in the model may affect the results and observations recorded. Similarly, Carson (2017) notes that while several controls and sensitivity analyses were implemented throughout her work, factors unaccounted for in the model proposed may be responsible for the observations made and results obtained. To corroborate her statement, she notes how a comparable quantitative, model-based investigation of the same phenomenon (i.e. targeted killings) by Jordan (2014) came to opposing conclusions as to the risk and potential impacts of operational blowback.

Constraints of survey instruments. As discussed in Chapter 5, several evaluations employing quantitative methods relied on surveys for collecting data. While this method enables researchers to effectively collect a range of data to be subsequently analysed qualitatively and quantitatively, some of its constraints are noted in the literature reviewed. For example, Khalil & Zeuthen (2014a; 2014b) note that many of the concepts used in CT and PCVE evaluation studies are ambiguous and open to interpretation. As such, respondents and researchers may interpret and administer questionnaires inconsistently, generating doubts as to the reliability and comparability of resultant data. To this end, the authors suggest embedding definitions of problematic concepts in survey instruments and protocols to minimise the risk of data inconsistencies. Similarly, Liht & Savage (2013) highlight the challenges of psychometric coding and the risks connected with designing survey instruments carrying unintended cognitive loads that may skew or bias data collected from respondents.

[121] Aldrich (2014).

[122] Johns et al. (2014).

[123] McNeal (2014); Webster et al. (2017).

Importance of triangulation and strengths of qualitative methods. Finally, lessons identified by authors focus also on the benefits and strengths of the methodologies employed in evaluation work. For instance, McNeal (2014) discusses the strengths of qualitative research in general and of the case study method in particular. Case studies, McNeal holds, are particularly suited for drawing conclusions as to the 'how' and 'why' of contemporary social phenomena the boundaries of which are difficult to discern from those of the context in which they exist. McNeal also reiterates the importance of using a volume of varied sources of evidence to be categorised, summarised and analysed using the same rigorous approach of a quantitative researcher. He suggests that this enables researchers to triangulate and validate findings with a degree of construct validity, external validity and reliability akin to that achieved using more complex research designs and methods.

6.7. Discussion and recommendations

This chapter has presented findings stemming from an analysis of lessons learned concerning the evaluation of CT and PCVE initiatives as identified by the authors of publications included in this study's CT and PCVE evaluations inventory. Lessons identified were found to relate to five recurring themes.

Firstly, authors highlighted the inherent complexities of the fields of CT and PCVE (see Section 6.2), including challenges stemming from the lack of clear and shared definitions for key CT and PCVE concepts, security concerns, complex stakeholder landscapes, a limited understanding of initiatives to be evaluated, and the presence of overtly small and homogeneous target groups. Secondly, authors discussed a number of challenges associated with measuring real-world phenomena connected to CT and PCVE (see Section 6.3), including the lack of outcome measures, outcome metrics, opportunities to measure long-term initiative effects, instruments accounting for social norms and expectations, and measures tracking exposure to interventions, which hampered their ability to design and conduct robust evaluations. These findings and lessons suggest that, in addition to the need for more evaluations in CT and PCVE to be conducted (see Section 5.4), **further research is needed on the dynamics, drivers and factors governing the phenomena of radicalisation, violent extremism and terrorism.** Furthermore, emphasis within research should also be placed on **developing new evaluation designs, frameworks and approaches for conducting evaluation in the CT and PCVE policy areas**.[124]

Turning to evaluation challenges, the studies reviewed highlighted a number of lessons learned as regards evaluation design (see Section 6.4) and specific methods employed during an evaluation (see Section 6.6). In particular, these included difficulties with evaluating initiatives lacking a clear theory of change, challenges in isolating initiatives' effects and attributing these to a single programme or initiative, constraints faced when establishing randomised experiments and quasi-experiments, the impossibility of adopting a longitudinal study approach, and constraints associated with data collection and analysis methods such as survey and models.

As discussed in Section 4.2, employing experimental or quasi-experimental designs is the most suitable approach for successfully addressing causal questions and establishing a link attributing isolated effects to

[124] Hirschfield et al. (2012); Khalil & Zeuthen (2014a; 2014b); Kurtz et al. (2016); Ris & Ernstofer (2017).

an initiative being evaluated.[125] In the fields of CT and PCVE, experimental and quasi-experimental designs are still far from becoming established or at least being as commonly used as they are in other fields with similar constraints, such as gangs' desistance.[126] The practical and ethical challenges of such designs broadly apply to all fields were they are employed and should not be seen as unique to CT and PCVE.[127] With this in mind, we believe **greater efforts could be made to ensure that wherever possible (quasi-)experimental designs are employed within CT and PCVE evaluation**. For example, experimental and quasi-experimental designs could be employed for the evaluation of group-focused preventative measures and initiatives that target non-radicals through educational, communications and capacity-building activities.[128] Furthermore, in the context of individual-focused initiatives that pursue disengagement and deradicalisation goals, such designs could be adjusted to ensure compliance with ethical requirements. For example, withdrawing an intervention from control groups could be temporary (so called wait list control) so that their participation in the intervention could be delayed until the measurements necessary for making comparisons between the groups are completed.

Lastly, studies reviewed discussed the practical difficulties encountered when implementing evaluations (see Section 6.5). In this regard, the analysis of lessons learned evidenced significant issues faced by researchers in terms of obtaining access to key stakeholders,[129] as well as to data required to complete a robust assessment of an initiative.[130] While similar challenges are faced in other policy domains,[131] issues connected to sensitivity and feasibility of access to the initiatives to be evaluated continue to hamper or, at best, significantly influence the design and approach of a number of recent CT and PCVE assessments.[132]

While in recent years researchers have been called on to provide answers as to the evidence base underpinning CT and PCVE work, and the impact and results produced by it, their access to key data and stakeholders continues to be limited.[133] Multiple approaches could be considered for implementation by agencies commissioning and implementing CT and PCVE initiatives and research to mitigate the impact of such constraints on evaluations and research conducted in these fields, including:

1. **Design and implement approaches for providing researchers with access to sensitive information on CT and PCVE initiatives and beneficiaries.** While CT and PCVE are fields characterised by significant political pressures and ethical and sensitivity issues, these dynamics are not unique to them. Furthermore, similar constraints should not be seen as insurmountable given the progress that has been made in adopting robust, empirical evaluation approaches in fields with comparable challenges, such as criminology and gangs' desistance.[134]

[125] Shadish et al. (2002).

[126] Davies et al. (2017).

[127] Resnik (2008); Fives et al. (2013); Nardini (2014).

[128] Feddes & Gallucci (2015).

[129] Sheikh et al. (2012); Noordergraaf et al. (2017); Webster et al. (2017).

[130] Krafchik & Ryszkowska (2011); Minor (2016).

[131] Davies et al. (2017); Fast & Neufeldt (2005).

[132] Jordan (2014); McNeal (2014); Carson (2017).

[133] Lum et al. (2006); Ris & Ernstofer (2017).

[134] Davies et al. (2017).

2. **Facilitate the establishment of mechanisms for data sharing among researchers.** In addition to the fact that limited empirical research is conducted within the fields of CT and PCVE, challenges in these fields are further exacerbated by the lack of data-sharing platforms among researchers. Research and evaluation commission agencies could consider generating guidelines for standardising data collection and recording procedures with a view to establishing a repository of anonymised data from CT and PCVE evaluations and research. Equally, existing platforms could be used or popularised in the CT and PCVE evaluation environment.[135]
3. **Collect regular baseline measurements and identify proxy measures and alternative indicators to mitigate the impact of data gaps.** As evidenced by publications reviewed over the course of this study, several CT and PCVE evaluators lament the lack of regular baseline measurements being conducted for metrics and indicators relevant to CT and PCVE goals.[136] Agencies commissioning CT and PCVE initiatives and evaluations should consider investing in identifying relevant baseline metrics to be recorded quantitatively at regular intervals. While individual programme implementers and evaluators could collect baseline data more systematically as part of their work, there is merit in considering a more structured and centralised approach for routine data to be collected on a regular basis at a higher level.
4. **Develop a framework for measurement in CT and PCVE, and identify alternative metrics and indicators** to (i) measure objective goals that may be used as proxies to investigate overall CT and PCVE trends within target populations; and (ii) to contribute towards developing a more consistent approach to evaluation in these fields.[137]

To conclude, a review of findings from this study and from previous comparable research efforts suggests that over the last decade CT and PCVE evaluation has gained traction and that the quality and robustness of studies undertaken has increased.[138] But several shortcomings, constraints and gaps continue to mar CT and PCVE evaluation. In particular, existing literature highlights a lack of capacity available to conduct a higher volume of robust CT and PCVE evaluations, as well as an emphasis on a quest for definitive metrics, findings or results meant to solve the riddles of CT and PCVE without an adequately developed evaluation capacity and theoretical underpinnings (see Section 6.2, 6.3 and 6.4).[139]

As discussed in Section 5.4, in the context of CT and PCVE evaluation, there may be merit in adopting approaches aimed at disentangling the complexities of this field as well as pursuing knowledge development focusing on the progressive accumulation of insights and advances tackling individual aspects and challenges of CT and PCVE evaluation.

[135] See, for instance, the Dataverse Project: https://dataverse.org/

[136] Aldrich (2012, 2014); Hirschfield et al. (2012); Vermeulen (2014).

[137] See, for example: Baruch et al. (forthcoming); Helmus et al. (2017); Feddes & Gallucci (2015).

[138] Lum et al. (2006); Christmann (2012); Romaniuk & Chowdhury Fink (2012); Feddes & Gallucci (2015); Gielen (2017).

[139] Feddes & Gallucci (2015); Marret et al. (2017); Gielen (2017).

To this end, **evaluators and commissioning agencies should consider embedding in their approach to CT and PCVE evaluation a 'co-production' ethos for knowledge development**. The concept of 'co-production' describes a situation in which inputs contributed from individuals who are not 'in' the same organisation (including both users and service providers) are transformed into goods and services.[140] Although the concept was first introduced in the 1970s, its use has gained significant traction in recent years, particularly in discussions concerning public policy, health and social care.[141] While different interpretations exist of what 'co-production' of knowledge entails in practice, this can be seen as a dynamic, experimental and reflective process sustained by different forms of engagement, interactions and social relations.[142] The purpose of moving towards co-production could be to contribute to: (i) generating new forms and approaches for initiatives to be implemented; (ii) generating values beyond economic value; and (iii) generating new insights and research practices that are relevant to different disciplines and practices within the spectrum of CT and PCVE research, contributing to a progressive accumulation of knowledge.[143]

Finally, it is noted that this chapter and its present, concluding section have presented a limited volume of practice-oriented recommendations pertaining to the design and execution of CT and PCVE evaluation. This is not surprising considering the broad and heterogeneous nature of the fields of CT and PCVE evaluation and the heterogeneity of the evaluations included in the study inventory. To be sure, the breadth of CT and PCVE strategies, policies and programmes and of their respective evaluations included in the inventory enabled the study team to identify a broad array of existing challenges and issues in these fields. However, the limited volume of comparable initiatives and evaluations, along with the importance of always embedding context-specific considerations and adjustments in the commissioning, designing and undertaking of CT and PCVE initiatives and evaluations, limited the study team's ability to provide practical recommendations and reflections. A broader discussion of existing tools and products that provide practical resources and recommendations for the undertaking of CT and PCVE evaluations is presented in the next chapter.

[140] Ostrom (1996).

[141] Needham (2008); Carr (2016).

[142] Filipe et al. (2017); Fast & Neufeldt (2005).

[143] Filipe et al. (2017).

7. Overall conclusions and recommendations

This chapter presents a summary of study results and recommendations. It first gives a synthesis of the study's findings, reviewing and responding to the research questions presented in Chapter 1. Following this, the chapter summarises the recommendations formulated throughout the report, focusing also on additional practical considerations and approaches that could be taken in the future to improve evaluations. Ways are suggested in which some of the challenges and shortcomings of CT and PCVE evaluations identified over the course of this study may be addressed by commissioners and implementers. The chapter concludes with observations and reflections formulated by the study team members as regards the research conducted and lessons learned from it.

7.1. Answering research questions

The following sections provide an overview of how the analysis conducted over the course of this study has contributed to answering the underpinning research questions presented in Chapter 1.

Research question 1. What is meant by counterterrorism policy in the Netherlands and other countries? What different types of counterterrorism measures, policies and interventions can be distinguished in the Netherlands and abroad?

Over the past two decades, CT has been a key policy concern for Western countries, including the Netherlands. Governments, policy- and decision-makers have come under pressure to address terrorist and violent extremist threats with a view to preventing attacks from occurring. The literature reviewed and experts consulted (see Chapter 3) indicate that the scope, purpose and activities characterising CT have evolved in recent years. This is both a response to changes in the threat landscape and a result of a growing understanding of the terrorism and violent extremism.

In the aftermath of the 9/11 attacks, CT policy in many Western countries focused on coercive measures, relying on the role and activities of law enforcement and security agencies to prevent and tackle terrorist attacks. In the last decade, however, there has been a growing recognition of the importance of adopting broader, holistic approaches to preventing and tackling terrorist activities. This has led to the development of initiatives designed to build resilience to violent extremist ideologies, and to address what we understand to be the root causes of violent radicalisation leading to terrorism. Such initiatives are broadly encompassed under the label of PCVE and are now considered an integral part of European CT policies and approaches.

CT and PCVE are highly interconnected and overlapping policy areas and distinguishing between the two is not always possible. For the purpose of this study, a high-level characterisation of the two work strands was adopted:

- PCVE was taken to encapsulate initiatives that address drivers of violence and extremism; that build resilience or immunity to extremist ideologies; and that deter or disrupt recruitment and mobilisation by extremist groups, including by supporting the reintegration of (former) extremists and offenders.
- CT was taken to encapsulate initiatives designed to deter, disrupt or isolate groups that use terror, including by training and equipping the security apparatus; by increasing the capacity to prepare, prevent, protect or respond to terrorist incidents; and by interdicting or prosecuting through law enforcement activities.

In addition to this differentiation, it was found that CT and PCVE work can be highly varied and that different structural characteristics contribute to shaping different measures and approaches. As a result, the study developed a taxonomy for analysing CT and PCVE initiatives, characterising them according to their type, goal, target ideology, level of intervention, type of activities, group of focus, implementer, geographic scope and foreseen duration (see Section 4.3). It was found that no single CT and PCVE blueprint appears to exist and that a wide array of approaches and programmes have been designed and implemented in recent years, spanning from broad communications campaigns for marginalised communities, to targeted military measures for terrorists living in conflict zones, to capacity-building interventions for individuals at risk of radicalisation.

Similarly to the trajectory and trends discussed above, it was found that in recent years Dutch authorities have designed and implemented a wide array of CT and PCVE legislation, policies and measures. In particular, the current Dutch CT strategy, designed to cover the 2016 to 2020 period, recognises the need for a comprehensive, multi-level CT approach that is threat-based, respects the rule of the law and emphasises both fixed and flexible measures. The strategy identifies five activities and areas for intervention:[144] procure (i.e. in a timely way gather and asses intelligence about (potential) threats to Dutch national security and national interests abroad); prevent (i.e. prevent and disrupt extremism and foil terrorist attacks before they occur); protect (i.e. protect people, property and vital processes from extremist and terrorist threats, both in the physical and virtual domains); prepare (i.e. prepare optimally for extremist and terrorist violence and its consequences); and pursue (i.e. enforce the law in the face of extremism and terrorism). Furthermore, and coherently with the framework laid out in the national CT strategy, it was found that a wide range of PCVE programmes have been developed and launched in the Netherlands in recent years, including initiatives that are both prison- and community-based.

[144] NCTV (2016a).

Research question 2. What evaluations of counterterrorism and preventing and countering violent extremism policies have been conducted over the last five years in the Netherlands and abroad?

A pivotal aspect of this study entailed the production of an inventory of CT and PCVE evaluations conducted in the Netherlands and abroad since January 2013. While it is recognised that the process of compiling such an inventory was to an extent arbitrary, the study team adopted an approach designed to reduce ambiguity about the selection of sources and reduce the risk of overlooking or excluding any relevant publications (see Chapter 2 and Annex B). Nonetheless, as PCVE, CT and evaluation are complex policy and research areas that do not lend themselves well to clear-cut definitions and inclusion criteria, should this study be repeated different research teams may come to different conclusions as regards the inclusion or exclusion of specific sources.

On the surface, the number of publications identified since 2013 by this study seems to suggest that CT and PCVE evaluation has gained traction in recent years and, given the inclusion criteria employed, that greater emphasis is being placed on collecting primary data for the purpose of evaluation. This appears to be consistent with previous observations made by Feddes & Gallucci (2015) and Gielen (2017), who describes this field has having 'taken flight' between 2016 and 2017. While the overall increase in published empirically based evaluations is a positive sign, close monitoring and investigation of their focus, quality and overall robustness is required.

Research questions 3 and 4. What can be said about counterterrorism and preventing and countering violent extremism evaluation characteristics? What are the differences and similarities between the identified evaluations of counterterrorism policies?

The analysis of the initiatives included in the study inventory and of their respective evaluations led to the identification of recurring characteristics, themes and noteworthy associations between initiatives and evaluation characteristics.

From the perspective of the CT and PCVE initiatives evaluated (see Section 5.2), the majority of publications included in the study inventory focus on initiatives undertaken at the level of programmes or interventions, whereas fewer publications focus on higher-level policies and strategies. In terms of initiative goals, those with a PCVE-specific or -relevant nature dominate (approximately 90 per cent of initiatives evaluated). In contrast, approximately 10 per cent of initiatives evaluated were CT-specific and only an additional 34 per cent had a clear CT-relevance. The initiatives evaluated were equally distributed across different levels of intervention (micro, meso, macro).

A small majority of initiatives evaluated by publications in the study's inventory focus exclusively on religious extremism, whereas approximately 34 per cent target different types of extremism. Only 12 per cent of initiatives reviewed exclusively targeted political extremism.

With regard to actors responsible for implementation, a small majority of initiatives was implemented by governmental actors, whereas non-governmental actors were responsible for 16 per cent and public-private partnerships for 30 per cent of initiatives respectively. Overall, it was observed that certain groups, namely communities and individuals vulnerable or at risk, were predominantly targeted in the sample of initiatives evaluated.

Looking at evaluation characteristics (see Section 5.3), it was found that evaluations included in the study's inventory comprise primarily process evaluations as well as impact- or outcome-focused evaluations. No economic evaluations were encountered in the study inventory. The majority of evaluations reviewed had no clear evaluation approach, although a number of theory-driven, realist and participatory evaluations were identified.

From an evaluation design perspective, the majority of publications included in the study's inventory focus on additionality. This design is used consistently for evaluating initiatives across different levels of intervention and for both CT- and PCVE-specific and -relevant work. Furthermore, a number of quasi-experimental and longitudinal designs was also adopted among impact or outcome evaluations.

From the point of view of methods, the majority of impact or outcome evaluations use quantitative methods alone or quantitative methods as part of a mixed approach. Conversely, most process evaluations rely solely on qualitative methods. Overall, methods employed typically include combinations of desk research, interviews, focus groups and case studies for the purpose of qualitative data collection, and surveys for quantitative data. As for analytical methods, unstated approaches and thematic analysis are most often used for qualitative data, whereas descriptive and advanced statistics are employed for quantitative analysis.

Looking at the timing of evaluations, the study's inventory comprises a majority of interim evaluations and a limited number of *ex post* evaluations and embedded (ongoing) evaluations. Different evaluation timings appear to be evenly distributed across different initiative types. No *ex ante* evaluations were recorded in the study's inventory.

Overall, when compared with results from previous reviews, the findings emerging from this study suggest that a growing volume of CT and PCVE evaluations is being undertaken and that the majority of these rely on primary data from multiple sources, perspectives and methods. While this finding is encouraging, there appear to be limits to the extent to which evaluation practice has advanced and grown evenly across all areas of CT and PCVE work (see Section 5.4). Furthermore, significant gaps and shortcomings continue to mar a number of evaluations. For example, some evaluations are characterised by designs that undermine their purported focus on and ability to come to robust conclusions about an initiative's impact.

Research question 5. What does the evaluation literature say about quality criteria for evaluations? To what extent do the identified counterterrorism and preventing and countering violent extremism evaluations meet these quality criteria?

Through a targeted review of literature (see Section 4.2), this study found that evaluation quality is a fluid concept. This is due to the fact that: (i) evaluation quality can be perceived differently by evaluators, stakeholders and decision-makers involved in an evaluative undertaking; and (ii) the concept of evaluation quality has been subject to change over time as standards for what constitutes or is considered good evaluation evolve and progress.

It was observed that different guidelines and principles exist for assessing evaluation quality and that usually these pertain to: (i) the technical quality of the information produced; (ii) the quality of the process used to obtain the information discussed; and (iii) the usefulness of the information produced.

However, an assessment of how well the articles included in this review meet the criteria established in the literature was not possible through a literature review only, and a comprehensive meta-evaluation of these sources would require more significant resources.

Given the evolving nature of the concept of quality in evaluation and the limitations of document review as the chosen method of making an assessment thereof, this report refrains from identifying commendable practices in the reviewed sources and from developing a prescriptive recipe for good CT and PCVE evaluations. A judgement of how well the articles included in this review meet the criteria established in the literature could be more appropriately made as a result of a systematic meta-evaluation of each source. This, according to the study's definition of evaluation, would require the collection and analysis of primary data to capture different perspectives on a particular evaluation against a set of criteria (e.g. validity, credibility, utility, cost-effectiveness, etc.). Since this was not possible, the study has aimed to capture quality assessments already performed, if any and if reported. In addition, what could be established is whether some minimal expectations for a promising evaluation have been met. These minimal expectations include (but are not limited to) capturing multiple perspectives (to limit bias), using grids, rubrics, scores or a set of indicators (to enhance transparency and objectivity), and providing recommendations (to increase the probability that evaluation results are used).

In the context of this study, and mindful of the limitations with regard to access to data, the study team has identified a series of customised metrics that capture information about the quality characteristics of an evaluation in the most factual way possible. This entailed focusing on review mechanisms employed to quality assess the evaluation, perspectives taken into account to draw conclusions, the consideration of unintended effects, and the presence and quality of recommendations formulated.

The majority of publications included in the study's inventory were subject to blind peer review, although in several instances it was not possible to determine which, if any, mechanism had been employed for review (see Section 5.3). The majority of evaluations also used multiple data sources or methods and considered different perspectives to draw conclusions, although almost a fifth relied on individual data sources. Furthermore, the vast majority of initiatives evaluated did not entail the use of grids, rubrics or scores for monitoring implementation work throughout its undertaking.

As regards formulating recommendations, it was found that the majority of evaluations did this, but that in no instance were the intended task-owners and timeframes for implementation indicated. As for unintended effects, half of evaluations reviewed investigated and reported on these, although this was found to be more common among evaluations of CT-specific and CT-relevant initiatives.

Research question 6. What practical lessons can be drawn on the basis of the existing evaluations regarding the evaluation of counterterrorism and preventing and countering violent extremism policy?

The study team adopted a broad definition of lessons learned with a view to capturing discussions of limitations and lessons learned *a priori* by evaluators. Some 61 lessons learned from 30 of the CT and PCVE evaluation publications reviewed were identified. Through a thematic analysis of lessons extracted, five overarching themes became apparent.

Firstly, the studies reviewed emphasised some of the inherent complexities of the fields of CT and PCVE and the impact that these have on evaluation practice (see Section 6.2). These include challenges stemming from lack of clear and shared definitions for key CT and PCVE concepts, security concerns, complex stakeholder landscapes, a limited understanding of initiatives to be evaluated, and the presence of overtly small and homogeneous target groups.

Secondly, the studies reviewed discussed a number of challenges associated with measuring real-world phenomena connected to CT and PCVE (see Section 6.3). In particular, authors lamented how the lack of outcome measures, outcome metrics, opportunities to measure long-term initiative effects, instruments accounting for social norms and expectations, and measures tracking exposure to interventions hampered their ability to design and conduct robust evaluations.

Thirdly, the studies reviewed highlighted a number of lessons learned as regards evaluation design (see Section 6.4). These include difficulties encountered with evaluating initiatives lacking a theory of change, challenges in isolating initiatives' effects and attributing these to a programme, constraints faced when establishing randomised experiments and quasi-experiments, and finally the impossibility of adopting a longitudinal study approach.

Fourthly, the studies reviewed discussed practical difficulties encountered when implementing evaluations (see Section 6.5). These include resource constraints, difficulty accessing information about interventions and effects, lack of access to data, and difficulties with sample size.

Lastly, the studies reviewed discussed the drawbacks and benefits of specific methods employed during an evaluation (see Section 6.6). These include lessons about the constraints of model-based investigations, the constraints of survey instruments, and the importance of triangulation and the strengths of qualitative methods.

7.2. Recommendations

The recommendations discussed throughout this report (see Sections 5.4 and 6.7) include a mix of more readily implementable measures, which could be adopted over the short and medium term, as well as more ambitious recommendations that lay out a possible vision to work towards in the field of CT and PCVE evaluation, both in the Netherlands and at a European level. The recommendations are aimed at stakeholders and organisations engaged in CT and PCVE evaluations through their design, implementation and commissioning. This section also discusses additional recommendations and suggestions stemming not only from evidence collected throughout this study, but also from a broader review and knowledge of the fields at hand and that practitioners involved with CT and PCVE policy, practice and evaluation may wish to consider.

With a view to sustaining and expanding CT and PCVE evaluation efforts, agencies funding initiatives and research in these fields should **continue to invest in the evaluation of planned and existing initiatives in these policy areas** (see Section 5.4). The timing of an evaluation in relation to the implementation of an intervention is an important question and it should be considered more carefully in CT and PCVE project and policy cycles. In many instances an evaluation can (usefully) happen before an intervention or policy is launched. For example, *ex ante evaluations* are carried out for European Union

spending programmes.[145] These studies aim to assess the expected results of a programme and provide recommendations on how its design could be improved to amplify its likely impacts.[146] Once the implementation starts (or it is well underway), the range of possible types of evaluation increases. These can include *evaluability assessments* that aim to establish the extent to which an activity or project can be evaluated in a reliable and credible fashion.[147] *Real-time evaluations* as well as *process evaluations* are other examples of studies that can provide useful insights and reflections to inform any adjustment to the intervention design or implementation.[148] When planning and conducting *impact evaluations* it is important to take into account when the results of the intervention are likely to start emerging. As such, timing an impact evaluation can be challenging since conducting it too early might fail to capture the changes generated, whereas conducting it after the implementation concludes can only inform possible future iterations of the evaluated interventions.

More broadly, **minimum quality and robustness requirements should be expected from future evaluations** (see Section 5.4), including the use of empirical data, multiple methodologies and adequate stakeholder engagement methods. As part of this push, commissioners and evaluators should **encourage greater transparency and clarity about evaluation methodologies and approaches used to allow better judgement of the evidence strength and learning** (see Section 5.4). Furthermore, in a field such as that of CT and PCVE evaluation, which is characterised by normatively and politically charged concepts, external pressures, as well as time and resource constraints, **agencies commissioning evaluations should consider adopting evaluative approaches aimed at disaggregating and, to the extent possible, isolating and addressing individually characteristics making CT and PCVE evaluation challenging** (see Section 5.4). For instance, evaluation commissioners should **consider encouraging the undertaking of rigorous real-time evaluations**, which have been used, inter alia, in the humanitarian context (see Section 5.4).

In parallel to this, known shortcomings and gaps in CT and PCVE evaluation practice should be addressed. Greater **efforts should be made to ensure that wherever possible experimental designs are employed within CT and PCVE evaluation** (see Section 6.7). In connection with this, it is further recommended that when conducting or commissioning an impact evaluation, process evaluations should be embedded in the approach to gain a more granular understanding of the initiative being implemented. Even robust impact evaluations have their limitations. One of them is the fact that they cannot explain why and how changes occur (or not). For example, many RCTs or quasi-experimental studies might produce so called null results (i.e. when a result does not support the working hypothesis and a specific intervention does not show an otherwise expected effect). There could be many explanations to such a result and this is where a process evaluation shedding more light on the implementation aspects of an initiative can be helpful. Addressing issues such as fidelity of the implementation, dosage of the

[145] Articles 30 and 140, Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the Union and repealing Council Regulation (EC, Euratom) No 1605/2002. ELI: http://data.europa.eu/eli/reg/2012/966/2016-01-01

[146] EC (2015).

[147] OECD DAC Network on Development Evaluation (2010).

[148] Ling (2012).

intervention received by participants, quality of delivery, reach and responsiveness among the participants, modalities or adaptations made to the intervention could illuminate an otherwise potentially simplistic question focusing only on whether an intervention worked or not.[149]

It is acknowledged that additional time will likely be required for CT and PCVE evaluation to grow and develop to the level of other fields. As such, while from a scientific robustness perspective we advocate for more (quasi-)experimental designs to be adopted, these might not always be feasible or desirable. To address this, a variety of other designs that could strengthen and improve the robustness of many CT and PCVE evaluations should be considered. For example, before-and-after studies (without a comparison group) or difference-in-differences approaches could be more systematically used. Also, different conceptual approaches could effectively support a chosen research design. We have emphasised the use of a theory of change in our definition of evaluation and sought to capture it in the evaluation framework. We expect that a CT and PCVE intervention should be underpinned by such a theory, which would need to be empirically tested through an evaluation. We also point to more specific approaches such as contribution analysis, which offers an alternative to causal attribution.[150]

Finally, we note and encourage the use of mixed-method designs that combine quantitative and qualitative techniques within a single study to allow for a more robust analysis, taking advantage of the strengths of each method to gain a better understanding of the research problem.[151] At the same time we caution against using an approach that simply relies on more than one method without careful consideration of the priority, application and sequence of their use.[152]

Many resources provide specific, practical and tailored guidance on planning and conducting CT and PCVE evaluations. The *IMPACT Europe Evaluation Toolkit* is one such example.[153] Its main purpose is to help professionals in designing and conducting evaluations in the PCVE field, and it aids the development of well-designed programmes that are easier to evaluate and more likely to achieve results. The toolkit is freely available online (see: http:www.impacteurope.eu/toolkit) and comprises three main sections:[154]

1. Evaluation Guide – an interactive instrument that helps practitioners in designing and conducting PCVE evaluations, offering guidance on different stages and aspects of evaluation design, undertaking and completion.
2. Interventions Database – a relational database presenting examples of current practices in the PCVE field.
3. Lessons Learned – a section providing examples of PCVE interventions that have been formally evaluated and discussing lessons learned from these evaluations.

[149] Humphrey et al. (2016).

[150] Mayne (2008).

[151] Tashakkori & Teddlie (1998).

[152] Ivankova et al. (2006).

[153] For further information, see: http://www.impacteurope.eu/toolkit

[154] Marret et al. (2017).

Another existing resource is the *RAND Program Evaluation Toolkit for Countering Violent Extremism*.[155] This product was designed to help PCVE practitioners overcome common challenges to evaluating their programmes and planning improvements to them. The toolkit guides users through the process of identifying the core components of a PCVE programme and developing a logic model to show connections between resources, activities, outcomes, evaluation measures and the need the programme addresses. The toolkit also helps practitioners design an evaluation and offers basic guidance on how to analyse and use evaluation data to inform programme improvement. Through checklists, worksheets and templates, users are taken step by step through the process of determining whether their programmes produce beneficial effects, ultimately informing the responsible allocation of scarce resources.

A third resource is *Learning and Adapting*, a handbook that provides professionals with guidance on different aspects of undertaking monitoring and evaluation in PCVE programming for the purpose of measuring effectiveness and impact. The handbook presents frameworks that can help policy-makers and practitioners understand the context in which PCVE monitoring and evaluation takes place, and the theories underpinning different monitoring and evaluation activities.[156]

We encourage practitioners and scholars active in the CT and PCVE fields to refer to these resources to access detailed, practical guidance on how specific methods, designs and approaches to evaluation may be adopted and implemented.

As discussed in Section 6.7, researchers' access to key data and stakeholders should also be improved, including by:

1. Designing and implementing approaches for providing researchers with access to sensitive information on CT and PCVE initiatives and beneficiaries.
2. Facilitating the establishment of mechanisms for data sharing among researchers
3. Collecting regular baseline measurements and identifying proxy measures and alternative indicators to mitigate the impact of data gaps.

In addition to evaluations, lessons extracted from the literature indicate that **further research is needed on the dynamics, drivers and factors governing the phenomena of radicalisation, violent extremism and terrorism** (see Section 6.7)**.** As part of these efforts, **mapping and stocktaking exercises akin to the present study should be undertaken regularly and with a comparable methodology** (see Section 5.4). Such **future exercises should also be provided with the resources and means to gain access to CT and PCVE initiatives' evaluators and beneficiaries**, with a view to providing a more in-depth and robust assessment of evaluations analysed (see Section 5.4). Furthermore, **future efforts akin to this study should also consider focusing on conducting a more in-depth comparative analysis of existing reviews** (see Section 5.4). Emphasis within future research should also be placed on **developing new evaluation designs, frameworks and approaches for conducting evaluation in the CT and PCVE policy areas** (see Section 6.7).

[155] Helmus et al. (2017).

[156] Dawson et al. (2014).

Lastly, as discussed in Section 5.4, in the context of CT and PCVE evaluation there may be merit in the adoption of approaches aimed at disentangling the complexities of this field and at pursuing an approach to knowledge development focusing on the progressive accumulation of insights and advances tackling individual challenges. To this end, **evaluators and commissioning agencies should consider embedding in their approach to CT and PCVE evaluation a 'co-production' ethos for knowledge development** (see Section 6.7). The purpose of moving towards co-production is to contribute to: (i) generating new forms and approaches for initiatives to be implemented; (ii) generating values beyond economic value; and (iii) generating new insights and research practices that are relevant to different disciplines and practices within the spectrum of CT and PCVE research, contributing to a progressive accumulation of knowledge.

7.3. Concluding observations

On the basis of the activities and results stemming from this study, a number of cross-cutting observations can be made to inform the planning and undertaking of future similar efforts.

Firstly, as discussed in Section 2.4, the methodology employed for the undertaking of this study relied primarily on the use of online databases indexing academic and other peer-reviewed publications. To avoid overlooking any potentially relevant publication, the study team endeavoured to build-in an acceptable degree of redundancy in its research approach. This was done by:

1) Searching multiple academic databases to ensure coverage of as diverse a body of journals and publications as possible within existing time and resource constraints.
2) Undertaking a range of additional research activities designed to ensure coverage of key journals and repositories that may have not have been indexed in academic databases and with a view to including in the study inventory any potentially relevant publication that would not have been triggered by the search strings employed. Activities under this strand included a manual review of several repositories and journals, as well as contacting a wide array of individual experts and practitioners with extensive knowledge of the fields at hand.

While the degree of redundancy observed in results obtained from systematic searches suggests that the search strings employed were adequate for the task at hand, it is possible that existing evaluations not indexed in databases searched may have been overlooked as a result of this approach. This may hold particularly true for evaluations published in non-peer-reviewed outlets, as well as for evaluations of projects commissioned and undertaken at a local level and thus not included in the national-level repositories reviewed by the study team.

While the study team endeavoured to engage with a broad range of experts and stakeholders in the fields of CT, PCVE and evaluation to mitigate this risk, future research efforts building on this study should consider broadening the scope of stakeholder and expert engagement activities with a view to ensuring possible additional publications not available on academic and national databases (e.g. evaluations conducted by local authorities) are captured.

Secondly, the study's analytical framework has been designed with a view to accommodating the specific purpose, needs, requirements and constraints of the present study (see Chapter 4). As a consequence, the

analytical framework does not seek to capture data pertaining to evaluation design and implementation that are normally not available through short manuscripts (such as those reviewed in this study), but that could otherwise be available from evaluations' authors and implementers. Pending availability of resources, future research efforts akin to this study should consider expanding the analytical framework proposed here and placing greater emphasis on engaging directly with evaluation authors and implementers to gain a more robust understanding of evaluation design choices and implementation activities. Further details about how the analytical framework could be expanded are discussed in Chapter 4 of this document.

Finally, placing additional emphasis on engaging with evaluators and commissioners may also be beneficial for the quality and depth of insights that studies similar to the one conducted here may generate. In particular, in light of the project's timeframe and resource constraints, the study team's engagement in this research effort was limited to a sample of domain experts and practitioners involved in CT, PCVE and evaluation activities. Given the broad and heterogeneous nature of these fields, this resulted in a limited number of practical insights pertaining to different activities and approaches, which did not allow for the formulation of practice-oriented recommendations. Future study efforts should consider placing greater emphasis on engaging in a systematic manner with evaluators and commissioning authorities to generate the evidence base required to sustain the formulation of a greater volume of practice-oriented recommendations and insights.

References

Alkin, M.C., & C.A. Christie. 2004. 'An Evaluation Theory Tree.' In *Evaluation Roots. Tracing Theorists' Views and Influences*, ed. M.C. Alkin, 12–65. Thousand Oaks, CA: SAGE Publications.

Arce M., D.G., & T. Sandler. 2005. 'Counterterrorism: A Game-Theoretic Analysis.' *Journal of Conflict Resolution* 49(2): 183–200.

Bakker, E., & R.L. de Bont. 2016. 'Belgian and Dutch Jihadist Foreign Fighters (2012–2015): Characteristics, Motivations, and Roles in the War in Syria and Iraq.' *Small Wars & Insurgencies* 27(5): 837–57.

Barkindo, A., & S. Bryans. 2016. 'De-Radicalising Prisoners in Nigeria: Developing a Basic Prison Based De-radicalisation Programme.' *Journal for Deradicalization* 7: 1–25.

Baruch, B., T. Ling, R. Warnes & J. Hofman. Forthcoming. 'Developing Knowledge on Counter-violent-extremism: A Measurement Framework for Evaluations.'

Bebarta, V.S., N. Garrett, M. Brenner, S.B. Mahon, J.K. Maddry, S. Boudreau, M. Castaneda & G.R. Boss. 2017. 'Efficacy of Intravenous Cobinamide Versus Hydroxocobalamin or Saline for Treatment of Severe Hydrogen Sulfide Toxicity in a Swine (Sus scrofa) Model.' *Academic Emergency Medicine* 24(9): 1088–98.

Bjørgo, T. & Horgan, J. (Eds.). 2008. *Leaving Terrorism Behind: Individual and Collective Disengagement*. New York: Routeledge

Bovenkerk, F. 2011. 'On Leaving Criminal Organizations.' *Crime, Law and Social Change* 55(4): 261–76.

Byman, D. 2007. 'US Counter–terrorism Options: A Taxonomy.' *Survival* 49(3): 121–50.

Cambridge University. 2017. 'The Cambridge Dictionary.' As of 2 November 2017: https://dictionary.cambridge.org/

Campbell, D., & J. Stanley. 1966. *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand McNally.

Carr, S. 2016. 'Can Co-production Really Transform UK Mental Health Services?' *Mad in America: Science, Psychiatry and Social Justice*, 29 August 2016. As of 15 June 2018: https://www.madinamerica.com/2016/08/can-co-production-really-transform-uk-mental-health-services/

Centre for Policy on Ageing. 2018. 'Policies on Ageing.' As of 15 June 2018: http://www.cpa.org.uk/cpa/policies_on_ageing.html

Chen, H., & Rossi, P. 1983. 'Evaluating with Sense: The Theory-driven Approach.' *Evaluation Review* 7(3): 283–302.

Christmann, K. 2012. *Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence*. Youth Justice Board Research Report.

Clarke, A. 1999. *Evaluation research: An Introduction to Principles, Methods and Practice*. London: SAGE Publications.

Conley-Tyler, M. 2005. 'A Fundamental Choice: Internal or External Evaluation?' *Evaluation Journal of Australasia* 4(1–2): 3–11.

Contandriopoulos, D., & A. Brousselle. 2012. 'Evaluation Models and Evaluation Use.' *Evaluation* 18(1): 61–77.

Council of the European Union. 2005. *The European Union Counter-Terrorism Strategy*. Brussels: Council of the European Union. As of 2 November 2017: https://register.consilium.europa.eu/doc/srv?l=EN&f=ST%2014469%202005%20REV%204

Cousins, J., & L. Earl. 1992. 'The Case for Participatory Evaluation.' *Educational Evaluation and Policy Analysis* 14(4): 397–418.

Davidson, E.J. 2004. *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: SAGE Publications.

Davies, M., J. Hofman & R. Warnes. 2017. *Exploring the Transferability and Applicability of Gang Evaluation Methodologies to Counter-violent Radicalization*. Santa Monica, CA: RAND Corporation. RR-2120-EC. As of 24 August 2018: https://www.rand.org/pubs/research_reports/RR2120.html

Davies, R. 2013. *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations*. Working Paper 40. London: UK Department for International Development.

Dawson, L., C. Edwards & C. Jeffray. 2014. *Learning and Adapting: The Use of Monitoring and Evaluation in Countering Violent Extremism – A Handbook for Practitioners*. London: Royal United Services Institute.

Decker, S., & D. Pyrooz. 2011. 'Gangs, Terrorism, and Radicalization.' *Journal of Strategic Security* 4(4): 151–66.

El-Said, Hamed. 2015. *New Approaches to Countering Terrorism: Designing and Evaluating Counter Radicalization and De-Radicalization Programs*. New York: Palgrave Macmillan.

European Commission. 2015. *Better Regulation Toolbox Accompanying the Better Regulation Guidelines*. SWD(2015) 111 final, Commission Staff Working Document. Luxembourg: The Publications Office of the European Union.

European Commission. 2017. *Radicalisation Awareness Network (RAN)*. As of 26 October 2017: https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network_en

European Parliament. 2015. *Resolution of 25 November 2015 on the Prevention of Radicalisation and Recruitment of European Citizens by Terrorist Organisations*. As of 2 November 2017:

http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2015-0410+0+DOC+XML+V0//EN

European Parliament. 2017. *The European Union's Policies on Counter-Terrorism, Relevance, Coherence and Effectiveness*. As of 26 October 2017: http://www.europarl.europa.eu/RegData/etudes/STUD/2017/583124/IPOL_STU(2017)583124_EN.pdf

Europol. 2017. *EU Terrorism Situation and Trend Report (TE-SAT) 2017*. Luxembourg: The Publications Office of the European Union.

Farrington, D.P., D.C. Gottfredson, L.W. Sherman & B.C. Welsh. 2002. 'The Maryland Scientific Methods Scale.' In *Evidence-Based Crime Prevention*, ed. L.W. Sherman, D.P. Farrington, B.C. Welsh & D.L. MacKenzie, 13–21. London: Routledge.

Fast, L.A., & R.C. Neufeldt. 2005. 'Envisioning Success: Building Blocks for Strategic and Comprehensive Peacebuilding Impact Evaluation'. *Journal of Peacebuilding & Development* 2(2): 24–41.

Feddes, A., & M. Gallucci. 2015. 'A Literature Review on Methodology used in Evaluating Effects of Preventive and De-radicalisation Interventions.' *Journal for De-Radicalisation* 5: 1–41.

Ferguson, N. 2016. 'Disengaging from Terrorism: A Northern Irish Experience.' *Journal for Deradicalization* 6: 1–23.

Filipe, A., A. Renedo & C. Marston. 2017. 'The Co-production of What? Knowledge, Values, and Social Relations in Health Care.' *PLoS Biology* 15(5): 1–6.

Fives, A., D.W. Russell, J. Canavan, R. Lyons, P. Easton & C. Devaney. 2013. 'The Ethics of Randomized Controlled Trials in Social Settings: Can Social Trials be Scientifically Promising and Must There Be Equipoise?' *International Journal of Research & Method in Education* 38(1): 56–71.

Frye, A.W., & P.A. Hemmer. 2012. 'Program Evaluation Models and Related Theories: AMEE Guide no. 67.' *Medical Teacher* 34(5): e288–e299.

Garrod, B., & T. Ling. 2018. *System Change Through Situated Learning. Pre-evaluation of the Health Innovation Network's Communities of Practice*. Santa Monica, CA: RAND Corporation. RR-1821-HIN. As of 20 June 2018: https://www.rand.org/pubs/research_reports/RR1821.html

Gielen, A.-J. 2017. 'Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How?' *Terrorism and Political Violence*. Published online: https://doi.org/10.1080/09546553.2017.1313736

Grünewald, F., & A. Binder. 2010. *Inter-agency Real Time Evaluation in Haiti: 3 Months After the Earthquake*. Paris: Group URD and Global Public Policy Institute.

Harris-Hogan, S., K. Barrelle & A. Zammit. 2015. 'What is Countering Violent Extremism? Exploring CVE Policy and Practice in Australia.' *Behavioural Sciences of Terrorism and Political Aggression* 8(1): 6–24.

Health Innovation Network. 2015. 'Patient Safety Collaborative – Cultivating Communities of Practice: A Call for Conveners – Frequently Asked Questions.' Healthinnovationnetwork.com. As of 12 July 2018: https://healthinnovationnetwork.com/wp-content/uploads/2017/01/Cultivating_Communities_of_Practice_-_a_Call_for_Conveners_-_FAQs_v2-1.pdf

Heath-Kelly, C. 2012. 'Counter-Terrorism and the Counterfactual: Producing the "Radicalisation" Discourse and the UK PREVENT Strategy.' *British Journal of Politics and International Relations* 15(3): 394–415.

Helmus, T.C., M. Matthews, R. Ramchand, S. Beaghley, D. Stebbins, A. Kadlec, M.A. Brown, A. Kofner & J.D. Acosta. 2017. *RAND Program Evaluation Toolkit for Countering Violent Extremism*. Santa Monica, CA: RAND Corporation. TL-243-DHS. As of 24 August 2018: https://www.rand.org/pubs/tools/TL243.html

HM Treasury. 2011. *The Magenta Book: Guidance for Evaluation*. London: HM Treasury.

Hoffman, B. 2002. 'Rethinking Terrorism and Counterterrorism Since 9/11'. *Studies in Conflict & Terrorism* 25(5): 303–16.

Humphrey, N., A. Lendrum, E. Ashworth, K. Frearson, R. Buck & K. Kerr. 2016. *Implementation and Process Evaluations (IPE) for Interventions in Education Settings: An Introductory Handbook*. Education Endowment Foundation.

Imison, C., N. Curry, H. Holder, S. Castle-Clarke, D. Nimmons, J. Appleby, R. Thorlby & S. Lombardo. 2017. *Shifting the Balance of Care: Great Expectations*. Research report. London: Nuffield Trust.

IMPACT Europe. 2017a. 'Evaluation Toolkit.' As of 2 November 2017: http://www.impact.itti.com.pl/index#/home

IMPACT Europe. 2017b. 'Lessons Learned: Monitoring and Evaluation.' As of 2 November 2017: http://www.impact.itti.com.pl/index#/inform/lessons/monitoring

Independent Evaluation Group – World Bank. 2007. *Sourcebook for Evaluating Global and Regional Partnership Programs. Indicative Principles and Standards*. Washington, D.C.: IEG-World Bank.

International Labour Organization. 2017. *ILO Policy Guidelines for Evaluation: Principles, Rationale, Planning and Managing for Evaluations*. 3rd edition. Geneva: Evaluation Office, International Labour Organization.

Ivankova, N.V., J.W. Creswell & S.L. Stick. 2006. 'Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods* 18(1): 3–20.

Kilburn, M.R., & T. Mattox. 2016. *Using Evidence to Strengthen Policies for Investing in Children*. Santa Monica, CA: RAND Corporation. TL-140-EC. As of 2 November 2017: https://www.rand.org/pubs/tools/TL140.html

Kleemans, E., C.K. Haarhuis, F. Leeuw & M. van Ooyen-Houben. 2007. 'Law Enforcement Interventions in the Netherlands: Mapping Interventions and "Browsing" for Evidence.' *Evidence & Policy* 3(4): 487–504.

Koehler, D. 2017. 'A Typology of "De-radicalisation" Programmes.' In *'De-radicalisation': Scientific Insights for Policy*, ed. L. Colaert, 63–81. Brussels: Flemish Peace Institute.

Leeuw, F.L. 2009. 'Evaluation Policy in the Netherlands.' In *Evaluation Policy and Evaluation Practice. New Directions for Evaluation*, ed. W.M.K. Trochim, M.M. Mark & L.J. Cooksy, 87–102.

Ling, T. 2012. 'Evaluating Complex and Unfolding Interventions in Real Time.' *Evaluation* 18(1): 75–87.

Lub, V. 2013. 'Polarisation, Radicalization and Social Policy: Evaluating the Theories of Change.' *Evidence and Policy* 9(2): 165–83.

Lum, C., L.W. Kennedy & A.J. Sherley. 2006. 'Are Counter-Terrorism Strategies Effective? The Results of the Campbell Systematic Review on Counter-Terrorism Evaluation Research.' *Campbell Systematic Reviews* 2: 1–49.

Madaleno, M., & S. Waights. 2014. *Guide to Scoring Methods Using the Maryland Scientific Methods Scale*. London: What Works Centre for Local Economic Growth.

Manjikian, M. 2017. 'Walking a Thin Line: The Netherland's Counterterrorism Challenge.' In *The Palgrave Handbook of Global Counterterrorism Policy*, ed. S.N. Romaniuk. London: Palgrave Macmillan.

Marret, J., J. Bellasio, H. Van den Berg, A. Van Gorp, D. van Hemert, L. Leone, R. Meijer, R. Warnes & R. Van Wonderen. 2017. *Innovative Methods and Procedures to Assess Counter-violent-radicalisation Techniques in Europe: Final Report Providing Background for Using and Further Developing the Validated Toolkit*. IMPACT Europe. As of 24 August 2018: http://impacteurope.eu/wp-content/uploads/2017/11/D6.2-Final-Synthesis-report.pdf

Marsden, S. 2017. *Reintegrating Extremists: Deradicalisation and Desistance*. London: Palgrave Macmillan.

Mastroe, C. 2016. 'Evaluating CVE: Understanding the Recent Changes to the United Kingdom's Implementation of Prevent.' *Perspectives on Terrorism* 10(2): 50–60.

Mayne, J. 2008. *Contribution Analysis: An Approach to Exploring Cause and Effect.* ILAC Methodological Brief. Institutional Learning and Change Initiative.

Murnane, R.J., & J.B. Willett. 2010. *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford: Oxford University Press.

Nardini, C. 2014. 'The Ethics of Clinical Trials'. *Ecancermedicalscience* 8: 387.

National Coordinator for Security and Counterterrorism (NCTV). 2011. *Nationale Contraterrorismestrategie 2011–2016*. As of 15 June 2018: https://www.nctv.nl/binaries/nationale-contraterrorismestrategie-2011-2015_tcm31-30099.pdf

National Coordinator for Security and Counterterrorism (NCTV). 2014. *The Netherlands Comprehensive Action Programme to Combat Jihadism. Overview of Measures and Actions*. As of 30 October 2017: https://english.nctv.nl/binaries/def-a5-nctvjihadismuk-03-lr_tcm32-83910.pdf

National Coordinator for Security and Counterterrorism (NCTV). 2016a. *National Counterterrorism Strategy for 2016–2020*. As of 30 October 2017: https://english.nctv.nl/binaries/LR_100495_rapportage_EN_V3_tcm32-251878.pdf

National Coordinator for Security and Counterterrorism (NCTV). 2016b. *Returning Jihadists May Pose Threat to the Security of the Netherlands*. As of 30 October 2017: https://english.nctv.nl/current_topics/news/2016/Returning-jihadists-may-pose-threat-to-the-security-of-the-Netherlands.aspx

National Coordinator for Security and Counterterrorism (NCTV). 2018. *Terrorist Threat Assessment for the Netherlands 47*. As of 10 July 2018: https://english.nctv.nl/binaries/Summary%20DTN%2047_tcm32-317170.pdf

Needham, C. 2008. 'Realising the Potential of Co-production: Negotiating Improvements in Public Services.' *Social Policy and Society* 7(2): 221–31.

Nelen, H., F. Leeuw & S. Bogaerts. 2010. *Antiterrorismebeleid en evaluatieonderzoek: Framework, toepassingen en voorbeelden*. Den Haag: Boom Juridische Uitgevers.

OECD. 2010. *Glossary of Key Terms in Evaluation and Results Based Management*. As of 15 June 2018: https://www.oecd.org/dac/evaluation/2754804.pdf

OECD DAC Network on Development Evaluation. 2010. *Evaluating Development Co-Operation. Summary of Key Norms and Standards*. 2nd edition. Paris: OECD Publishing.

Ostrom, E. 1996. 'Crossing the Great Divide: Coproduction, Synergy, and Development.' *World Development* 24: 1073–87.

Patton, M.Q. 2001. 'Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned.' *American Journal of Evaluation* 22: 329–36.

Patton, M.Q. 1998. 'Discovering Process Use.' *Evaluation* 4(2): 225–233

Patton, M.Q. 2008. *Utilization-Focused Evaluation*. 4th edition. Thousand Oaks, CA: SAGE Publications.

Pawson, R., & N. Tilley. 1997. *Realistic Evaluation*. London: SAGE Publications.

Picciotto, R. 1999. 'Towards an Economics of Evaluation.' *Evaluation* 5(1): 7–22.

Polastro, R., A. Nagrah, N. Steen & F. Zafar. 2011. *Inter-agency Real Time Evaluation of the Humanitarian Response to Pakistan's 2010 Flood Crisis.* Madrid: DARA.

Radicalisation Awareness Network (RAN). 2017. *Preventing Radicalisation to Terrorism and Violent Extremism Approaches and Practices*. Luxembourg: The Publications Office of the European Union.

Ramsay, G. 2015. 'Why Terrorism Can, But Should Not Be Defined.' *Critical Studies on Terrorism* 8(2): 211–28.

Reding, A., A. Van Gorp, K. Cox, A. Walczak, C. Giacomantonio & S. Hoorens. 2014. *Handling Ethical Problems in Counterterrorism*. Santa Monica, CA: RAND Corporation. RR-251-WODC. As of 24 August 2018: https://www.rand.org/pubs/research_reports/RR251.html

Resnik, D.B. 2008. 'Randomized Controlled Trials in Environmental Health Research: Ethical Issues'. *J Environ Health* 70(6): 28–30.

Ris, L., & A. Ernstofer. 2017. 'Borrowing a Wheel: Applying Existing Design, Monitoring, and Evaluation Strategies to Emerging Programming Approaches to Prevent and Counter Violent Extremism'. As of 30 October 2017: http://cdacollaborative.org/wordpress/wp-content/uploads/2017/04/Applying-Existing-DME-Strategies-to-Emerging-PCVE-Approaches.pdf

Rogers, P., A. Petrosino, T.A. Huebner & T.A. Hacsi. 2000. Program Theory Evaluation: Practice, Promise, and Problems. *New Directions for Evaluation* 87: 5–13.

Romaniuk, P., & N. Chowdhury Fink. 2012. *From Input to Impact: Evaluating Terrorism Prevention Programs*. Center on Global Counterterrorism Cooperation. As of 26 October 2017: http://globalcenter.org/wp-content/uploads/2012/10/CGCC_EvaluatingTerrorismPrevention.pdf

Romaniuk, P. 2015. *Does CVE Work? Lessons Learned from the Global Effort to Counter Violent Extremism*. Center on Global Counterterrorism Cooperation. As of 25 June 2018: http://www.globalcenter.org/wp-content/uploads/2015/09/Does-CVE-Work_2015.pdf

Schmid, A.P. 2004a. 'Terrorism – The Definitional Problem.' *Case Western Reserve Journal of International Law* 36(2): 374–419.

Schmid, A.P. 2004b. 'Frameworks for Conceptualising Terrorism.' *Terrorism and Political Violence* 16(2): 197–221.

Schmid, A.P. 2013. 'Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review.' *ICCT Research Paper* 4(2): 1–97.

Schmid, A., & A.J. Jongman. 1988. *Political Terrorism: A New Guide to Actors, Authors, Concepts, Data Bases, Theories and Literature*. Piscataway, NJ: Transaction Publishers.

Schuurman, B., & E. Bakker. 2014. 'Reintegrating Jihadist Extremists: Evaluating a Dutch Initiative, 2013–2014.' *Behavioral Sciences of Terrorism and Political Aggression* 6(2): 66–85.

Schwartz, R., & J. Mayne. 2005. 'Assuring the Quality of Evaluative Information: Theory and Practice.' *Evaluation and Program Planning* 28(1): 1–14.

Scriven, M. 1967. 'The Methodology of Evaluation.' In *Perspectives on Curriculum Evaluation*, ed. R.W. Tyler, R.M. Gagné and M. Scriven, 39–83. Chicago, IL: Rand McNally.

Scriven, M. 1991. *Evaluation Thesaurus*. Beverly Hills, CA: SAGE Publications.

Scriven, M. 1997. 'Truth and Objectivity in Evaluation.' In *Evaluation for the 21st Century: A Handbook*, ed. E. Chelimsky & W.R. Shadish, 477–500. Thousand Oaks, CA: SAGE Publications.

Scriven, M. 2011. 'Evaluating Evaluations: A Meta-evaluation Checklist.' As of 8 June 2018: http://michaelscriven.info/images/EVALUATING_EVALUATIONS_8.1.11.pdf

Scriven, M. 2015. 'Key Evaluation Checklist (KEC).' As of 25 June 2018: http://www.michaelscriven.info/images/MS_KEC_8-15-15.doc

Shadish, W.R., D.T. Campbell & T.D. Cook. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shadish, W.R., T.D. Cook & L.C. Leviton. 1991. 'Good Theory for Social Program Evaluation.' In *Foundations of Program Evaluation: Theories of Practice*, ed. W.R. Shadish, T.D. Cook & L.C. Leviton, 36–67. Newbury Park, CA: SAGE Publications.

Sherman, L.W., D. Gottfredon, D. MacKenzie, J. Eck, P. Reuter & S. Bushway. 1997. 'Preventing Crime: What Works, What Doesn't, What's Promising.' College Park University of Maryland (NCJ No. 165366). As of 8 June 2018: https://www.ncjrs.gov/pdffiles/171676.pdf

Sielbeck-Bowen, K.A., S. Brisolara, D. Seigart, C. Tischler & E. Whitmore. 2002. 'Exploring Feminist Evaluation: The Ground From Which We Rise.' *New Directions for Evaluation* 96: 3–8.

Stake, R.E. 1983. 'Program Evaluation, Particularly Responsive Evaluation.' In *Evaluation Models*, part of the Evaluation in Education and Human Services book series, Vol. 6, 287–310. Dordrecht: Springer.

Stake, R.E., & T.A. Schwandt. 2006. 'On Discerning Quality in Evaluation.' In *The Sage Handbook of Evaluation*, ed. I.F. Shaw, J.C. Greene & M.M. Mark, 404–18. London: SAGE Publications.

Stufflebeam, D.L. 2001. 'The Metaevaluation Imperative.' *American Journal of Evaluation* 22 (2): 183–209.

Stufflebeam, D.L., & A.J. Shinkfield. 2007. *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.

Streefkerk, J.W., M.P. van Esch-Bussemakers, M.A. Neerincx & R. Looije. 2008. 'Evaluating Context-Aware Mobile Interfaces for Professionals.' In *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, ed. J. Lumsden. Hershey, PA: IGI Global.

Tashakkori, A., & C.B. Teddlie. 1998. *Mixed Methodology: Combining Qualitative and Quantitative Approaches.* Applied Social Research Methods Series, Vol. 46. Thousand Oaks, CA: SAGE Publications.

Tuck, H., & T. Silverman. 2016. *The Counter-Narrative Handbook*. London: Institute for Strategic Dialogue. United Nations (UN). 2014. *Security Council Unanimously Adopts Resolution Condemning Violent Extremism, Underscoring Need to Prevent Travel, Support for Foreign Terrorist Fighters*. SC/11580, 24 September 2014. As of 15 June 2018: https://www.un.org/press/en/2014/sc11580.doc.htm

United Nations Security Council. 2014. *Resolution 2178 (2014)*. As of 30 October 2017: http://www.un.org/ga/search/view_doc.asp?symbol=S/RES/2178%20%282014%29

UK Government. 2017. *Terrorism Act 2006*. As of 30 October 2017: https://www.legislation.gov.uk/ukpga/2006/11/contents

US Department of Justice. 2017. *The USA PATRIOT Act: Preserving Life and Liberty*. As of 30 October 2017: https://www.justice.gov/archive/ll/highlights.htm

Utilization-Focused Evaluation. 2018. *Developmental Evaluation*. As of 15 June 2018: http://www.utilization-focusedevaluation.org/developmental-evaluation/

Van Dongen, T. 2010. 'Mapping Counterterrorism: A Categorisation of Policies and the Promise of Empirically Based, Systematic Comparisons.' *Critical Studies on Terrorism* 3(2): 227–41.

Van Ginkel, B., & E. Entenmann. 2016. *The Foreign Fighters Phenomenon in the European Union: Profiles, Threats & Policies*. The Hague: International Centre for Counter-Terrorism.

Van Hemert, D, H. van den Berg, T. van Vliet, M. Roelofs, M. Huis in 't Veld, J. Marret, M. Gallucci & A. Feddes. 2014. *Innovative Method and Procedure to Assess Counter-violent-radicalisation Techniques in Europe: Synthesis Report on the State-of-the-Art in Evaluating the Effectiveness of Counter-violent Extremism Initiatives*. IMPACT Europe. As of 30 October 2017: http://impacteurope.eu/wp-content/uploads/2015/02/D2.2-Synthesis-Report.pdf

Weggemans, D., E. Bakker & P. Grol. 2014. 'Who Are They and Why Do They Go? The Radicalisation and Preparatory Processes of Dutch Jihadist Foreign Fighters.' *Perspectives on Terrorism* 8(4): 100–10.

Weinberg, L., A. Pedazhur & S. Hirsch-Hoefler. 2010. 'The Challenges of Conceptualising Terrorism.' *Terrorism and Political Violence* 16(4): 777–94.

Weiss, C.H. 1995. 'Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families.' In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, ed. J. Connell, A. Kubisch, L.B. Schorr & C.H. Weiss, Vol. 1, 65–92. New York, NY: Aspen Institute.

Wenger, E., R. McDermott & W. Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Boston, MA: Harvard Business School Press.

Wensink, W., B. Warmenhoven, R. Haasnoot, R. Wesselink, B. van Ginkel, S. Wittendorp, C. Paulussen, W. Douma, B. Boutin, O. Guven & T. Rijken. 2017. *The European Union's Policies on Counter-Terrorism Relevance, Coherence and Effectiveness*. Brussels: European Union.

Williams, M.J., & S.M. Kleinman. 2014. 'A Utilization-focused Guide for Conducting Terrorism Risk Reduction Program Evaluations.' *Behavioral Sciences of Terrorism and Political Aggression* 6(2): 102–46.

Annex A. CT and PCVE evaluations inventory

ID

01. Aldrich, D.P. 2012. 'Radio as the Voice of God: Peace and Tolerance Radio Programming's Impact on Norms.' *Perspectives on Terrorism* 6(6): 34–60.

02. Aldrich, D.P. 2014. 'First Steps Towards Hearts and Minds? USAID's Countering Violent Extremism Policies in Africa.' *Terrorism and Political Violence* 26(3): 523–46.

03. Aly, A., E. Taylor & S. Karnovsky. 2014. 'Moral Disengagement and Building Resilience to Violent Extremism: An Education Intervention.' *Studies in Conflict & Terrorism* 37(4): 369–85.

04. Argomaniz, J., & A. Vidal-Diez. 2015. 'Examining Deterrence and Backlash Effects in Counter-terrorism: The Case of ETA.' *Terrorism and Political Violence* 27(1): 160–81.

05. Boyd-MacMillan, E.M. 2016. 'Increasing Cognitive Complexity and Collaboration Across Communities: Being Muslim Being Scottish.' *Journal of Strategic Security* 9(4): 79–110.

06. Broadbent, R. 2013. 'Using Grass Roots Community Programs as an Anti-extremism Strategy.' *Australian Journal of Adult Learning* 53(2): 187–210.

07. Carson, J.V. 2017. 'Assessing the Effectiveness of High-Profile Targeted Killings in the "War on Terror": A Quasi-Experiment.' *Criminology and Public Policy* 16(1): 191–220.

08. Choudhury, T., & H. Fenwick. 2011. 'The Impact of Counter-terrorism Measures on Muslim Communities.' *International Review of Law, Computers & Technology* 25(3): 151–81.

09. Demant, F., & B. de Graaf. 2010. 'How to Counter Radical Narratives: Dutch Deradicalization Policy in the Case of Moluccan and Islamic Radicals.' *Studies in Conflict & Terrorism* 33(5): 408–28.

10. Dunn, K.M., R. Atie, M. Kennedy, J.A. Ali, J. O'Reilly & L. Rogerson. 2015. 'Can You Use Community Policing for Counter Terrorism? Evidence from NSW, Australia.' *Police Practice and Research* 17(3): 196–211.

11. Feddes, A.R., L. Mann & B. Doosje. 2015. 'Increasing Self-esteem and Empathy to Prevent Violent Radicalization: A Longitudinal Quantitative Evaluation of a Resilience Training Focused on Adolescents with a Dual Identity.' *Journal of Applied Social Psychology* 45(7): 400–11.

12. Finn, M., B. Momani, M. Opatowski & M. Opondo. 2016. 'Youth Evaluations of CVE/PVE Programming In Kenya in Context.' *Journal of Deradicalization* 7: 164–224.

13. Gestel, B. van & C.J. de Poot. 2014. *Evaluatie Wet opsporing terroristische misdrijven*. Den Haag: WODC. As of 14 June 2018: https://www.wodc.nl/onderzoeksdatabase/evaluatie-antiterrorismewetgeving-in-werking.aspx

14. Haider, R., & E. Martinez. 2014. 'Engaging Transit Riders in Public Awareness Programs.' *Journal of Public Transportation* 17(1): 21–41.

15. Hirschfield, A., K. Christmann, A. Wilcox, M. Rogerson & K. Sharratt. 2012. *Process Evaluation of Preventing Violent Extremism Programmes for Young People*. London: Youth Justice Board.

16. Johns, A., M. Grossman & K. McDonald. 2014. '"More Than a Game": The Impact of Sport-based Youth Mentoring Schemes on Developing Resilience Toward Violent Extremism.' *Social Inclusion* 2(2): 57–70.

17. Jordan, J. 2014. 'The Effectiveness of the Drone Campaign Against Al Qaeda Central: A Case Study.' *Journal of Strategic Studies* 37(1): 4–29.

18. Khalil, J., & M. Zeuthen. 2014a. 'A Case Study of Counter Violent Extremism (CVE) Programming: Lessons from OTI's Kenya transition initiative.' *Stability: International Journal of Security & Development* 3(1): 1–12.

19. Khalil, J., & M. Zeuthen. 2014b. Qualitative Study on Countering Violent Extremism (CVE) Programming Under the Kenya Transition Initiative (KTI). Washington D.C.: USAID.

20. Krafchik, M., & Y. Ryszkowska. 2011. *Evaluation of Young and Safe Project: London Borough of Lambeth*. Inspira Consulting Evaluation Report.

21. Kundnani, A. 2009. *Spooked! How Not to Prevent Violent Extremism*. London: Institute of Race Relations.

22. Kurtz, J., R. Wolfe & B. Tesfaye. 2016. 'Does Youth Employment Build Stability? Evidence From an Impact Evaluation of Vocational Training in Afghanistan.' In *Expanding Research on Countering Violent Extremism*, ed. S. Zeiger, 105–15. Abu Dhabi: Hedayah.

23. Lakhani, S. 2012. 'Preventing Violent Extremism: Perceptions of Policy from Grassroots and Communities.' *Howard Journal of Criminal Justice* 51(2): 190–206.

24. Lamb, J.B. 2013. 'Preventing Violent Extremism: A Policing Case Study of the West Midlands.' *Policing* 7(1): 88–95.

25. Liht, J., & S. Savage. 2013. 'Preventing Violent Extremism Through Value Complexity: Being Muslim Being British.' *Journal of Strategic Studies* 6(4): 44–66.

26. McNeal, G.S. 2014. 'Targeted Killing and Accountability.' *Georgetown Law Journal* 102(3): 681–794.

27. Minor. 2016. *MAXIME Berlin - Interkulturelles und interreligiöses Präventionsprojekt zur Toleranzund Demokratieentwicklung Evaluation 2014 – 2016*. Berlin: Violence Prevention Network e.V.

28. Mitts, T. 2017. *Do Community Engagement Efforts Reduce Extremist Rhetoric on Social Media?* As of 24 August 2018: http://dx.doi.org/10.2139/ssrn.2940290

29. Moffett, K., & T. Sgro. 2016. 'School-based CVE strategies.' *ANNALS of the American Academy of Political and Social Science* 668(1): 145–64.

30. Möller, K., & F. Neuscheler. 2016. *Bericht zur Evaluation des Programms„ ‚KOMMENT' – MentorInnen-Qualifizierung zum Umgang mit vorurteilsmotivierten und fremdenfeindlichen Müttern und Vätern in Sachsen*. Berlin: Violence Prevention Network e.V.

31. Möller, K., & F. Neuscheler. 2018. *Bericht über zentrale Ergebnisse der Evaluation der Beratungsstelle Hessen - Religiöse Toleranz statt Extremismus (Managementfassung)*. Berlin: Violence Prevention Network e.V.

32. Noordegraaf, M., S. Douglas, A. Bos & W. Klem. 2016. *Gericht, gedragen en geborgd interventievermogen? Evaluatie van de nationale contraterrorisme-strategie 2011–2015*. Den Haag: WODC. As of 14 June 2018: https://www.wodc.nl/onderzoeksdatabase/2608-evaluatie-nationale-contraterrorisme-strategie-2011-2015.aspx

33. Noordegraaf, M., S. Douglas, A. Bos & W. Klem. 2017. 'How to Evaluate the Governance of Transboundary Problems? Assessing a National Counterterrorism Strategy.' *Evaluation* 23(4): 389–406.

34. O'Toole, T., D. Nilsson & T. Modood. 2012. 'Balancing Tolerance, Security and Muslim Engagement in the United Kingdom: The Impact of the "Prevent" Agenda.' *Critical Studies on Terrorism* 5(3): 373–89.

35. Octavia, L., & E. Wahyumi. 2015. *Countering & Preventing Radicalization in Indonesian Pesantren: Final Evaluation*. Washington, D.C.: Search for Common Ground.

36. Palloks, K., & A. Steil. 2015. *Bericht der Evaluation/wissenschaftlichen Begleitung des Modell-Projekts „Verantwortung über-nehmen – Eltern stärken. Arbeit mit rechtsextrem orientierten Eltern" (Rexel) von Violence Preven-tion Network, gefördert durch das BMFSFJ und die Bundeszentrale für politische Bildung*. Berlin: Violence Prevention Network e.V.

37. Roodnat, J., & Q. Eijkman. 2016. *Waak voor het stempel terrorist*. Utrecht: Hogeschool Utrecht and Senze.

38. Schuurman, B., & E. Bakker. 2016. 'Reintegrating Jihadist Extremists: Evaluating a Dutch Initiative, 2013–2014.' *Behavioral Sciences of Terrorism and Political Aggression* 8(1): 66–85.

39. Sheikh, S., S. Sarwar & E. King. 2012. *Evaluation of the Muslim Council of Wales' Prevent Work*. Merthy Tydfil: Office for Public Management.

40. Veldhuis, T.M., E.H. Gordijn, S.M. Lindenberg & R. Veenstra. 2011. *Terrorists in Prison: Evaluation of the Dutch Terrorism Wing*. Groningen: University of Groningen.

41. Veldhuis, T. 2015. *Captivated by Fear: An Evaluation of Terrorism Detention Policy*. Groningen: University of Groningen.

42. Vermeulen, F. 2014. 'Suspect Communities – Targeting Violent Extremism at the Local Level: Policies of Engagement in Amsterdam, Berlin, and London.' *Terrorism and Political Violence* 26(2): 286–306.

43. Webber, D., M. Chernikova, A.W. Kruglanski, M.J. Gelfand, M. Hettiarachchi, R. Gunaratna, M. Lafreniere & J.J. Belanger. 2017. 'Deradicalizing Detained Terrorists.' *Political Psychology* 39(3): 539–56.

44. Webster, S., J. Kerr & C. Tompkins. 2017. *A Process Evaluation of the Structured Risk Guidance for Extremist Offenders*. London: HM Prison & Probation Service.

45. Williams, M.J., J.G. Horgan & W.P. Evans. 2015. 'The Critical Role of Friends in Networks for Countering Violent Extremism: Toward a Theory of Vicarious Help-seeking.' *Behavioral Sciences of Terrorism and Political Aggression* 8(1): 45–65.

46. Williams, M.J., J.G. Horgan & W.P. Evans. 2016. *Evaluation of a Multi-Faceted, U.S. Community-Based, Muslim-Led CVE Program*. Washington D.C.: Department of Justice.

47. Wilner, A., & B. Rigato. 2017. 'The 60 Days of PVE Campaign: Lessons on Organizing An Online, Peer-To-Peer, Counter-Radicalization Program.' *Journal of Deradicalization* 12: 227–68.

48. Winter, H., N. Struiksma, N. Woestenburg & R. Bottema. 2015. *Inventarisatie en analyse van de CT-strategie 2011–2015.* Den Haag: WODC. As of 14 June 2018: https://www.wodc.nl/onderzoeksdatabase/2346-voorbereiding-evaluatie-contraterrorismestrategie-2011-2015.aspx

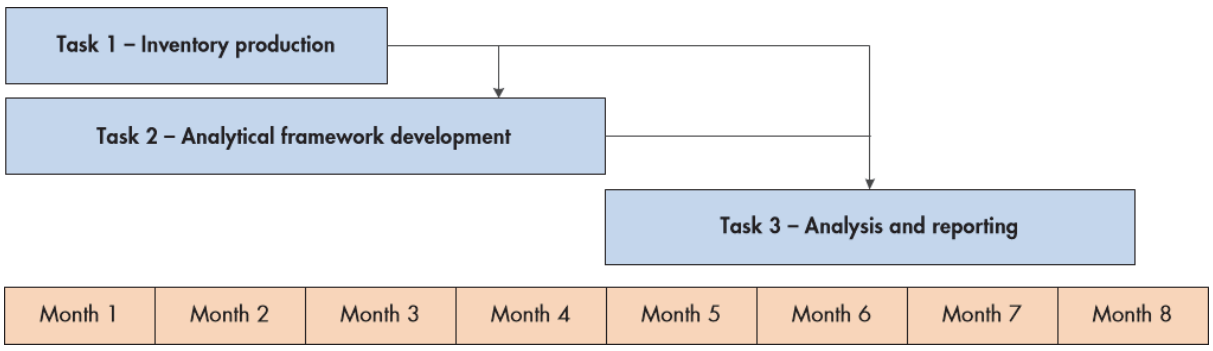
Annex B. Methodology

This study's approach revolved around three interconnected research Tasks:

- **Task 1 – Production of the CT and PCVE evaluations inventory.** This entailed the development of an inventory of evaluations of CT and PCVE strategies, policies and interventions conducted in the Netherlands and abroad since 1 January 2013 (See Annex A).
- **Task 2 – Development of an analytical framework.** This entailed the development of an analytical framework to be later employed for assessing evaluations collected in the study's evaluations inventory.
- **Task 3 – Analysis of evaluations and reporting.** This entailed the analysis of the CT and PCVE evaluations inventory produced in Task 1 through the lenses of the analytical framework developed in Task 2.

This annex provides a detailed overview of the methodology and approach employed by the research team to achieve the objectives of the three project Tasks. Figure B.1 offers a schematic overview of the project's sequence, duration and dependencies.

Figure B.1. Overview of study approach



B.1. Task 1 methodology and approach

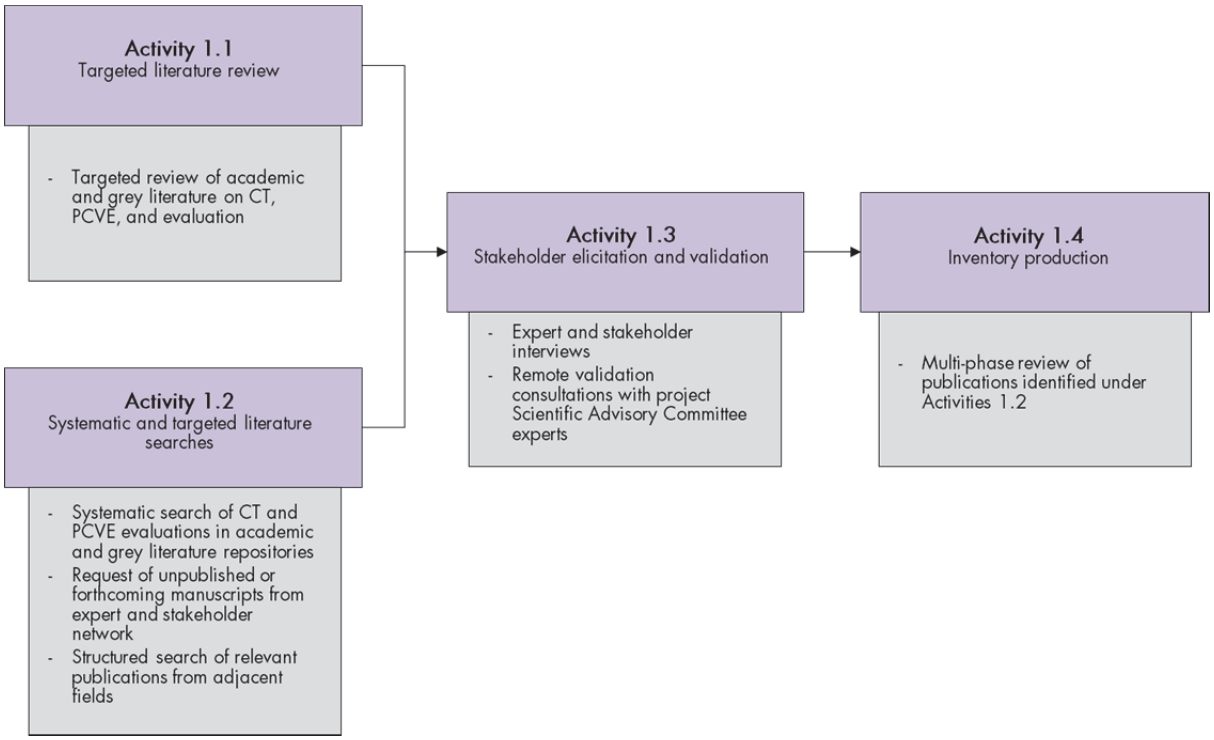
The purpose of Task 1 was to produce an inventory of evaluations of CT and PCVE strategies, policies and interventions conducted in the Netherlands and abroad since 1 January 2013. To achieve this, the study team undertook four Activities:

- Activity 1.1: Targeted literature review.
- Activity 1.2: Systematic and targeted literature searches.
- Activity 1.3: Stakeholder elicitation and validation activities.

- Activity 1.4: Inventory production.

Figure B.2 provides a schematic overview of Task 1's approach and of the research activities underpinning it. The following sections describe each activity in more detail.

Figure B.2. Overview of Task 1 methodology and approach



B.1.1. Activity 1.1 – Targeted literature review

In Activity 1.1, the study team conducted a non-systematic review of academic and grey literature on CT, PCVE and evaluation. The purpose of this activity was twofold:

1. To refine and consolidate the study's inclusion criteria to be employed for determining whether a manuscript or publication should be included in the study inventory.
2. To produce operational definitions for key concepts and underpinning phenomena, with a view to guiding the review process and the determination of whether sources are within or outside the scope of the study inventory (Activity 1.4).

Literature reviewed was identified through a non-systematic snowballing technique (i.e. a technique whereby, based on an initial set of key sources, additional studies and resources are identified through their citations) and was limited to sources available in the English language.

B.1.2. Activity 1.2 – Systematic and targeted literature searches

In Activity 1.2, the study team conducted systematic and targeted literature searches through academic and grey literature repositories, as well as through professional networks. The purpose of this activity was twofold:

1. To identify publications of evaluations of CT and PCVE strategies, policies and interventions in English, Dutch, French and German published since 1 January 2013.
2. To identify publications from fields adjacent to CT and PCVE that may present transferable lessons learned about evaluation approaches.

Systematic searches

The study team conducted systematic searches for CT and PCVE evaluation publications through online academic journal databases. Searches focused on publications released since 1 January 2013 and written in English, Dutch, German and French. In addition to Dutch and English (as the lingua franca for scholarly communication), German and French were selected on the basis of the following criteria:

1. Feasibility and availability of language skills within the research team.
2. The languages' expected body of evaluation literature.
3. The anticipated number of CT and PCVE interventions available in the countries where these languages are used.

Table B.1 provides an overview of the academic databases searched. The search engine settings on each database were adjusted so as to investigate only the title, abstract and keywords of publications in the databases, rather than their full texts. This was done because previous experience of study team members in a similar undertaking indicated that searching through online academic databases with the 'Full text search' option enabled would result in a very large volume of results, at times potentially unmanageable, and that many such results would have limited to no relevance to the scope of this study.[157]

[157] For further details, see: Van Hemert et al. (2014); Feddes & Gallucci (2016).

Table B.1. Academic databases searched

Database	Link
Social Science Research Network (SSRN)	https://www.ssrn.com/en/
JSTOR	https://www.jstor.org/
Scopus	https://www.elsevier.com/solutions/scopus
EPPI-Centre	https://eppi.ioe.ac.uk/
Web of Science	http://wokinfo.com/
PubMed	https://www.ncbi.nlm.nih.gov/pubmed/
Campbell Collaboration	https://www.campbellcollaboration.org/
Narcis (Dutch only)	https://www.narcis.nl/
Picarta (Dutch only)	http://umlib.nl/picarta_go
WorldCat (Dutch only)	https://www.worldcat.org/

It should be noted that in the case of the EPPI-Centre database, search parameters were set to cover all of the active databases maintained on this website, namely: the Database of Promoting Health Effectiveness Reviews, the Trials Register of Promoting Health Interventions, and Bibliomap.

Table B.2 gives an overview of the search strings deployed for the majority of academic databases, including JSTOR, Scopus, EPPI-Centre, Web of Science and PubMed. Keywords are clustered in three thematic areas:

1. Keywords referring to terrorism, violent extremism and associated phenomena.
2. Keywords used to describe strategies, policies and other types of interventions.
3. Keywords referring to the field of evaluation.

Several of the keywords used across different search strings are preceded by or have been truncated with an asterisk (i.e. a wildcard) to ensure that database search engines take into account through a single search string:

- Words with multiple spellings (e.g. radicalisation and radicalization).
- Words in singular and plural forms (e.g. policy and policies).
- Words that may have a prefix or suffix attached (e.g. counter-terrorist and terrorist-related).
- Words with common roots (e.g. evaluation; evaluate, evaluating, evaluative, etc.).

It should be noted that each database employs a different syntax and set of rules for how search strings, wildcards and Boolean operators should be used to compile and run searches. For all instances where the details of these divergences are not relevant from a substantive research perspective, these are not discussed in the present document.

The English search terms were then translated into Dutch, French and German. It should be noted that the sets of search strings in different languages are not literal translations, and in some cases additional search terms were included or different terms were used. This was done to ensure that the search strings in Dutch, French and German would yield the most relevant results in the context of the study's objectives. For example, the term 'foreign fighter' literally translates to '*buitenlandse strijder*' in Dutch. But in the Dutch public discourse, as well as the literature and media, terms such as '*syriëganger*', '*syriëstrijder*' or '*jihadganger*' tend to be more common. Similarly, terms in the policy or evaluation literature do not necessarily translate literally. All search terms were reviewed and edited with these considerations in mind.

Table B.2. Search strings employed on the JSTOR, Scopus, EPPI-Centre, Web of Science and PubMed databases

Language	Search strings
English	(*terroris* OR *radicali* OR extremis* OR "foreign fight*") AND (intervention* OR policy OR policies OR program* OR strategy OR strategies OR initiative*) AND (evaluat* OR assess* OR effect*OR *success*)
Dutch	(*terroris* OR *terreur* OR *radicalise* OR *extremis* OR syriëganger* OR syriëstrijder* OR "buitenlandse strijder*" OR jihadganger*) AND (*Interventie* OR *beleid* OR *programma OR *strategie* OR *maatregel* OR *initiatief) AND (*evalu* OR *beoordel* OR *effect* OR *succes* OR assess* OR *meting)
French	(*terroris* OR *radicalis* OR extremis* OR (combattant* AND (étranger* OR terroriste* OR français* OR djihadiste*)) AND (intervention* OR politique* OR programme* OR stratégie* OR initiative*) AND (évalu* OR effet* OR effect* OR succès OR réussi* OR résultat*)
German	(*Terroris* OR *Radikali* OR *Extremis* OR (ausländisch* AND Kämpfer) OR (terroristisch* AND Kämpfer) OR IS-Rückkehrer) AND (Maßnahme* OR Strategie* OR*programm* OR *Taktik*) AND (Evaluation* OR *evaluier* OR effekt* OR Wirkung OR Wirksamkeit OR erfolg* OR Ergebnis* OR Bewertung)

In the case of the Social Science Research Network (SSRN) database, it was not possible to use truncated words or Boolean operators. Therefore, an ad hoc set of search strings was developed for this database, as presented in Table B.3.

Table B.3. Search strings employed on the SSRN academic database

Language	Search strings
English	(Evaluation OR Effectiveness) AND (terrorism OR terrorist OR radicalisation OR radicalization OR "foreign fighter" OR extremism)
Dutch	(evaluatie OR effectiviteit) AND (terrorisme OR terreur OR radicalisering OR extremisme)
French	(évaluation OR efficacité OR) AND (Terrorisme OR terroriste OR radicalisation OR extremism OR "combattant étranger" OR "combattante étrangère" OR "combattant terroriste" OR "combattante terroriste" OR "combattant français" OR "combattante française" OR "combattant djihadiste" OR "combattante djihadiste")
German	(Evaluation OR Bewertung OR Evaluierung OR Wirkung OR Wirksamkeit) AND (Terrorismus OR Terrorist OR Radikalisierung OR "ausländischer Kämpfer" OR "terroristischer Kämpfer" OR Extremismus)

Similarly, as the Campbell Library only includes systematic reviews assessing interventions or policies in social sciences, the search strategy was adapted to ensure that any relevant publication would be found. To this end, the study team ran searches on this database with only individual terms contained in the cluster for keywords referring to terrorism, violent extremism and associated phenomena.

Lastly, due to the limited number of results available in Dutch on the academic databases identified above, three additional sources providing a better coverage of Dutch language literature were identified and searched: Narcis, Picarta, and WorldCat.[158] Search strings employed to investigate these databases are presented in Table B.4.

Table B.4. Search strings employed on Dutch-focused academic databases

Database	Search strings
Narcis	(terroris* OR terreur* OR radicalise* OR extremis* OR syriëganger* OR syriëstrijder* OR jihadganger*) AND (Interventie* OR beleid* OR programma OR strategie* OR maatregel* OR initiatief) AND (evalu* OR beoordel* OR effect* OR succes* OR assess* OR meting)
Picarta	Same as Table B.2
WorldCat	Searches conducted employing search strings as in Table B.2 yielded no results. An additional set of searches was run employing the following search string to broaden the scope of results: (terroris* AND evalu*)

Targeted review of academic journals and grey literature repositories

In addition to searches on generalist academic databases, the study team conducted a review of publications released since 1 January 2013 through a selected number of (i) academic journals relevant to

[158] Please refer to Table B.1 for the databases' addresses.

the scope of this study; and (ii) national and international repositories of grey literature. This was done to ensure that:

1. No relevant publication released under a specialist journal was overlooked because it did not meet the search criteria used for database searches.
2. Relevant publications not indexed on academic databases but available through national and international institutions active in the fields of CT and PCVE were included in the study inventory.

Tables B.5 and B.6 list the academic journals and grey literature repositories reviewed.

Table B.5. Overview of academic journals reviewed

Database	Link
Terrorism and Political Violence	http://www.tandfonline.com/loi/ftpv20
Studies in Conflict and Terrorism	http://www.tandfonline.com/loi/uter20
Journal of Terrorism Research	https://jtr.st-andrews.ac.uk/
CTC Sentinel	https://ctc.usma.edu/publications/sentinel
Perspectives on Terrorism	http://www.terrorismanalysts.com/pt/index.php/pot
Journal of Deradicalization	http://journals.sfu.ca/jd/index.php/jd
Behavioural Sciences of Terrorism and Political Aggression	http://www.tandfonline.com/loi/rirt20
Critical Studies on Terrorism	http://www.tandfonline.com/loi/rter20
Dynamics of Asymmetric Conflict	http://www.tandfonline.com/toc/rdac20/current
Survival	http://www.tandfonline.com/loi/tsur20
International Security	http://www.mitpressjournals.org/forthcoming/isec
American Journal of Evaluation	http://journals.sagepub.com/home/aje
Evaluation	http://journals.sagepub.com/home/evi
Evaluation Review	http://journals.sagepub.com/home/erx
Evidence and Policy	http://www.ingentaconnect.com/content/tpp/ep

Table B.6. Overview of grey literature repositories reviewed

Institution	Link
WODC Publications Database	https://english.wodc.nl/publications/
UK Home Office Research Database	https://www.gov.uk/government/publications
French Ministry of Interior Publications Database	https://www.interieur.gouv.fr/Publications/Rapports-de-l-IGA
Center for Evidence Based Crime Policy	http://cebcp.org/
German National Center for Crime Prevention	https://www.nzkrim.de/english/
Organization for Security and Co-operation in Europe	http://www.osce.org/resources/publications
Global Counter Terrorism Forum Violent Extremism (Hedayah)	http://www.hedayahcenter.org/publications/89/report
Global Counter Terrorism Forum (GCTF) Tools	https://www.thegctf.org/Tools/Overview
UK College of Policing - What Works Crime Reduction	http://whatworks.college.police.uk/Pages/default.aspx
US National Criminal Justice Reference Service	https://www.ncjrs.gov/
Alliance for Peacebuilding – Monitoring and Evaluation of CVE	http://www.allianceforpeacebuilding.org/2015/12/monitoring-evaluation-of-cve/
Design Monitoring and Evaluation for Peacebuilding	http://www.dmeforpeace.org/learn/resources/

For all journals and most institutional repositories the study team endeavoured to review all publications released after 1 January 2013. In a limited number of instances, due to the lack of a complete chronological index of publications, or to the unmanageable number of publications hosted, a different review approach was adopted:

- The UK Home Office Research Database hosts 80,112 publications released since 1 January 2013. To narrow down the number of publications to be assessed, only those classified as 'Research and analysis' containing the keyword 'Evaluation' were reviewed.
- The US National Criminal Justice Reference Service indexes 2,044 publications within the timeframe relevant to the project. As for the previous repository, only those publications containing the word 'Evaluation' were reviewed.
- The French Ministry of Interior Publications Database does not allow searching of publications. In light of this, all publications tagged under the following categories were reviewed: security and prevention of delinquency, justice, civil liberties, urbanism, immigration, management of the territory, territorial communities, European funds, transport safety, civil safety and risk prevention.
- The Design Monitoring and Evaluation for Peacebuilding repository hosts 1,071 potentially relevant resources and publications. Only resources tagged as 'Evaluation reports' were reviewed.

Solicitation of unpublished or draft manuscripts from external experts and practitioners

The study team sent direct requests via email for unpublished or forthcoming manuscripts and evaluations to a wide network of CT and PCVE policy-makers, practitioners, experts and evaluators. Each expert identified by the study team was initially contacted with a request and, in case of no response, was contacted a second and final time after two weeks. A total of 52 experts from 21 countries from across North America, Europe, Asia and Oceania were contacted; overall, 23 responded to our solicitations, providing the details of 30 manuscripts.

Structured search of relevant publications from adjacent fields

The study team conducted targeted searches to identify publications discussing evaluation approaches and lessons learned from fields that can be considered adjacent to CT and PCVE from the perspective of evaluation challenges. The fields selected for review were those of criminology, gangs' desistance, peacebuilding and cult exit. Previous research by RAND Europe indicates that evaluations from established academic fields can be used to inform the development of evaluation practice in relatively novel research areas.[159]

The study team conducted searches in English through the academic and grey literature database Google Scholar,[160] reviewing the first 20 results for each search string run. The purpose was not to gather articles and manuscripts discussing individual evaluations conducted in the fields selected; rather, we collected only studies discussing collections or summaries of lessons learned and guidance about evaluations in the fields of focus. Table B.7 provides an overview of the search strings employed.

Table B.7. Overview of search strings for identifying adjacent fields publications

Search string
"cult exit" AND (intervention* OR program* OR handbook OR guidelines) AND (evaluation OR assessment)
peacebuilding AND (intervention* OR program* OR handbook OR guidelines) AND (evaluation OR assessment)
Gangs AND (intervention* OR program* OR handbook OR guidelines) AND (evaluation OR assessment)
Criminology AND (intervention* OR program* OR handbook OR guidelines) AND (evaluation OR assessment)

B.1.3. Activity 1.3 – Stakeholder elicitation and validation activities

The study team conducted a series of stakeholder and expert engagement activities. Different elicitation techniques were used to engage stakeholders and experts, pursuing a multiplicity of purposes relevant to the scope of Task 1, including:

- Acquire information on how CT and PCVE can be conceptualised and what are the differences / overlap / tension lines between activities in these two fields.

[159] Davies et al. (2017).

[160] For further information, see: https://scholar.google.co.uk/

- Acquire information on challenges, barriers and enablers characterising CT and PCVE evaluations.
- Acquire information on any manuscript or publication that should be included in the study inventory.

It should be noted that, to minimise the risk of reporting fatigue from stakeholder and experts, the study team combined stakeholder engagement activities conducted for Task 1 and Task 2 into a single set of engagements. This is reflected in the content of the elicitation and validation tools and materials employed by the study team and presented in Annex C. The following sections discuss in greater detail the elicitation and validation activities conducted within the scope of Task 1.

Experts and stakeholders interviews

One technique employed to engage experts and stakeholders was that of semi-structured interviews. The primary purpose of Task 1 interviews was that of:

- Understanding how CT and PCVE can be conceptualised and what are the differences / overlap / tension lines between activities in these two fields.
- Producing an initial mapping of challenges, barriers and enablers characterising CT and PCVE evaluations, capturing any lessons learned from interviewees as regards these.
- Gathering information on any public or non-public manuscript or publication that should be included in the study inventory.

Semi-structured interviews were selected due to their inherent flexibility, which ensures that key focus areas and themes are discussed with all interviewees, whilst allowing interviewers to adapt the questionnaire deployed to the knowledge and expertise of the interviewee engaged. This flexibility has implications for the comparability of data recorded across different interviews. However, it also provides opportunities to discuss and validate with new interviewees emerging findings and results identified through previous consultations and research activities. Semi-structured interviews also leave scope for interviewees to add other information that had not been covered in the interview protocol. Annex C presents the interview protocol that was employed during Task 1 interviews.

Interviews were conducted with seven academics, experts and practitioners from organisations and institutions active in CT and PCVE in the Netherlands and abroad, including Europe, North America and developing countries in Asia and the Middle East. When selecting interviewees, the study team endeavoured to engage experts active in different areas of CT, PCVE and evaluation from both the Netherlands and abroad.

Remote consultations with the project Scientific Advisory Committee

In addition to semi-structured interviews, remote consultations with the project SAC were held at key junctures in Task 1 activities with a view discussing emerging results and validating upcoming research activities. As regards Task 1, the purpose of this was to:

1. Validate search strings designed for the undertaking Activity 1.2.
2. Validate the list of proposed interviewees to be engaged under Activity 1.3.
3. Validate the study's inventory inclusion criteria for CT and PCVE evaluation literature.

During the undertaking of Task 1, two written project updates were issued to the study SAC, the first on 7 December 2017 and the second on 2 February 2018. On both occasions, SAC members were asked to provide their written feedback and comments within 10 days of receipt of the project update.

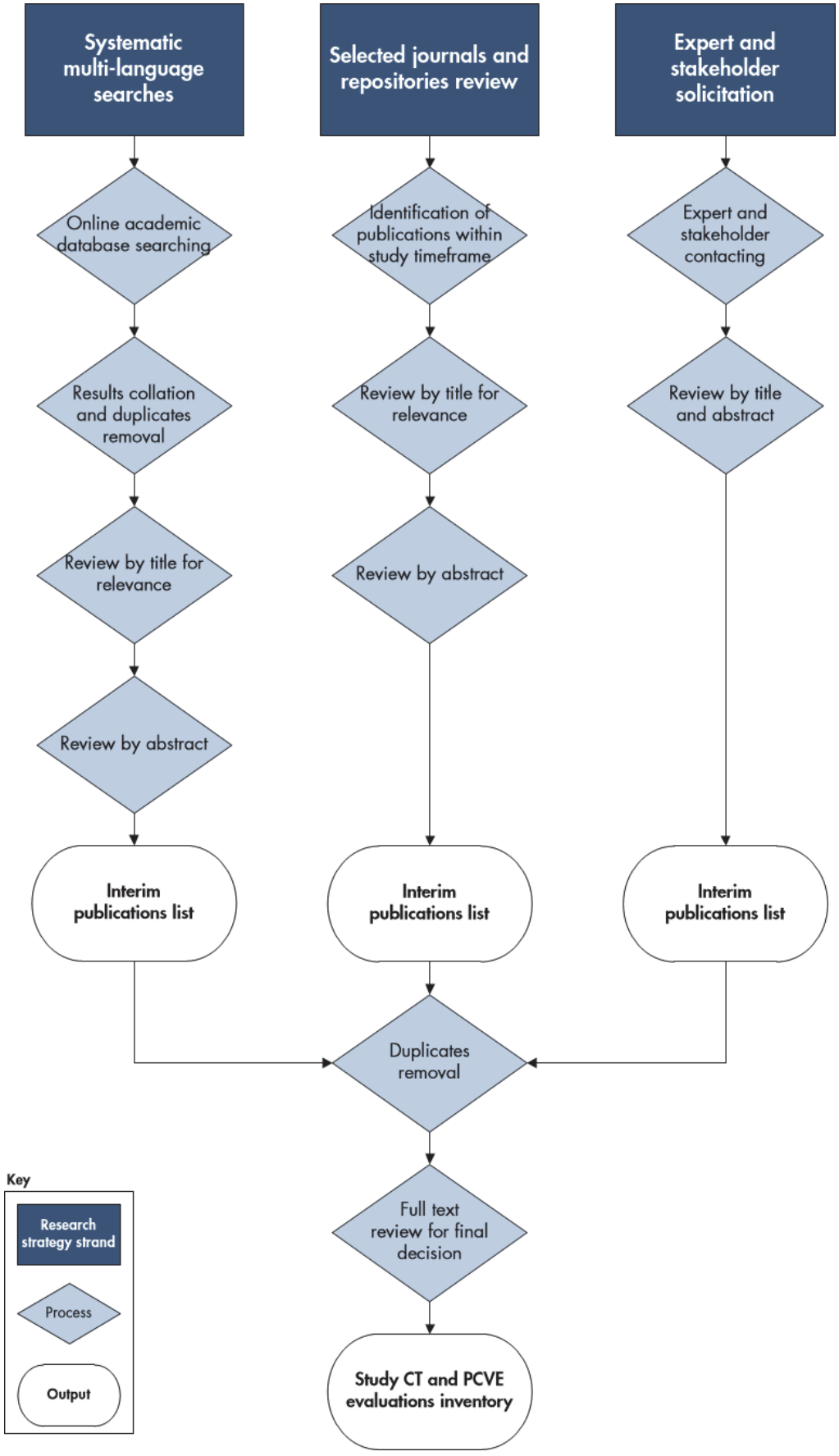
B.1.4. Activity 1.4 – Inventory production

Following the systematic and targeted searches for CT and PCVE evaluation literature and the solicitation of manuscripts from experts conducted in Activity 1.1, the study team reviewed the sources identified according to a multi-step process, with a view to finalising the study CT and PCVE evaluation inventory.

Figure B.3 gives a schematic overview of the review process that led to the finalisation of the study's CT and PCVE inventory. The figure highlights how publications identified through the three different research strands pursued under Activity 1.1 were reviewed and collated to produce the final inventory.

The following sections discuss the undertaking of the review and inventory production process, including intermediate results from each of the inventory research and production phases.

Figure B.3. Inventory production overview



Strand 1 – Reviewing results from systematic multi-language searches

Step 0 – Collating results and removing duplicates

The bibliographic details of all publications identified through systematic searches were saved into a database using the reference manager software EndNote.[161] The database collated all results obtained from the different academic databases consulted, clustering them according to the four search languages that were employed. The study team then used EndNote's automated functions to remove duplicate results encountered across all language clusters.

Table B.8 provides an overview of results obtained on different academic databases by language of search and before the removal of duplicates.

Table B.8. Overview of results by academic database and search language employed

Database	English	Dutch	French	German
JSTOR	84	0	7	0
Scopus	2,214	0	9	3
Web of Science	1,479	0	5	1
PubMed	453	0	1	0
EPPI-Centre	1	0	0	0
SSRN	174	0	0	23
Campbell Collaboration	1	0	0	0
Narcis	-	14	-	-
Picarta	-	18	-	-
WorldCat	-	8	-	-
TOTAL	**4,406**	**40**	**22**	**27**

Step 1 – Review by title for relevance

In order to assess the relevance of publications identified to the scope of the study, all publications were first screened by reviewing their title. This task was conducted by three researchers who screened different subsets of the publications list. This led to a first filtering of non-relevant publications that had been captured through searches. For example, a publication by Bebarta et al. (2017) on the efficacy of different treatments for severe hydrogen sulfide toxicity in a swine model was removed; it had originally been captured as its abstract contains keywords employed in the study search strings (i.e. terrorism, assessed, successful).

[161] For further details on this software, see: http://endnote.com/

In all instances, an inclusive approach was taken throughout the review process, opting to leave in any potentially relevant source unless it could be determined beyond doubt to be outside of the scope of the study. For example, during the first review phase, the study team left in all sources with a generic title (e.g. 'Iraq', 'The Tip of the Iceberg') or that had been captured by EndNote with an incomplete title record.

Step 2 – Review by abstract for relevance

A second round of screening of results obtained through systematic searches was conducted by assessing the relevance of each publication's abstract according to the inventory inclusion criteria defined under Activity 1.1. This review was conducted by three researchers who screened different subsets of the publications list.

In particular, study team members reviewed sources in order to identify those that appeared to comply with the following conditions:

1. The manuscript or publication refers to a study in which a CT or PCVE strategy, policy, programme or intervention is evaluated.
2. The CT or PCVE strategy, policy, programme or intervention being evaluated is (i) underpinned by an explicit or implicit theory of change; and (ii) evaluated using a qualitative or a quantitative evaluation approach that entails the collection and analysis of primary data to investigate a set of clearly defined evaluation questions.
3. The manuscript was written or published after 1 January 2013.

This second review round led to a further reduction in the number of publications included in the study repository. As during Step 1, only those manuscripts whose abstracts placed them beyond doubt outside the scope of the study were removed from the interim inventory. In instances where abstracts were deemed to fall in a grey area, they were then reviewed by a second member of the study team. If it was not possible to determine with full certainty the relevance or lack thereof of a publication, it was left in the interim inventory.

Table B.9 provides an overview of results obtained from academic databases as a whole and by language of search during the intermediate review stages described above. The second column shows the total number of potentially relevant sources identified after having merged all sources identified through academic databases in a given language and having removed duplicate results (Step 0 above). The third column shows the number of potentially relevant sources identified after a first screening of results on the basis of publications' titles (Step 1 above). The fourth column shows the total number of potentially relevant sources identified through academic databases for each language after a review of publications by abstract (Step 2 above).

Table B.9. Overview of results of online databases consulted after intermediate review stages

Language	Yield with no duplicates	Yield after reviewing by title	Yield after reviewing by abstract
English	3,318	1,037	38
Dutch	38	17	7
French	18	5	0

German	23	10	4

Strand 2 – Reviewing publications from selected journals and repositories

In parallel to the review by title and abstract of publications identified through systematic searches, a review of publications released since 1 January 2013 through the selected list of academic journals and grey literature repositories presented in Section B.1.1.

For most of the academic journals and grey literature repositories selected for review it was not possible to download in bulk bibliographic data of potentially relevant publications into reference manager software, as was done for sources identified through systematic searches. All sources under this research strand were therefore reviewed individually and directly on the journal or repository website.

Each publication was first screened for relevance by looking at its title. If this was deemed potentially relevant, the publication was also reviewed by abstract (consistent with the study inclusion criteria previously discussed). Publications that were deemed relevant or potentially relevant were downloaded and included in a separate, interim list of publications to be further assessed for inclusion in the study inventory. As was done for sources identified through systematic searches, publications were excluded only if it was possible to determine beyond doubt that they were not relevant to the scope of the present study.

It should be noted that the review of publications under this research strand was conducted by a different member of the study team who was not involved with the review of publications identified through systematic searches (the only exception to this being inventories in Dutch, German and French, which were reviewed by study team members involved with the first research strand).

Table B.10 gives an overview of the total number of publications reviewed on each academic journal or grey literature repository. The table also shows the number of publications selected as potentially relevant from each journal or repository after a review by title and then by abstract of each source.

Table B.10. Overview of academic journals and grey literature repositories reviewed and associated results

Database	Total publications reviewed	Results included after 'review by title'	Results included after 'review by abstract'
Terrorism and Political Violence	253	30	3
Studies in Conflict and Terrorism	289	30	3
Journal of Terrorism Research	90	3	0
CTC Sentinel	352	6	0
Perspectives on Terrorism	343	6	0
Journal of Deradicalization	103	19	5
Behavioural Sciences of Terrorism and Political Aggression	68	10	4
Critical Studies on Terrorism	168	14	0

Database	Total publications reviewed	Results included after 'review by title'	Results included after 'review by abstract'
Dynamics of Asymmetric Conflict	64	1	1
Survival	480	0	0
International Security	168	0	0
American Journal of Evaluation	236	0	0
Evaluation	200	2	1
Evaluation Review	100	0	0
Evidence and Policy	160	1	1
WODC Publications Database	448	6	3
UK Home Office Research Database	1,066	10	1
French Ministry of Interior Publications Database	259	1	0
Center for Evidence Based Crime Policy	171	0	0
German National Center for Crime Prevention	25	3	0
Organization for Security and Co-operation in Europe	628	1	0
Global Counter Terrorism Forum Violent Extremism (Hedayah)	37	5	1
Global Counter Terrorism Forum (GCTF) Tools	6	0	0
UK College of Policing – What Works Crime Reduction	63	0	0
US National Criminal Justice Reference Service	102	0	1
Alliance for Peacebuilding – Monitoring and Evaluation of CVE	21	8	2
Design Monitoring and Evaluation for Peacebuilding	64	14	1
TOTAL	**5,964**	**170**	**28**

Strand 3 – Reviewing submissions from experts and stakeholders

Lastly, as indicated in Section B.1.1, the study team contacted a number of expert and stakeholders from the fields of CT, PCVE and evaluation asking for unpublished or forthcoming manuscripts relevant to the scope of the study.

While no unpublished manuscripts were received, some of the experts contacted did provide the details of publically available sources. Each publication received was screened for relevance by title and by abstract. As for the other research strands, if it was not possible to determine beyond doubt the non-relevance of a source to the scope of the study, it was left in for further review. Some experts also submitted potentially relevant publications published outside the temporal scope of the study. In agreement with WODC and the SAC, the study team included these publications in the interim inventory, assuming that since they had been flagged by experts they were likely to be of particular relevance or importance to the purpose of the study.

Table B.11 shows the response rate obtained to solicitations sent to external stakeholders and experts, as per the methodology discussed in Section 2.1. A total of seven manuscripts collected through these solicitations were tentatively selected for inclusion in the project repository.

Table B.11. Overview of stakeholders and experts solicitation response rate and publications received

Overview of stakeholder solicitations	
Stakeholders contacted	52
Responses received	21
Manuscripts received	30
Publications tentatively included in repository	30

Finalising the inventory

After the steps detailed above, the resulting groups of publications were collated into a single, interim inventory list of sources. At this stage, a study team member who had not previously been involved with the screening process removed duplicates and conducted a second screening of abstracts. Following this, the full text of each of the publications included in the interim inventory was reviewed to determine its relevance to the study. If the reviewer could not determine with certainty whether a publication should be included or not, a separate full-text review was conducted by a second reviewer. Consensus decisions were taken for these sources and the study team endeavoured to err on the side of inclusion.

For example, although primary data were not collected during evaluative efforts, two studies by Jordan (2014) and Carson (2017) assessing the outcomes of drone and targeted killing campaigns were included as it was felt that their design and approach could contribute to shedding light on how the constraints around evaluating sensitive CT policies may be mitigated. Conversely, a study by Barkindo & Bryans (2016) on a prison-based intervention in Nigeria was excluded as this was not perceived to adopt a systematic evaluation approach, but rather to be based on non-structured observations, interactions and reflections. Similarly, the study team excluded from the Task 3 structured analysis a number of publications that focused primarily on investigating push and pull factors for violent extremism or overall ways in which CT and PCVE evaluations may be conducted, rather than discussing a single instance or

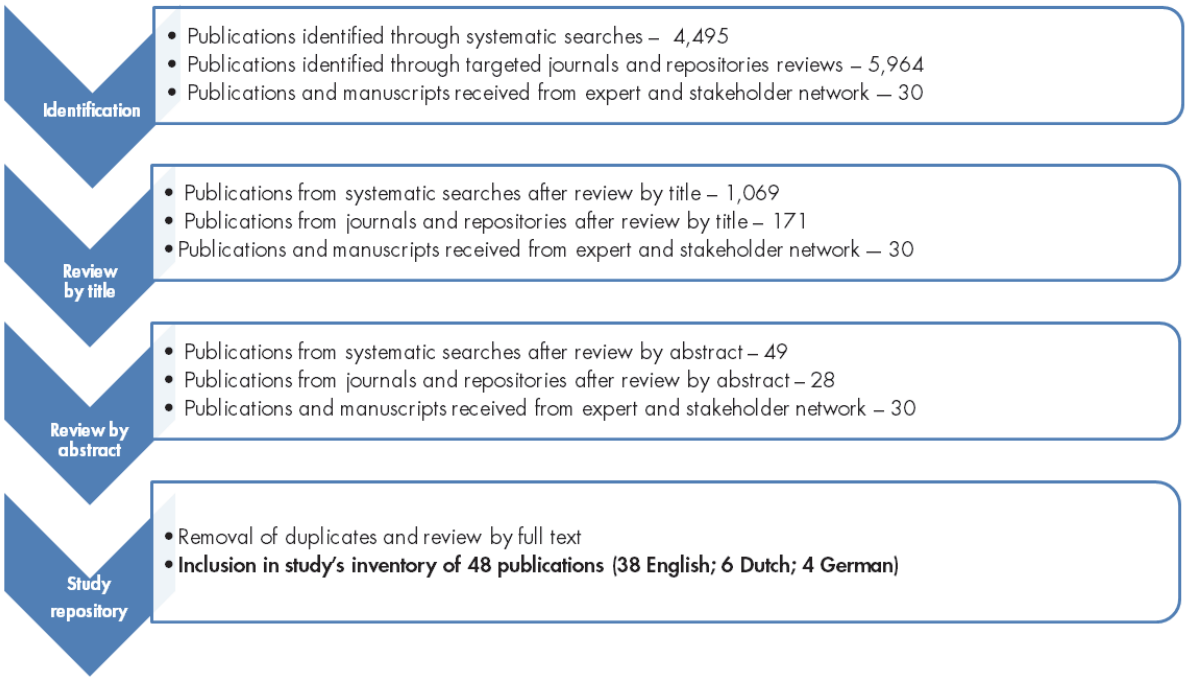
experience.[162] The exclusion of such publications from the study inventory should not be taken as an indication of their quality or contribution to the scientific debate, but merely as an assessment of their relevance to the inclusion criteria and parameters employed in the context of this study.

More broadly, it is recognised that the inventory selection and production process as described above was to an extent arbitrary. The study team appreciates that the issues and phenomena investigated (i.e. PCVE, CT, evaluation) are complex and do not lend themselves well to clear-cut definitions and inclusion/exclusion criteria. This was inevitably at odds with the basic premise of the study, which rests on a set of operational definitions and inclusion criteria that guided the work of reviewers.

The study team endeavoured to reduce ambiguity and the extent to which the review process may have led to the exclusion of relevant sources by adopting a transparent, multi-step and multi-strand review approach. Throughout the review process sources were excluded only when it could be determined beyond doubt that they were not relevant.

Figure B.4 provides a schematic overview of the inventory production process and of intermediate results and yield recorded.

Figure B.4. Overview of the inventory production process



B.2. Task 2 methodology and approach

The purpose of Task 2 was to develop an analytical framework to assess CT and PCVE evaluations. To achieve this, the study team undertook two Activities:

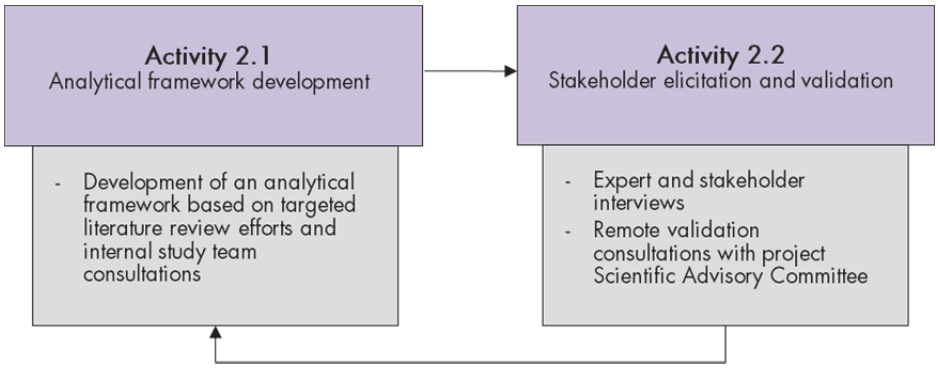
- Activity 2.1: Analytical framework development

[162] Williams & Kleinman (2013); Ferguson (2016); Mastroe (2016).

- Activity 2.2: Stakeholder elicitation and validation.

Figure B.5 provides a schematic overview of Task 2's approach and of the research activities underpinning it. The following sections describe each Activity in detail.

Figure B.5. Overview of Task 2 methodology and approach



B.2.1. Activity 2.1 – Analytical framework development

In Activity 2.1, the research team developed the study's analytical framework for assessing CT and PCVE evaluations was conducted in an iterative manner to allow for refinement through the use of expert consultation and validation activities.

Firstly, the concept and structure of the analytical framework were developed through an internal study team session designed to identify the overarching approach, purpose and structure of the framework. In the following phase, the analytical framework concept and approach was fleshed out on the basis of a targeted literature review effort and further refined through periodic study team internal consultations. Upon completion of the first analytical framework draft, the study team commenced conducting stakeholder elicitation and validation activities, as discussed in the following section.

B.2.2. Activity 2.2 – Stakeholder elicitation and validation

Throughout Task 2, in parallel to the finalisation of work on Activity 2.1, the study team conducted a series of stakeholder and expert engagement activities. In the context of Task 2, stakeholders and experts were asked to provide feedback on drafts of the analytical framework with a view to validating and finalising its content.

To minimise the risk of reporting fatigue from experts and stakeholders, the study team combined the engagement activities conducted for Task 1 and Task 2 of this study into a single set of activities. Section 2.1.3 and Annex C of this document provide further details as to the content and purpose of elicitation and validation activities undertaken.

B.3. Task 3 methodology and approach

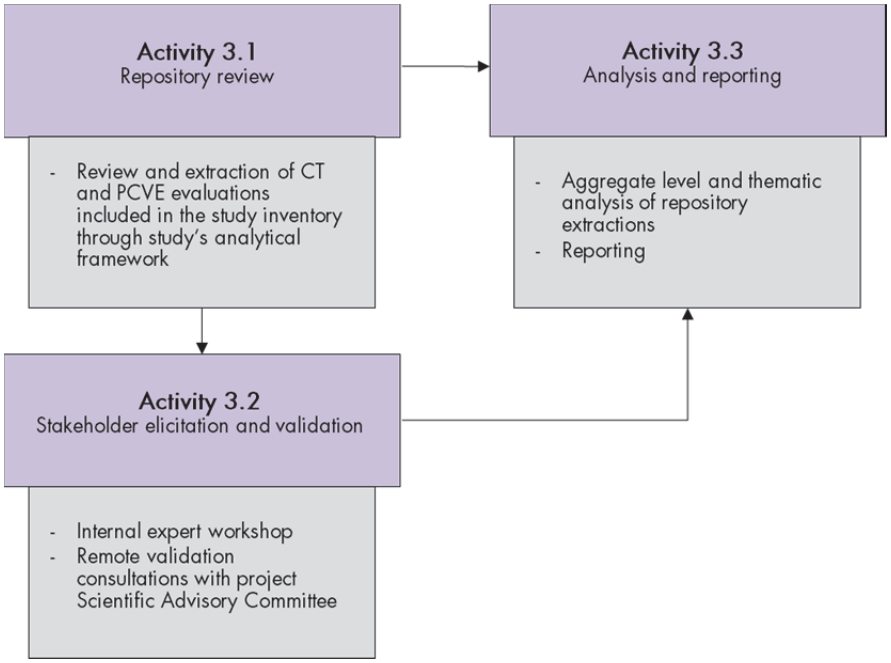
The purpose of Task 3 was to analyse the CT and PCVE evaluations inventory produced in Task 1 through the lenses of the analytical framework developed in Task 2 and to present overall findings, lessons

learned and recommendations stemming from the analysis in a synthesis report. To achieve this, the study team undertook three Activities:

- Activity 3.1: Repository review
- Activity 3.2: Stakeholder elicitation and validation
- Activity 3.3: Analysis and reporting.

Figure B.6 provides a schematic overview of Task 3's approach and of the research activities underpinning it. The followings sections describe each Activity in detail.

Figure B.6. Overview of Task 3 methodology and approach



B.3.1. Activity 3.1 – Repository review

The research team analysed publications included in the study inventory presented in Annex A according to the analytical framework developed in Task 2. To do so, the analytical framework was implemented as an analysis matrix in an electronic spreadsheet. In the analysis matrix, each column represented a data category from the framework and each row was used to capture data from an individual publication.

For each of the data categories of the analytical framework, reviewers analysing a publication were asked to complete two cells on adjacent columns. The first cell was designed to capture text or a summary of text extracted directly from the publication. The second cell was used by reviewers to provide a high-level synthesis of the value to be recorded for each data category. This was done to facilitate in subsequent phases an aggregate-level, quasi-quantitative analysis of the literature reviewed. Where possible, a list of pre-determined values to select from was included in the spreadsheet to ensure consistency. As part of this process, reviewers also extracted lessons learned concerning the evaluation as identified by the authors of the source being reviewed.

For example, with regard to evaluation methods, the spreadsheet-based implementation of the framework allowed researchers to capture information about the different data collection methods employed by

extracting relevant excerpts from the publication in a dedicated cell. Researchers were asked to include all relevant details pertaining to data collection methods in this cell (e.g. number of interviews conducted, whether interviews were conducted remotely or in person, etc.). Next to the first cell, the spreadsheet provides a space for researchers to synthesise information about what data collection methods were used (e.g. interviews, focus groups, survey) in a synthetic manner, facilitating comparability across the inventory.

At the start of the extraction phase, each study team member involved with this strand of work was asked to review and extract two publications. The extractions were then reviewed by a second senior researcher who would validate the extraction and coding approaches taken, providing feedback in case any adjustment should be made. A total of four researchers conducted the review and extraction of the sources included in the study inventory.

Upon conclusion of the first review and extraction of all sources, a study team member reviewed and consolidated the coding prepared by different researchers across all sources reviewed, not least to clean data entries. Lastly, a second and final round of validation of data points extracted in relation to the evaluation approach, evaluation design and evaluation type categories for each publication was undertaken by a senior researcher.

B.3.2. Activity 3.2 – Stakeholder elicitation and validation

Throughout Activity 3.1 and following its completion, the study team conducted two stakeholder and expert engagement activities.

Internal expert workshop

The study team engaged with peer reviewers tasked with reviewing the study's work and outputs in the context of RAND's Quality Assurance system through an elicitation and validation workshop. This took place on 12 June 2018 both in person and over teleconference and was designed to:

- Introduce project reviewers to the study and provide them with an overview of its approach, activities and emerging results.
- Elicit reviewers views as to:
    1) How do emerging findings from the study relate to those identified by other publications in the field of CT and PCVE evaluation?
    2) How do emerging findings concerning evaluation in the fields of CT and PCVE relate to evaluation practice in other (comparable) fields?
    3) What recommendations could be put forward for consideration, taking both research and practitioner perspectives into account?

Remote consultations with the project Scientific Advisory Committee

In addition to the study team internal workshop, a remote consultation with the project SAC and an in-person validation meeting were held at critical junctures during Task 3 to discuss emerging results and validate upcoming research activities.

B.3.3. Activity 3.3 – Analysis and reporting

Upon conclusion of the repository review in Activity 3.1, the data extracted from the inventory's publications were analysed at an aggregate level. In particular, compatibly with and mindful of limitations stemming from the size of the sample available, data collected under different sections of the analytical framework were examined through descriptive statistics and cross tabulations. The analysis was first conducted by a member of the study team. Emerging results were then discussed through an internal workshop with other study team members and peer reviewers.

Two researchers conducted in parallel a thematic analysis of the lessons learned extracted from publications, coding them using a bottom-up approach. The two coding and analysis results were compared and discussed by the research team in order to come to a consensus as to the content and implications of findings stemming from the inventory reviewed.

Finally, the implications inferred by different study team members from the results of the analysis of the inventory's extractions and lessons learned helped to formulate the initial recommendations presented in the concluding chapters of this study. Initial recommendations were discussed through a series of internal meetings as well as during the internal study validation workshop with peer reviewers and during the final validation meeting with the project SAC.

B.4. Limitations

This section briefly outlines various limitations to the methodology and results of the work discussed in the report.

Systematic searches

The study team endeavoured to adopt a transparent, traceable, well-documented and repeatable process for the systematic searching and reviewing of CT and PCVE evaluations published over the last five years in English, Dutch, French and German. Search and review strategies and approaches employed were designed with a view to maximising results obtained and to ensure a sufficient degree of redundancy that would allow for as many relevant sources as possible to be identified and included in the study inventory. However, due to time and resource constraints, an element of compromise was required, in particular with regard to searches on academic databases:

1. Search strings were limited to a certain number of key terms across all languages employed to the detriment of other potentially relevant ones.
2. Only a certain number of academic databases could be investigated.
3. Search engines were set to review publications by title, abstracts and keywords only, rather than by full text.

As such, it is possible that a limited number of academic publications (which did not contain keywords used, were not indexed on the databases searched, or were not written in one of the four languages selected) may have been overlooked.

More broadly, only a limited number of non-academic repositories and websites could be reviewed in the framework of the study. As such, the study inventory building approach may be biased in favour of academic journal publications and may not be able to account for:

1. Publications and repositories belonging to institutions not selected for targeted reviews.
2. Publications released through quasi- or non-peer reviewed journals and outlets not indexed in academic databases and not selected for targeted reviews.
3. Non-academic publications written in languages other than those employed for the purpose of this study.

Expert consultation

In light of the project's timeframe and resource constraints, the study team's engagement was limited to a sample of domain experts and practitioners involved in CT, PCVE and evaluation activities.

Review process

As discussed in Section 2.1.4, it is recognised that the inventory selection and production process was to an extent arbitrary. The issues and phenomena investigated (i.e. PCVE, CT, evaluation) are complex and do not necessarily lend themselves well to clear-cut definitions and inclusion or exclusion criteria. Should this study be repeated, different research teams may come to different conclusions as regards the inclusion or exclusion of a range of sources. Bearing these limitations in mind, the authors trust the study to offer relevant and reliable insights on the status of CT and PCVE evaluations based on a review of the selected sources.

The study team endeavoured to reduce ambiguity and the extent to which the review process may have led to the exclusion of relevant sources by adopting a transparent, multi-step and multi-strand review approach. Furthermore, throughout the review process sources were excluded only when it could be determined beyond doubt that they were not relevant.

Finally, 11 sources included in the study inventory date from before January 2013 and do not comply with the study's inventory third inclusion criteria. These publications were brought to the attention of the study team by external experts and stakeholders contacted during the study's undertaking. In agreement with WODC and the SAC, the study team decided to include these publications, given that they had been highlighted from experts in the fields of CT, PCVE and evaluation. It should be noted, however, that the pre-2013 sample of publications included is neither exhaustive nor systematic, as publications from before January 2013 were not searched for and reviewed under other strands of the research protocol employed.

Limitations of the analytical framework

The analytical framework presented in Chapter 4 of this report has been designed with a view to accommodating the specific purpose, needs, requirements and constraints of the present study. In particular, the framework aims to analyse in a comprehensive and homogeneous manner a broad set of evaluation reports pertaining to different types of initiatives, spanning from high-level strategies to grassroots programmes, and covering the entire spectra of CT and PCVE work. For the framework's overarching taxonomy and for the values envisioned under each analytical category to be designed, this

entailed finding a balance between comprehensiveness and retaining a manageable and comparable level of detail that would ensure the study's capacity to offer meaningful, aggregate-level analysis over the course of Task 3.

Furthermore, the analytical framework has been designed with a view to conducting the review, assessment and extraction of lessons learned from evaluation manuscripts based exclusively on data contained in them. Additional review of underpinning primary data or consultation activities with evaluation implementers, beneficiaries or practitioners was not feasible within the context of this study.

As a consequence, the analytical framework is limited in relation to the information it seeks to capture, particularly as regards the quality of the evaluation's design and conduct. This is in recognition of the fact that detailed data pertaining to evaluation design and implementation activities are unlikely to be available in short academic manuscripts and quasi-peer reviewed evaluation reports. Only a limited assessment of the quality and robustness of evaluations conducted was thus possible. The analytical framework would need to be significantly expanded in the event that it was to be used to perform an in-depth critical appraisal of CT and PCVE evaluations for which access to evaluation implementers and beneficiaries to collect primary data was possible. Further details about how the analytical framework could be expanded are discussed in Chapter 4 of this document.

Fundamental rights compliance

It was beyond the scope of this study to advise on what constitutes acceptable or suitable CT and PCVE strategy, policy or programming and to assess whether initiatives being evaluated and their evaluations complied with human, civil and any other fundamental rights. The inclusion of any particular CT or PCVE initiative in the study's inventory should not be taken as an endorsement of such practices by the study team, nor an indication that an assessment of the suitability, relevance or fundamental rights compliance of such initiatives has been undertaken in the context of this study.

Ethical considerations and human rights form a basis for sound evaluations. In the Netherlands and in a broader European context all evaluations should be designed and conducted in accordance with the rights and principles set out in the European Convention on Human Rights and the Charter of Fundamental Rights of the European Union. Compliance of CT and PCVE strategies, policies and interventions with human and civil rights should be assessed as part of an evaluation, alongside other relevant ethical issues.[163]

[163] Marret et al. (2017).

Annex C. Stakeholder elicitation details and tools

C.1. Interviewees list

Table C.1. Task 1 and 2 interviewees list

ID	Name	Function	Institution
1	Hans Nelen	Professor of Criminology	Department of Criminal Law and Criminology, Maastricht University
2	Milena Uhlmann	Research Associate	Migration, Integration and Asylum Research Centre – Research field I 'International Migration and Migration Governance', German Federal Office for Migration and Refugees
4	Phyllis Dininio	Technical Director Democracy, Governance and Conflict Area	Management Systems International
5	Anonymous	Anonymous	Anonymous
6	Anonymous	Anonymous	Anonymous
7	Anonymous	Anonymous	Anonymous
6	Anonymous	Anonymous	Anonymous

C.2. Stakeholder interview protocol

An Inventory of evaluations in counterterrorism policy and interventions

Interview Protocol

Introduction

- Introduction to self and RAND Europe
- RAND Europe has been commission by the Research and Documentation Centre (WODC) of the Dutch Ministry of Justice and Security to:
  - 'conduct a study aimed at developing an inventory of evaluations of policy and programmes pertaining to CT and PCVE; and analysing evaluations gathered in the study inventory to identify any lessons and recommendations that could help advance evaluation practice in the fields of CT and PCVE'
- Do you have any questions about the study before we begin?
  - Make sure interviewee has a copy of the information sheet and has completed a consent form

About interviewee

- *[If not academic interviewee]* Please could you tell us about your organisation, its work, and your current roles and responsibilities?
- Please could you tell us about your background and previous experience with regard to CT and PCVE?

Interview Questions

The interview should focus broadly on two areas:

- Understanding what constitutes counterterrorism (CT) and preventing and countering violent extremism (PCVE) and what are the differences / overlap / tension lines between these two; and
- CT and PCVE evaluations, understanding challenges, barriers, and enablers characterising these and capturing any lessons learned from interviewees as regards these.

1. First, we would like to discuss and test with you some of the assumptions we have made in our analytical framework. This is designed to help us review and analyse different evaluations. At this stage, we are interested in discussing with you some of the categories designed to cluster CT and PCVE interventions within the framework.

   Discuss the 'Goal', 'Type of activities' and 'Group of focus' categories and values with interviewee (one by one) and elicit his/her views and feedback on the usefulness/comprehensiveness of the values suggested under each category.
   - For background, this is what this section of the analytical framework section is designed to do. This section comprises a set of criteria designed to facilitate the capturing of fundamental details about the CT/PCVE strategy, policy, programme or initiative being

evaluated, including its nature, purpose, type of activities, group and unit of focus, status, scope (temporal and geographic), and duration.

Section 3 – Initiative characteristics		
Goal		
(Main goal of the strategy, policy or programme evaluated)	**Procure**	(The aim is to gather and assess intelligence about (potential) extremist or terrorist threats)
	Prevent	(The aim is to avert an extremist or terrorist threat from occurring in the first place, including by stifling the potential radicalization of vulnerable individuals)
	Protect	(The aim is to protect people, property and vital processes from extremist and terrorist threats, be they physical or virtual)
	Prepare	(The aim is to prepare for extremist and terrorist violence and its consequences, including through internal coordination and planning)
	Repress / Prosecute	(The aim is to contain a terrorist or extremist threat through a legal and human rights-compliant use of force, coercion, or punishing measures)
	Disengage	(The aim is to remove engagement of targets from violent activities, this does not necessarily entail a change in beliefs or attitude)
	Deradicalise	(The aim is to alter and remove the commitment of targets to an extremist ideology or belief)
Type of activities		
(Type of activities conducted or planned under the strategy, policy or programme evaluated)	**Educational**	(Activities related to the provision of knowledge or skill – e.g. classroom interventions or vocational courses)
	Communications	(Communications and information activities designed to provide alternative or counter-narratives as referred to extremist and terrorist propaganda – e.g. social media campaigns)
	Capacity building (e.g. training, counselling, coaching and mentoring)	(Collaborative activities comprising two parties developing a formal or informal relationship aimed at the transfer of skills, knowledge and attitude with the objective of development and growth of the mentee – e.g. counselling and mentoring programmes for radicalised individuals)
	Legal	(Measures and initiatives relating to the law – e.g. adoption of targeted legal measures for citizens travelling to conflict zones)
	Policing	(Law enforcement activities and measures aimed at preventing and repressing unlawful extremist and terrorist activities – e.g. community policing)
	Intelligence	(Activities entailing the collection, analysis and exploitation of information in support of CT and PCVE objectives)
	Military	(Activities conducted by national armed forces and the defence establishment in support of CT and PCVE objectives)
Group of focus		
(Primary group of focus of the strategy,	**Vulnerable / at-risk subjects**	(Subjects that are anticipated to be vulnerable to extremist or terrorist rhetoric and propaganda and that could start a process of

policy or programme evaluated)		radicalisation, potentially leading to their adoption of extremist views to support or carry out violent or terrorist activities to support their views)
	Radicalised subjects	(Subjects that have developed radical views or mindset but have not yet committed criminal offences or violent activities in support of their views)
	Offenders	(Radicalised subjects who have committed criminal offences and/or violent activities in support of their views, but have not yet been apprehended by law enforcement agencies)
	Prisoners	(Extremist and/or terrorist offenders who have been arrested and are currently in-custody in prison or in a probation programme)
	Close networks	(Close networks of individuals tied by family links or other forms of close relation and/or affection to subjects that belong to one of the above categories, i.e. vulnerable subjects, radicalised subjects, offenders and prisoners)
	Communities	(Broader communities that may comprise subjects that belong to one of the above categories, i.e. vulnerable subjects, radicalised subjects, offenders, prisoners and close networks)
	First-line workers	(First-line workers active within different lines of work (e.g. health, education, law enforcement, social work) who engage directly with individuals belonging to the categories of vulnerable subjects, radicalised subjects, offenders and prisoners)
	Other – specify	

2. In your opinion, to what extent is evaluation an established practice in the fields of CT and PCVE?
    - *Prompts:To what extent is evaluation practice a common/recurring feature of CT and PCVE strategies/programmes etc.?*
      What are barriers of evaluation in CT and PCVE at a macro/strategic level?
      What are enablers of evaluation in CT and PCVE at a macro/strategic level?
3. Do you have personal experience in conducting, commissioning or contributing to CT/PCVE evaluations?
    - *[Prompts for interviewees with direct experience of CT/PCVE evaluation]*
      Based on your experience, are there any lessons you learned as regards individual evaluations
        - *subject and scope; objectives; approach and methods; recommendations?*
      What barriers and enablers affected the design and conduct of your evaluation(s)?
4. In your opinion, what recommendations or changes are needed to advance evaluation practice in CT and PCVE?
    - *Prompts: Try to elicit recommendations or views for advancing evaluation practice both at a macro/strategic level, and micro/intervention level*

Interview Close

- We have now covered the questions we wanted to ask you. Do you have any further comments or observations?

- Can you recommend any relevant literature sources that we should consult as part of this study?
- If we have further questions, can we contact you again?
- Thank you very much.

C.3. Stakeholder interview consent form

Participant Consent Form

An Inventory of evaluations in counterterrorism policy and interventions

1. I confirm that I have read and understood the information sheet for the above study. ☐
2. I confirm that I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
3. I understand that my participation is voluntary and that I am free to withdraw up until the point where my data can reasonably be extracted without giving any reason. ☐
4. I agree to participate in the above study. ☐
5. I understand that I may request to clarify or withdraw any statements made during the course of the interview. ☐
6. My preference regarding how my data is attributed in the report and any resulting publications:
    a. Full attribution: I agree to be named and quoted in the report for the client and in any resulting publications. ☐
    b. Partial anonymity: I only agree for my organisational affiliation (role, organisation) to be associated with my quotes. The following details may be used: ☐
    c. Full anonymity: I consent to my data being used anonymously. I do not agree to be named or for my organisational affiliation to be included in reporting resulting from the research. ☐
7. I consent to my interview being audio recorded for the purpose of note taking. ☐

Name of participant (PLEASE PRINT) **Date** (add date) **Signature** (add signature)

Annex D. Analytical framework

This annex provides a detailed overview of the analytical framework employed in this study. Table D.1 provides an overview of the table structure used to present the different sections of the analytical framework with a view to facilitating its interpretation. Tables D.2, D.3 and D.4 present Sections 1, 2 and 3 of the analytical framework respectively.

Table D.1. Analytical framework table structure

Section title		
Category #1		
(Category #1 definition)	**Possible category value #1**	(Definition of possible criterion value #1)
	Possible category value #2	(Definition of possible criterion value #2)
Category #2		
(Category #2 definition)	**Sub-category #2.1**	
	(Definition of sub-category #2.1)	Possible sub-category value #1
		Possible sub-category value #2
		Possible sub-category value #3

Table D.2. Analytical framework: *Section 1 – Evaluation methodological characteristics*

Section 1 – Evaluation methodological characteristics		
Purpose		
(The main reason for / question that the evaluation aims to address: impact, process or value for money)	**Process**	(Implementation and delivery mechanisms of the policy)
	Impact / outcome	(Outcomes of the policy and difference it made)
	Economic	(Comparison of benefits and costs of the policy)
Methods		
(A strategy and tools to implement an evaluation, including how to collect and interrogate data)	**Conceptual approach**	**Extract relevant text from publication – example values:**
	(A conceptual strategy for designing and conducting evaluation efforts)	Realist evaluation; contribution analysis; etc.
	Designs	**Extract relevant text from publication – example values:**
	(A general framework for the collection and analysis of data)	Randomised Controlled Trial (RCT); Quasi-randomness: Instrumental variable (IV), Regression Discontinuity Design (RDD); Difference-in-Differences (DiD); Propensity Score Matching (PSM); Longitudinal design; Cross-sectional analysis; Before-and-after; Additionality (no control/comparison group, no sound counterfactual, e.g. ethnographic / case / comparative studies); etc.
	Data collection methods	**Extract relevant text from publication – example values:**
	(Techniques for collecting data)	Data mining; Desk-based research; Focus groups; Interviews; Surveys; etc.
	Data analysis methods	**Extract relevant text from publication – example values:**
	(Individual techniques for analysing data)	Benchmarking; Case studies; Cost-benefit analysis; Cost-effectiveness analysis; Descriptive statistics; Literature review; Logic models / Theory of change; Network analysis; Qualitative comparative analysis; Stakeholder analysis; etc.

Evaluator		
(An individual or entity conducting the evaluation)	**External**	(External from the organisation implementing the policy)
	Internal	(A dedicated team / unit / individual embedded within the organisation implementing the policy)
	Unclear – explain	Add text clarification as to why unclear
Timing		
(The timing of an evaluation in relation to the life cycle of the examined policy)	***Annex E. Ex ante***	(An evaluation conducted before the policy, programme or intervention was launched)
	Interim	(An evaluation conducted during the policy implementation: at the start, midway or towards the end of the policy cycle)
	Annex F. Ex post	(An evaluation conducted after the policy was finalised)
	Embedded (ongoing) evaluation	(Evaluation throughout the life cycle of the policy)
	Unclear	Add text clarification as to why unclear
Subject scope		
(The scope of the evaluation compared to its subject, i.e. the policy, programme, initiative, intervention, etc.)	**Multiple**	(Evaluation covers multiple policies or interventions)
	Entire	(Evaluation covers the entire policy)
	Selective	(Evaluation covers only some aspects or elements of a wider policy)
	Unclear	Add text clarification as to why unclear
Temporal scope		
(The temporal scope of an evaluation in relation to the life cycle of the examined policy)	**Multiple subsequent policies**	
	Full life cycle	
	Part of the life cycle	
	Not specified	

Geographical scope		
(The geographical scope of the evaluation)	**Transnational**	
	National	
	Regional	
	Local	
	Unclear	Add text clarification as to why unclear
Evaluation criteria		
(Standards against which the policy is assessed)	**Relevance**	(Relevance measures the extent to which the policy is suited to the priorities and needs of the target group, recipient and client)
	Effectiveness	(Effectiveness is a measure of the extent to which the policy attains its objectives)
	Efficiency	(Efficiency measures the outputs in relation to the inputs. It is an economic term which signifies that the policy uses the least costly resources possible in order to achieve the desired results)
	Impact	(The positive and negative changes produced by the policy, directly or indirectly, intended or unintended)
	Sustainability	(Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn)
	Other – please specify	(E.g. scalability; transferability; coherence, etc.) – add text explanation
	Unclear	Add text clarification as to why unclear
Evaluation questions		
(Are the questions the evaluation seeks to answer stated?)	**Stated**	
	Unclear	

Table D.3. Analytical framework: *Section 2 – Evaluation quality characteristics*

Section 2 – Evaluation quality characteristics		
What method was used to quality assess the evaluation?		
(Methods used in the study to ensure quality)	**External blind review**	
	Independent peer review	
	Self-assessment	
	Project-embedded independent review	(e.g. steering group / committee, advisory panel)
	Other – please specify	
	Unclear	Add text clarification as to why unclear
Is the evaluation based on multiple perspectives?		
(Evidence and data come from multiple groups of stakeholders)	**Yes – please specify**	
	No	(e.g. only from participants)
Does the evaluation employ grids, rubrics, scores or indicators?		
(Investigates whether a framework is used for assessing different levels of performance during implementation)	**Applied**	
	To some extent – please specify	
	Not applied	
Does the evaluation identify any lessons learned?		
(Key conclusions, observations or points discussed in relation to design and conduct of the evaluation)	**Stated**	
	Unclear	Add text clarification as to why unclear

Does the evaluation feature recommendations for policy improvements?	
(Recommendations formulated as a result of the evaluation)	**Recommendations addressed at specific task-holders AND time-bound** **Recommendations addressed at specific task-holders OR time-bound** **No task-holders and timeframe indicated for the recommendations** **No recommendations identified**
Does the evaluation feature a focus on unintended effects and consequences of the strategy, policy or programme being evaluated?	
(Considerations on possible or registered side effects of the strategy, policy or programme being evaluated)	**Yes – please specify** **No**

Table D.4. Analytical framework: *Section 3 – Initiative characteristics*

Section 3 – Initiative characteristics		
Type		
(What type of initiative is being evaluated: strategy, policy, programme or intervention)	**Strategy**	(A plan of action adopted officially by an organisation, institution or group of people designed to achieve long-term, overall aims as regards a particular topic or issue area and whose implementation rests on a number of subordinated policies and programmes)
	Policy	(A course or principle of action adopted officially by an organisation, institution or group of people, potentially in support of the achievement of a higher-level strategic approach or vision as regards a particular topic or issue area)
	Programme or intervention	(A set of actions taken to improve a situation, resolve a conflict, transfer skills or generate a positive transformation in an individual, group or social context)
Goal		
(Main goal of the strategy, policy or programme evaluated)	**Procure**	(The aim is to gather and assess intelligence about (potential) extremist or terrorist threats)
	Prevent	(The aim is to avert an extremist or terrorist threat from occurring in the first place, including by stifling the potential radicalization of vulnerable individuals)
	Protect	(The aim is to protect people, property and vital processes from extremist and terrorist threats, be they physical or virtual)
	Prepare	(The aim is to prepare for extremist and terrorist violence and its consequences, including through internal coordination and planning)
	Repress/prosecute	(The aim is to contain a terrorist or extremist threat through a legal and human rights-compliant use of force, coercion or punishing measures)
	Disengage	(The aim is to remove engagement of targets from violent activities, this does not necessarily entail a change in beliefs or attitude)
	Deradicalise	(The aim is to alter and remove the commitment of targets to an extremist ideology or belief)

Ideology		
(Type of extremist ideology addressed by programme. A free text box will permit to capture the words used by the text to describe the target ideological group in greater level of detail – e.g. Islamist foreign fighters)	**Religious extremism**	
	Political extremism	
	Other	e.g. environmentalism
Level of intervention		
(Level of intervention of the strategy, policy or programme evaluated)	**Macro**	(The macro level refers to activities or measures implemented at the level of entire social systems – e.g. nations; regions; cities)
	Meso	(The meso level refers to activities or measures implemented at the level of small-to-medium sized groups, including both institutions (e.g. schools, hospitals) and affective social groups (communities, peer groups)
	Micro	(The micro level refers to activities or measures implemented at the level of individuals)
Type of activities		
(Type of activities conducted or planned under the strategy, policy or programme evaluated)	**Educational**	(Activities related to the provision of knowledge or skill – e.g. classroom interventions or vocational courses)
	Communications	(Communications and information activities designed to provide alternative or counter-narratives as referred to extremist and terrorist propaganda – e.g. social media campaigns)
	Capacity building (e.g. training, counselling, coaching and mentoring)	(Collaborative activities comprising two parties developing a formal or informal relationship aimed at the transfer of skills, knowledge and attitude with the objective of development and growth of the mentee – e.g. counselling and mentoring programmes for radicalised individuals)
	Legal	(Measures and initiatives relating to the law – e.g. adoption of targeted legal measures for citizens travelling to conflict zones)

	Policing	(Law enforcement activities and measures aimed at preventing and repressing unlawful extremist and terrorist activities – e.g. community policing)
	Intelligence	(Activities entailing the collection, analysis, and exploitation of information in support of CT and PCVE objectives)
	Military	(Activities conducted by national armed forces and the defence establishment in support of CT and PCVE objectives)
Group of focus		
(Primary group of focus of the strategy, policy or programme evaluated)	**Vulnerable / at-risk subjects**	(Subjects that are anticipated to be vulnerable to extremist or terrorist rhetoric and propaganda and that could start a process of radicalisation, potentially leading to their adoption of extremist views to support or carry out violent or terrorist activities to support their views)
	Radicalised subjects	(Subjects that have developed radical views or mindset but have not yet committed criminal offences or violent activities in support of their views)
	Offenders	(Radicalised subjects who have committed criminal offences and/or violent activities in support of their views, but have not yet been apprehended by law enforcement agencies)
	Prisoners	(Extremist and/or terrorist offenders who have been arrested and are currently in-custody in prison or in a probation programme)
	Close networks	(Close networks of individuals tied by family links or other forms of close relation and/or affection to subjects that belong to one of the above categories, i.e. vulnerable subjects, radicalised subjects, offenders and prisoners)
	Communities	(Broader communities that may comprise subjects that belong to one of the above categories, i.e. vulnerable subjects, radicalised subjects, offenders, prisoners and close networks)
	First-line workers	(First-line workers active within different lines of work (e.g. health, education, law enforcement, social work) who engage directly with individuals belonging to the categories of vulnerable subjects, radicalised subjects, offenders and prisoners)

	Other – specify
Implementer	
(Main organisation or actor responsible for the implementation of the strategy, policy or programme evaluated)	**International organisation** **Governmental actor** **Non-governmental actor** **Public-private partnership**
Status	
	Completed **Ongoing**
Period of implementation	
(Actual time of implementation of a strategy, policy or programme)	**Please record available information**
Geographic scope	
	Transnational **National** **Regional** **Local**
Foreseen duration of programme	
(Expected duration of the total implementation of a strategy, policy or programme)	**3+ years** **1–3 years** **6–12 months** **<6 months**

Length of implementation	
(Duration of the intervention, or lifetime of the strategy, policy or programme at the time of the evaluation)	**3+ years** **1–3 years** **6–12 months** **<6 months**