



INDEED

Evidence - Based Model for Evaluation of
Radicalisation Prevention and Mitigation

Deliverable 1.2

D1.2 Report outlining identified, analysed and recommended research approaches, methods and tools for evidence-based evaluation coming from the area of PVE/CVE and De-radicalisation and other selected disciplines

September 2022 (M13)

Author: Stephan Klose (VUB)

Abstract:

This report aims to prepare the ground for the development of an evidence-based approach to evaluation in the field of PVE/CVE and De-radicalisation (P/CVE/DeRad). The report carries out three key tasks. First, it provides a multi-disciplinary review of the literature surrounding the term "evidence-based practice" (EBP) which seeks to generate a better understanding of the term's usage as well as challenges for its practical application in different disciplines. Second, it undertakes a multi-disciplinary review of evaluation designs which outlines their key strengths and weaknesses, as well as their usage within the field of P/CVE/DeRad. Finally, the report introduces the term evidence-based evaluation (EBE) as "a process of planning and implementing evaluations which integrates available external evidence, professional expertise and stakeholder values, preferences and circumstances". This approach, the report suggests, does not only represent an application of EBP in the field of evaluation, but also a potential catalyst for the uptake of EBP (in P/CVE/DeRad and beyond) insofar as it promises to promote the usage of robust evaluation designs as well as the principles of EBP.



This project has received funding by the European Union's
Horizon 2020 research and innovation programme
H2020-SU-SEC-2020 under grant agreement no 101021701



Information table

Project Acronym	INDEED
Deliverable Number	1.2
Deliverable Title	Report outlining identified, analysed and recommended research approaches, methods and tools for evidence-based evaluation coming from the area of PVE/CVE and De-radicalisation and other selected disciplines
Version	1.2
Status	Version Re-Submitted to EC
Responsible Partner	VUB
Main author	Stephan Klose (VUB)
Contractual Date of Delivery	30.07.2022
Type	Report (R)
Actual Date of Delivery	16.09.2022
Re-submission Date of Delivery	12.01.2023 14.06.2023
Dissemination Level	Public

This document reflects only the author's views and not that of the Research Executive Agency. The Research Executive Agency is equally not responsible for any use that may be made of the information contained in this document. This document may not be reproduced or copied without permission. © Copyright in this document remains vested in the Project Partner.





Document history

Version Number	Date	Status	Author	Description
0.1	31.07.2022	Draft	Stephan Klose (VUB)	First draft
0.2	01.08.2022-12.08.2022	First version	WP1 Partners	Review by WP1 Partners and input
0.3	28.08.2022	First version	Stephan Klose (VUB)	Pre-final version of the report
0.4	28.08.2022-08.09.2022	Second version	Stephan Klose (VUB)	Pre-final version of the report
0.5	09.09.2022	WP1 Partners accepted	WP1 partners, INDEED coordinator	Review by WP1 partners
0.6	09.09.2022	PC accepted	Marzena Kordaczuk-Was (PPHS)	Review by Project Coordinator
0.7	12.09.2022	Final version	Stephan Klose (VUB)	Finalised the report
1.0	16.09.2022	Submitted to EC	Marzena Kordaczuk-Was (PPHS); Natalia Jarmuzek-Troczyńska (PPHS)	Final approval and submission
1.1	12.01.2023	Re-Submitted to EC	Stephan Klose (VUB) Marzena Kordaczuk-Was (PPHS); Natalia Jarmuzek-Troczyńska (PPHS)	Editing, minor content improvements and re-submission
1.1.1	08.06.2023	Revision draft	Stephan Klose (VUB)	Revision incorporating comments received from reviewers at INDEED mid-term review
1.2	14.06.2023	PC accepted; Re-submitted to EC	Marzena Kordaczuk-Was (PPHS); Natalia Jarmuzek-Troczyńska (PPHS)	Final approval and re-submission





Executive Summary

This report aims to prepare the ground for the development of an evidence-based approach to evaluation in the field of PVE/CVE and Deradicalisation (hereinafter P/CVE/DeRad). **The report is divided into three parts.**

The first part of the report provides an overview of the evolution of the evidence-based movement from its inception in the field of medicine in the 1990s to its later application across the scientific spectrum. The report clarifies key terms of the evidence-based movement, including the central concept of evidence-based practice (EBP) which is defined as “a decision-making process which integrates available external evidence, professional expertise and client values, preferences and circumstances”. It further discusses challenges to the implementation of EBP, both in P/CVE/DeRad and across academic disciplines, as well as the question of “what constitutes evidence?”. The report highlights that the application of EBP in the field of P/CVE/DeRad faces multiple challenges, which include conceptual difficulties as well as a lack of robust primary evaluations.

The second part of the report provides an overview of (meta-)evaluation designs and their usage in the field of P/CVE/DeRad. It systematically delineates, compares and discusses the strengths and weaknesses of notable evaluations designs which include systematic reviews, meta-analyses, realist reviews, randomized controlled trials, quasi-experimental designs, longitudinal and cross-sectional studies, theory-based designs, stakeholder-oriented designs and economic evaluation designs. This part of the report further reviews the application of such designs in the field of P/CVE/DeRad. It highlights that most evaluations in P/CVE/DeRad take the form of a cross-sectional study in which the evaluator collects data from recipients (and other stakeholders) of an intervention at one specific point in time without the use of a control group. Longitudinal and (quasi-)experimental designs, as well as theory-driven evaluations, by contrast, remain in short supply which significantly limits the strength of the evidence base in P/CVE/DeRad.

The third and concluding part of the report makes the case for and conceptualizes an evidence-based approach to evaluation in P/CVE/DeRad. Drawing on the definition provided for EBP, the report defines “evidence-based evaluation” (EBE) as “a process of planning and implementing evaluations which integrates available external evidence, professional expertise and stakeholder values, preferences and circumstances”. The report further discusses the applicability of EBE while highlighting challenges to its implementation, such as the often-underdeveloped evidence base in P/CVE/De-Rad and evaluation research, the frequent lack of sufficient time and resources for conducting evaluations as well as the sensitivity of data which creates challenges for the realization of participatory evaluation practices. The report, finally, stresses the utility and added value of an EBE tool (to be developed in WP3 of INDEED) which can guide the planning and implementation of EBE in the field of P/CVE/De-Rad.





Table of Contents

Executive Summary.....	4
1 Introduction	8
1.1 Background and objectives	8
1.2 Methodology.....	9
1.3 Strengths and limitations.....	9
1.4 Report structure.....	9
2 Part I: The evidence-based movement and its role in PVE/CVE and De-radicalisation	11
2.1 The evidence-based movement: origins and concepts	11
2.1.1 The movement's origins	11
2.1.2 Evidence-based practice and evidence-based interventions.....	12
2.1.3 What counts as evidence?.....	13
2.2 Challenges to the implementation of evidence-based practice.....	14
2.2.1 Creating the evidence base: advances and challenges	14
2.2.2 Challenges to the implementation of evidence-based practice	15
2.3 Evidence-based PVE/CVE and De-radicalisation.....	17
2.3.1 What is evidence-based PVE/CVE and De-radicalisation?	17
2.3.2 Creating an evidence base for PVE/CVE and De-radicalisation	18
2.3.3 Implementing an evidence-based approach to PVE/CVE and De-radicalisation.....	19
3 Part II: Evaluation designs and their usage in PVE/CVE and De-radicalisation	20
3.1 Meta-evaluation designs	20
3.1.1 Narrative and scoping review	20
3.1.2 Systematic reviews and meta-analyses	21
3.1.3 Realist and EMMIE-based reviews	23
3.1.4 Summary	24
3.2 Evaluation designs for primary studies	25
3.2.1 Randomized controlled trials	26
3.2.2 Natural experimental studies and quasi-experimental studies.....	27
3.2.3 Longitudinal and cross-sectional designs	28
3.2.4 Case series, case reports and case-control designs.....	28
3.2.5 Theory-based designs	29
3.2.6 Stakeholder-oriented designs	30
3.2.7 Economic evaluation designs	31
3.2.8 Summary	32
3.3 Evaluation designs in the field of PVE/CVE and De-radicalisation	34
3.3.1 Meta-evaluation designs used in the field of PVE/CVE and De-radicalisation	34
3.3.2 Evaluation designs for primary studies used in the field of PVE/CVE and De-radicalisation	37
4 Part III: Towards an evidence-based approach to evaluation in PVE/CVE and De-radicalisation.....	40
4.1 Taking stock and moving ahead	40
4.2 Evidence-based evaluation: definition and key principles	40
4.3 Evidence-based evaluation: What it is and what it isn't	41
4.4 The relationship between evidence-based practice and evidence-based evaluation ..	42
4.5 Challenges to the implementation of EBE.....	43
4.6 Towards an evidence-based evaluation tool.....	43
5 Conclusion	46
6 Reference List	47





List of Figures

Figure 2: Overview of key concepts	13
Figure 3: Key challenges to the implementation of EBP	17
Figure 4: Prevalence of evaluation designs, 1990-2014 (based on Feddes and Galluci, 2015/6) and 2013-2017 (based on Bellasio et al., 2018)	38
Figure 5: Number of evaluations by design, 2013-2017 (based on Bellasio et al., 2018)	39
Figure 6: Evidence-based evaluation	41

List of Tables

Table 1: Overview of meta-evaluation designs.....	24
Table 2: Overview of evaluation designs for primary studies	32
Table 3: Overview of systematic reviews published by the Campbell Collaboration in the field of PVE/CVE and De-radicalisation	35
Table 4: Prevalence of evaluation designs in PVE/CVE and De-radicalisation between 1990 and June 2014 (based on Feddes and Galluci, 2015/6)	37
Table 5: Prevalence of evaluation designs in PVE/CVE and De-radicalisation between 2013 and 2017 (based on Bellasio et al., 2018)	37

List of Acronyms

Acronym	Definition
INDEED	Strengthening a comprehensive approach to prevent and counteract radicalisation based on a universal evidence-based model for evaluation of radicalisation prevention and mitigation
P/CVE	Preventing violent extremism/ countering violent extremism
P/CVE/DeRad	Preventing violent extremism/ countering violent extremism and de-radicalisation
CVE	Countering violent extremism
EBP	Evidence-based practice
EBE	Evidence-based evaluation
EMMIE	Effect, Mechanism, Moderators, Implementation and Economic Costs
EBM	Evidence-based medicine
EBSW	Evidence-based social Work
EBI	Evidence-based initiative
RCT	Randomized controlled trial
RAN	Radicalisation Awareness Network
EUCPN	European Crime Prevention Network
GRADE	Grading of Recommendations Assessment, Development and Evaluation





CONSORT	Consolidated Standards of Reporting Trials
QES	Quasi-Experimental study
NES	Natural experimental study
CMOs	Context-Mechanism-Outcome
ToC	Theory of Change
RCS	Repeated cross-sectional
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
WP	Work Package





1 INTRODUCTION

The introduction briefly situates the deliverable within the INDEED project before outlining the report's key objectives, methodology, limitations and structure.

1.1 BACKGROUND AND OBJECTIVES

This report represents a deliverable of the INDEED project which aims to strengthen the knowledge, capabilities and skills of PVE/CVE and De-radicalisation (hereinafter P/CVE/DeRad) first-line practitioners and policy makers in designing, planning, implementing and evaluating initiatives based on an evidence-based approach.

INDEED aims to develop:

1. A universal Evidence-Based Evaluation Model (EBEM) for evaluating radicalisation prevention and mitigation initiatives.
2. A practical EBEM-based evaluation tool.
3. A collection of user-friendly repositories (repositories of radicalisation factors and pathways into radicalisation; factors strengthening resilience to radicalisation, repositories of evidence-based practices) for practical use by practitioners and policy makers.
4. Targeted curricula and trainings (offline/ online).
5. Lessons learnt and policy recommendations.

All results will be integrated and openly accessible in the INDEED multilingual Toolkit for practitioners and policy makers in the field for the entire lifecycle of PVE/CVE and De-radicalisation initiatives, from design to evaluation.

This report falls into the realm of INDEED's Work Package (WP) 1 which aims to prepare the ground for the creation of an Evidence-Based Evaluation Model in the field of P/CVE/DeRad. The main objectives of WP1 are:

1. To support the development of an evidence-based evaluation model (EBEM) for radicalisation prevention and mitigation initiatives.
2. To gather and analyse evaluation approaches, models and tools.
3. To provide updated knowledge on existing risk and protective factors.

The report (deliverable 1.2) primarily addresses objective 1 of WP1 as well as the above-noted objectives 1 and 2 of INDEED. The deliverable is described as follows:

- A report outlining identified, analysed and recommended research approaches, methods and tools for evidence-based evaluation coming from the area of P/CVE/DeRad and other selected disciplines.

In view of the deliverable description and the ambitions of WP1, the key objectives of the report have been defined as follows:

1. To provide a multi-disciplinary review of (challenges to) the implementation of evidence-based practice (EBP), including in P/CVE/DeRad.
2. To outline, delineate and systematically compare different evaluation designs suggested in the evaluation literature, and to discuss their utilization in the field of P/CVE/DeRad.
3. To conceptualize the term evidence-based evaluation (EBE), and to discuss its applicability and challenges for its implementation in the field of P/CVE/DeRad.
4. To discuss the value an EBE tool can add to already existing evaluation tools in P/CVE/DeRad and the broader field of crime prevention.





1.2 METHODOLOGY

The report is based on extensive desk research which was conducted by WP1 partners between November 2021 and June 2022. During this period, WP1 partners systematically collected information on 1) the usage of evaluation designs, methods and tools as well as 2) approaches and challenges to the implementation of EBP across several academic disciplines, including criminology, education, medicine, public health, social work and P/CVE/DeRad. The implementation of the data collection process followed the timeline, tools and task distribution outlined in D1.1 (Methodological guidelines for Partners on data collection).

The report was implemented in close coordination with WPs 2 and 3, and benefited from the feedback provided by WP1 partners both during the data collection and the drafting phase. The report moreover benefited from the organization of a Research Forum on 24 May 2022 in Brussels which brought together experts in the field of evaluation and evidence-based practice for an extensive discussion of the challenges addressed in the report. A summary of the Research Forum can be found on the [INDEED website](#).

1.3 STRENGTHS AND LIMITATIONS

The report, in addressing the objectives outlined above, promises to make several key contributions, both to the literature on evidence-based practice and evaluation in P/CVE/DeRad as well as the further development of INDEED. Specifically, **the report makes three sets of contributions:**

First, the report presents a uniquely comprehensive and systematic overview of the evolution of EBP, including of challenges to its implementation, across several disciplines which serves to define (and avoid misconceptions of) EBP in P/CVE/DeRad. **Second**, the report provides an extensive overview of evaluation designs and methods, and their usage in P/CVE/DeRad, which offers insights into the development and current state of the field's evidence base. **Finally**, the report applies the concept of EBP to the field of evaluation in P/CVE/DeRad by developing and discussing the concept of EBE which will pave the way for the development of INDEED's EBEM and EBE tool (to be developed in WP3).

While the report has several strengths, it also includes notable limitations. Importantly, it must be stressed that while the report attempts to be comprehensive in both its multi-disciplinary review of EBP as well as its overview of evaluation designs, it inevitably only provides a selective account. To partially mitigate this constraint, the report explicitly directs readers to alternative attempts at cataloguing (certain types of) evaluation designs where these are available. It moreover transparently lists the disciplines which serve as the foundation for its multi-disciplinary review of the evolution of EBP. Secondly, it must be noted that this report only provides a tentative and preliminary discussion of the term EBE which will be further developed and specified in INDEED's WP3 ("Development of the Evidence-Based Evaluation Model (EBEM) for radicalisation prevention and mitigation and an Evaluation Tool dedicated to the PVE/CVE/De-radicalisation initiatives").

1.4 REPORT STRUCTURE

This report is divided into three parts. In its **first part**, the report outlines the evolution of the evidence-based movement across several disciplines. It discusses the concept of EBP and the question of "what constitutes evidence?". Furthermore, it outlines key challenges to the realization of EBP across multiple disciplines, including in P/CVE/DeRad.





The **second part** of the report zooms in on the subject of evaluation. It provides an extensive overview – and a comparative discussion of the strengths and weaknesses – of (meta-) evaluation designs. Moreover, it analyses the usage of such designs in the field of P/CVE/DeRad to provide a better understanding of the field's current evidence base.

Finally, in its third part, the report applies the concept of EBP to the realm of evaluation by introducing the notion of EBE as “a process of planning and implementing evaluations which integrates available external evidence, professional expertise and stakeholder values, preferences and circumstances”. The report further discusses the applicability of EBE while highlighting challenges to its implementation. Finally, it stresses how an EBE tool which guides the planning and implementation of EBE can add value to existing evaluation tools in the field of P/CVE/De-Rad.





2 PART I: THE EVIDENCE-BASED MOVEMENT AND ITS ROLE IN PVE/CVE AND DE-RADICALISATION

The first part of this report provides an overview of the evolution of the evidence-based movement. The review is divided into three sections. The **first section** examines the origins and key concepts of the evidence-based movement, as well as the question of "what counts as evidence?". The **second section** reflects on key challenges for the implementation of EBP, including the creation of a solid evidence base. Finally, the **third section** discusses the state of the evidence-based movement in the field of P/CVE/DeRad, including challenges to the implementation of EBP in this area.

2.1 THE EVIDENCE-BASED MOVEMENT: ORIGINS AND CONCEPTS

The first section of this report draws on multiple disciplines – notably medicine, public health, education, social work and criminology – to provide an overview of the evolution of the evidence-based movement. After discussing the origins of the movement in the field of medicine, this section reviews the movement's key concepts before addressing the question of "what counts as evidence?".

2.1.1 THE MOVEMENT'S ORIGINS

The evidence-based movement, whose philosophical origins date back to the 19th century and earlier (Sackett, 1997: 3), **first took off in the field of medicine** where in 1992 the evidence-based Medicine Working Group, in a landmark article published in the Journal of the American Medical Association (JAMA), proclaimed evidence-based medicine (EBM) as a "new paradigm for medical practice" (Guyatt et al., 1992: 2420).

Proponents of this new paradigm described EBM as a medical decision-making process in which practitioners competently identify and consult the best available "external evidence" (i.e. the best available medical research) rather than rely on their experience, intuition or traditional training (Evidence-based Working Group, 1992: 2420). As famously put by David Sackett, EBM concerns the "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett, 1997: 3).

While EBM emphasizes that practitioners should consult external evidence in their decision-making, it, importantly, does not promote a (top-down) process in which external evidence determines professional practice. Quite to the contrary, EBM advocates for a professional decision-making which integrates "the best research evidence with [...] clinical expertise and [a] patient's unique values and circumstances" (Strauss et al., 2018: 18).

Following this notion, **EBM, broadly speaking, embraces three key principles**. **First**, the practice of EBM commits clinicians to make decisions informed by the best available external evidence. This requires practitioners to consult, and to acquire the skills to appraise, the best available medical research, rather than rely on (often outdated) textbooks or the counsel from senior colleagues (Sackett et al., 1996: 71; Strauss et al., 2018: 20).

Second, EBM commits clinicians to consider the values, preferences and circumstances of individual patients in the application of acquired external evidence. It thus advocates for a patient-centered approach to medicine which values "compassion, sensitive listening skills [...] and understanding of [a] patient's illnesses in the context of their experience, personalities and culture" (Guyatt and Busse, 2006: 30, see also Greenhalgh et al., 2014: 3).

Third, EBM commits clinicians to draw on and develop their professional expertise. EBM, as





highlighted by Strauss et al. (2018: 18), in this regard, involves the usage of “clinical skills and past experiences to rapidly identify each patient’s unique health state and diagnosis [and to] integrate evidence with patient values”.

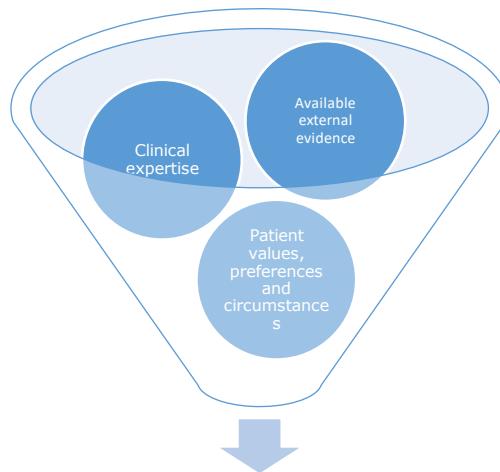


Figure 1: The key ingredients of evidence-based medicine

Guided by these three principles, EBM has, since the 1990s, developed into a forceful movement which has been described as one of the greatest 15 medical milestones since 1840 by the British Medical Journal (BMJ) (Montori and Guyatt, 2008: 1814). In the late 1990s, this movement rapidly expanded beyond the field of medicine to other disciplines, particularly other fields related to human services, such as education, social work or criminology.

2.1.2 EVIDENCE-BASED PRACTICE AND EVIDENCE-BASED INTERVENTIONS

As the evidence-based movement rapidly expanded in the late 1990s and early 2000s, the terms “evidence-based” and “evidence-based practice” (EBP) quickly gained popularity across and beyond academic circles. In many disciplines, such as **public health** (Jenicek, 1997: 189), **education** (Davies, 1999: 117) or **social work** (Gambrill, 1999: 346), EBP was introduced akin to EBM as a decision-making process which integrates 1) available external evidence, 2) professional expertise and 3) client values, preferences and circumstances.

Ever since, EBP has remained controversial, especially for its association with the notion that “non-research factors are not important” (Parrish, 2018: 408; see also Cowen, 2019). This association, as suggested by Parrish (2018: 408), has partly been the result of misconceptions of EBP circulating within and outside of academia, including in policy documents. A particularly important source of conceptual confusion, in this regard, has been the often-fuzzy distinction between EBP and “best practices” – sometimes (and rather unhelpfully) labelled evidence-based practices” (EBPs).

As stressed in the field of evidence-based social work (EBSW), these two concepts require careful separation (Spensberger et al., 2019). While EBP, as noted above, refers to a decision-making process which integrates available external evidence, professional expertise, and client values, preferences and circumstances, EBPs instead denote “concrete intervention models based on strong research evidence” (Wike et al., 2019: 506). To better distinguish between EBP and EBPs, some proponents of the evidence-based movement have suggested to replace the notion of EBPs with the concept of empirically supported treatments, evidence-supported interventions or





evidence-based interventions (Spensberger et al., 2019). For the sake of clarity and consistency, and in line with INDEED's terminology, this report will use the term evidence-based initiatives (EBIs) to speak about EBPs in the remainder of this text.



Figure 1: Overview of key concepts

2.1.3 WHAT COUNTS AS EVIDENCE?

From its inception, the evidence-based movement has been closely associated with the promotion of a clear hierarchy of study designs, so-called "**levels of evidence**", which can provide academics and practitioners alike with guidance on the reliability of study results (Burns et al., 2011). This hierarchy has been traditionally topped by randomized controlled trials (RCTs) – or systematic reviews thereof – which have been widely regarded as a "**gold standard**" method for evaluating the efficacy of a specific intervention (Sackett, 1996: 72). The evidence-based movement, in this regard, has stood firmly in the tradition of promoting RCTs – a tradition first pioneered by Archie Cochrane in the field of medicine in the 1970s and 80s (Brownson, Fielding and Green, 2017: 3; Claridge and Fabian, 2005: 552).

Today, this tradition is perhaps most clearly preserved in the name and mission of the British charitable organization Cochrane (formerly the Cochrane Collaboration) which acts as the most important provider of systematic reviews – focused on RCTs and quasi-experimental research – in all areas of health care. The Campbell Collaboration, which extends the work of Cochrane to the social sciences, moreover, has emerged as another key organization for the evidence-based movement whose mission it is to strengthen the evidence base in fields like education, social work and criminology through systematic reviews of (primarily) RCTs and quasi-experimental studies. The significance of RCTs for the evidence-based movement is further reflected in the prevailing standards for identifying EBIs. As noted by Wike et al. (2019: 506), EBIs are "based on strong research evidence, most typically defined as a minimum of two randomized controlled trials". RCTs, within the evidence-based movement, thus have long played a privileged role in the creation of an evidence base, which some scholars have argued to be unwarrantedly elevated (see, for instance, Worral, 2007).

While proponents of EBP have emphasized the importance of RCTs, they, however, have also long argued that RCTs cannot answer every question. From criminology (Clear, 2009) to medicine (Burns et al., 2011) to education (Davies, 1999), scholars have highlighted that the most suitable study design ultimately depends on the **type of evidence** sought. Rychetnik et. al. (2004), in this context, distinguished between **three basic types of evidence**: **type 1** evidence which suggests that "something should be done", **type 2** evidence which indicates "which intervention or policy should be done", and **type 3** evidence which suggests "how something should be done". Following Rychetnik et al. (2002: 122-3), the latter type of evidence in particular calls for non-RCT designs which are better suited to evaluate the implementation, unintended consequences, sustainability or contextuality of an intervention.





More recently, Lancaster, Rhodes and Rosengarten (2020: 4), further highlighted that in times of an acute crisis, such as during the COVID-19 pandemic, also “limited anecdotal, indirect and circumstantial findings ought not to be ignored, given [...] the consequences of inaction”. As further stressed by Greenhalgh et al. (2020), “in the face of a pandemic the search for perfect evidence may [indeed] be the enemy of good policy”. As these examples show, **discussions within the evidence-based movement about “what should count as evidence”, while often centered on (quasi-)experimental research (and particularly RCTs), are generally more nuanced.**

At the same time, however, there remains to this day **considerable debate and confusion about how different types of evidence should best be integrated**. Critics of the evidence-based movement rightly stress that the movement continues to largely privilege experimental over non-experimental, and quantitative over qualitative methods, and that oftentimes the notion of multiple lines of evidence is more rhetorical than rooted in practice (Shahjahan, 2011: 190). **The question of “what should count as evidence”, thus, remains far from resolved.**

2.2 CHALLENGES TO THE IMPLEMENTATION OF EVIDENCE-BASED PRACTICE

Following up on the introduction to the evidence-based movement provided above, the following section will reflect on the **key challenges for implementing EBP in practice**. After reviewing challenges to the creation of an evidence base for EBP across multiple disciplines, this section discusses **the problems practitioners face in implementing EBP in their daily work**.

2.2.1 CREATING THE EVIDENCE BASE: ADVANCES AND CHALLENGES

The creation of an evidence base constitutes one of the core tasks and challenges of the evidence-based movement. Thus far, a key tenet of the implementation of this task, next to the promotion of RCTs, has been the expansion of systematic reviews through organizations like Cochrane and the Campbell Collaboration which help practitioners navigate – and draw conclusions from – the existing scientific literature. The expansion of such reviews has been particularly welcomed in fields faced with an overwhelming number of (experimental) studies such as medicine and education. As highlighted by Beerkens (2018: 282), “the issue [in the field of education] in practice is often not the lack of evidence but too much evidence with no professional consensus, which suggests a need for good synthesis studies that could inform policies”.

While some fields struggle with the sheer number of studies, others, in turn, have been faced with the opposite problem: a lack of evidence, particularly a lack of (quasi-)experimental studies. Clear (2009: 5-6), in this regard, remarked that “as impressive as systematic reviews [produced by the Campbell Collaboration in the field of criminology] are, they all include language that bemoans a weak empirical basis [...] many summaries offer global conclusions but, in doing so, refer to studies with disparate, and even opposing, findings”. In disciplines like criminology, thus, the main challenge remains the promotion of empirical studies which can provide a solid basis for meaningful synthesis.

As these examples indicate, the overall strength of the evidence base can vary greatly from one discipline to another. Disparities, however, also exist within and across disciplines. Rychetnik et al. (2002: 125), for instance, noted that “[as] evidence is often gathered on simple interventions and from groups that are easy to reach in a population [...] little evidence exists on interventions for disadvantaged groups”. Empirical studies have, furthermore, long focused on evaluating interventions in particular geographical areas (e.g. the global North, North America or urban areas) which questions the transferability of relevant findings. In the field of criminology, Blaustein (2016: 166), in this context, warned, for instance, against the unreflective transfer of





(putatively evidence-based) policing methods from the global North to areas with a different security environment, such as Latin America.

Brownson et al., (2017: 4), moreover, highlighted that the evidence base in a given field is often non-conducive to practitioners' needs and circumstances. In the worst case, existing studies (and reviews thereof) can be misleading, either because the number and quality of studies produced on competing interventions is heavily skewed, or because study results are outright corrupted by vested interests. The latter problem has been particularly severe in the field of medicine where industry-funded RCTs have, for instance, shown exaggerated treatment effects for SSRIs (a group of antidepressants such as 'Prozac') and atypicals (a group of drugs used to treat bipolar disease) which, in the absence of independent RCTs or RCTs for alternative (but less profitable) treatments such as cognitive behavioural therapy or exercise, have led clinicians to prescribe seemingly evidence-based but inefficient treatments for mental illnesses (Every-Palmer and Howick, 2014; see also Kemm, 2006; Greenhalgh et al., 2014; Ioannidis, 2016).

Altogether, despite a significant increase in empirical studies (and syntheses thereof) across multiple disciplines, the evidence-based movement remains confronted with the problem of an often weak, patchy and (in some cases) corrupted evidence base. In view of such challenges, proponents of EBP have called for measures to promote empirical research, particularly in the global South and on disadvantaged groups, to control the influence of vested interests, and to create networks between practitioners and researchers to ensure that empirical studies speak more clearly to practitioner needs (see, for instance, Heneghan et al., 2017).

2.2.2 CHALLENGES TO THE IMPLEMENTATION OF EVIDENCE-BASED PRACTICE

The uptake of EBP across different disciplines has, thus far, often remained rather low, including in areas where practitioners not only know of the approach but also view it largely positively (Parrish, 2018; Shapira, Enosh and Havron, 2017). Even in the field of medicine, the usage of EBP can be underwhelming. A recent cross-sectional study, for instance, concluded that only 14.2% of French and Swiss healthcare professionals practice EBM on a daily basis while 65% don't practice it despite knowing or having heard about it (Lafuente-Lafuente et al., 2019).

A key reason for the often-disappointing uptake of EBP lies in the quality and relevance of existing research. A systematic review by Sadeghi-Bazargani et al. (2014: 795-796), in this regard, suggests that the by far most common barrier to the practice of evidence-based medicine is a "research barrier" which finds expression, for instance, in conflicting research findings or the poor generalizability and limited practical significance of available studies. Another reason for the low uptake of EBP can be found in practitioners' often-limited knowledge of and experience with EBP (Shapira, Enosh and Havron (2017: 187). Proponents of the evidence-based movement, in this context, have long emphasized the need to strengthen the role of EBP in professional education and university curricula (Rosen, 2003). A study by Wike et al. (2019: 513) in the field of social work suggested in this regard that the use of EBP and EBIs are associated with "having a MSW [Master of Social Work] field placement that emphasizes EBP [as well as] having MSW faculty who emphasized the importance of EBP in teaching".

While the inclusion of EBP into university curricula has overall advanced significantly over the past decades, notable differences remain across disciplines and national contexts. In the field of medicine, "EBM is [now] a component of the foundation years training programme in the UK, the focus of graduate assessment in the USA [...] a requirement of practicing physicians in Canada [...] and increasingly popular worldwide at both a graduate and undergraduate level" (Ahmadi et al., 2015). In the field of social work, by contrast, EBP, while included into a majority of master programs in the US (Wike et al., 2019: 504), has yet to be incorporated into the university curricula of many other countries, such as Israel (Shapira, Enosh and Havron, 2017: 187).





Next to the inclusion of EBP into university programmes, its incorporation into professional education remains another challenge for the evidence-based movement which is of particular significance in fields like public health where more than 50% of workers (in the United States) enter the field without formal training in a related discipline (Brownson et al., 2009: 182). Much debate exists about how such education should best be organized, whereby some argue that "a multi-faceted approach that entails a combination of methods like lectures, computer sessions, small group discussions, journal clubs, and assignments [is] more likely to improve learners' EBP knowledge" (Kyriakoulis et al., 2016) while others (Wilke et al., 2019: 505) stress that to be effective training for practitioners should best be combined with ongoing consultation and supervision. As further highlighted by Spensberger et al. (2019: 35), the right method may depend on the type of learners, whereby "[some] (e.g. novices) [may] benefit more from teacher-centred instruction, while others [including practitioners may] learn more from student-centred instruction".

While EBP training methods remain debated, EBP proponents agree that, next to education, the workplace environment constitutes a second critical arena for promoting the uptake of EBP. As highlighted by Wike et al. (2019: 513), the use of EBP and EBIs is not only associated with a conducive educational environment but also an "organizational culture that emphasizes EBP". In the same vein, Shapira, Enosh and Havron (2017: 195) stress that "working in a non-EBP-friendly environment might take the wind out of [practitioners'] sails". Altogether, scholars across multiple disciplines have long suggested that the uptake of EBP hinges greatly on whether their work environment provides practitioners with the time, resources, training and encouragement needed to effectively implement EBP in day-to-day situations (see for instance, Brownson, Fielding and Green, 2017; Van der Zwet, 2018).

Besides the educational and organizational environment, proponents of EBP have moreover highlighted the significance of a conducive political environment for the uptake of EBP and EBIs, particularly in areas where problems require large-scale interventions, and where such interventions are subject to public scrutiny and ideological debate (Brownson et al., 2009: 189-90). In the field of criminology, Welsh and Farrington (2001: 168), for instance, stressed that "it is all too common that evidence becomes secondary to the political and policy considerations of the day".

Finally, a challenge for the uptake of EBP can be found in the dissemination and formatting of evidence. The evidence-based movement has long been criticized for its reliance on a linear and unidirectional dissemination of evidence through journal articles, systematic reviews, online databases and practice guidelines (Mullen et al., 2007: 328; Brownson et al., 2017: 5). Such forms of dissemination, as stressed by Rosen (2003: 201), place the burden on practitioners and create expectations which are "not only unfair [but] also unrealistic". To address these challenges, some scholars have called for a more participatory approach to the implementation of EBP which draws on community-based participatory research (Kohatsu et al., 2004: 420) or an organizational excellence model that engages local partnerships between research and practice (Van der Zwet, 2018: 133). Brownson et al. (2017), moreover, called on researchers to not only disseminate their work through traditional channels (e.g. academic conferences, peer-reviewed journals) but to also to engage with platforms through which practitioners generally learn (e.g. webinars, workshops).





1. Low quality and relevance of the existing evidence base
2. Poor dissemination and formatting of evidence
3. Lack of EBP trainings and education
4. Non-conducive work-place or political environment

Figure 2: Key challenges to the implementation of EBP

2.3 EVIDENCE-BASED PVE/CVE AND DE-RADICALISATION

After the previous two sections have reviewed the key tenets of and challenges faced by the evidence-based movement across different disciplines, **the final section of this part of the report zooms in on the emergence of an evidence-based approach to P/CVE/DeRad**. Following a brief discussion of how an evidence-based approach to P/CVE/DeRad may be best defined, **this section will review relevant challenges in P/CVE/DeRad for the creation of an evidence base and the implementation of EBP**.

2.3.1 WHAT IS EVIDENCE-BASED PVE/CVE AND DE-RADICALISATION?

Over the past decade, calls have markedly increased for the implementation of an evidence-based approach in the field of P/CVE/DeRad (Aiello, Puigvert and Schubert, 2018; Baruch et al., 2018; Feddes and Gallucci, 2015; Marsden, 2020; Nehlsen et al., 2020; Pistone et al., 2019). While, in the wake of such calls, the term "evidence-based" has become ubiquitous, both within and outside the field's academic conversations, its definition and the meaning and implications of an evidence-based approach to P/CVE/DeRad, however, have often remained ambiguous.

Scholars advocating for an evidence-based approach to P/CVE/DeRad, drawing on criminology and particularly the work of Sherman (1998), have notably foregrounded the need to identify and utilize EBIs (for a discussion, see Nehlsen et al., 2020: 6). Less attention, by contrast, has been given to the encouragement of EBP – the practice of integrating available evidence with professional expertise and client values, preferences and circumstances. The conversation in P/CVE/DeRad about the implementation of an evidence-based approach has consequently revolved primarily around the question of "what works?" (see, for instance, Feddes and Gallucci, 2015: 3).

This focus is understandable considering the flurry of political initiatives and the lack of empirical evaluations thereof which characterize the P/CVE/DeRad landscape (Marsden, 2020; Pistone et al., 2019). However, as noted by Freese (2014) in the adjacent field of counterterrorism, a too strong emphasis on the question of "what works?" can also come at a cost. Specifically, Freese (2014: 50) highlighted that "although it may be difficult to push for evidence-based practice before sufficiently rigorous research is available to drive these practices, we should [also] recognize that, just as in business, demand drives supply". Emphasizing EBP as the end goal of implementing an evidence-based approach to P/CVE/DeRad, and highlighting the role of practitioners in this process, thus may act as an important catalyst for strengthening the field's empirical base.

It should further be noted that practitioners, as suggested by a study in field of social work (Van der Zwet, 2018), may show a more positive attitude towards an evidence-based approach if this approach means implementing EBP rather than EBIs – a sentiment that may be heightened if EBIs, as in P/CVE/DeRad, lack firm empirical grounding. Against this backdrop, advocates of an evidence-based approach to P/CVE/DeRad may benefit from taking inspiration not only from a "what works" agenda but also from the concept of EBP defined as a decision-making process





which integrates 1) available external evidence, 2) professional expertise and 3) client values, preferences and circumstances.

2.3.2 CREATING AN EVIDENCE BASE FOR PVE/CVE AND DE-RADICALISATION

A key challenge for the realization of an evidence-based approach to P/CVE/DeRad is the development of a solid evidence base which allows for the meaningful identification of EBIs as well as the implementation of EBP. Many efforts are currently underway to meet this challenge, including the expansion of empirical research and the development of systematic and scoping reviews aimed at evaluating the “state of the art” (Marsden, 2020).

A key role in this process is played by the Campbell Collaboration, an international research network which publishes systematic reviews, and which provides guidance on standards for primary data collection (Marsen, 2020: 2). In 2015, driven by a five-country partnership – including Australia, Canada, New Zealand, the UK and the US – the Campbell Collaboration set up a dedicated CVE programme which since has led to the publication of a wide range of systematic reviews in P/CVE/DeRad, covering topics such as the effectiveness of counter-narratives (Carthy et al., 2020), the significance of risk and protective factors (Wolfowicz et al., 2021), and the effectiveness of police programmes to strengthen community connectedness (Mazerolle et al., 2020) (see section [3.3.1](#) for a full list).

While many of these reviews, in the absence of robust primary studies, can only provide limited guidance on “what works” in the field of P/CVE/DeRad, they play an important part in illuminating the field’s state of the art. Recent reviews have especially highlighted the field’s lack of robust (including experimental) primary evaluations which has largely persisted in spite of a growing number of empirical studies (see section [3.3.2](#) for further analysis) (Feddes and Galluci 2015/6; Gielen, 2017; Lum et al., 2008; Pistone et al., 2019). Against this backdrop, scholars have emphasized both the need for as well as key challenges to the implementation of strong evaluation designs (see, for instance, Baykal et al., Baruch et al., Bellasio et al., 2018; Feddes and Galluci, 2015; Nehres et al., 2020).

In view of the latter, scholars have highlighted in particular that rigorous evaluations in P/CVE/DeRad are often challenged by security and ethical concerns, the inaccessibility of target groups and an unavailability of (often highly sensitive) data (Bellasio et al., 2018: 61-62). Other hurdles include conceptual difficulties which complicate the identification and isolation of suitable indicators for measuring the effectiveness of P/CVE/DeRad interventions (Clement et al. (2021: 5) as well as poor evaluation planning, scarcely implemented [theories of change](#) and time and resource constraints (Bellasio et al., 2018: 63-66). Finally, the often-low quality of evaluations in P/CVE/DeRad has been attributed to unstable funding schemes and a lack of trust which keep organizations, especially smaller grassroots organizations, from adopting complex and robust evaluation practices (Baykal et al., 2021: 8; Clement et al., 2021: 13-4).

Altogether, scholars have shown that the lack of robust evaluation designs in P/CVE/DeRad is due not only to structural challenges but also very practical ones (such as poor evaluation planning) which should be addressed more effectively. Bellasio et al. (2018: 69), in this regard, argued that “greater efforts should be made to ensure that wherever possible [\(quasi-\)experimental designs](#) are employed within CT and PCVE evaluations”. Apart from the promotion of strong evaluation designs, some scholars further emphasized that a key challenge for developing a stronger evidence base in P/CVE/DeRad lies in the creation of effective linkages between primary and meta-evaluation designs (Gielen, 2017). Traditional systematic reviews, which privilege the inclusion of (quasi-)experimental designs can, in the absence of such studies in many areas of P/CVE/DeRad, often only provide limited insights into primary research findings (see also [table 3](#)) (Nehlsen et al., 2020). Against this backdrop, some scholars have stressed the need for and utility of alternative reviews, such as [realist reviews](#), which do “not value one evaluation method over the other [but rather build on the premise] that each evaluation study





can be valuable in terms of analyzing relevant contexts, mechanisms, and outcomes" (Gielen, 2017: 4).

2.3.3 IMPLEMENTING AN EVIDENCE-BASED APPROACH TO PVE/CVE AND DE-RADICALISATION

Proponents of the evidence-based movement in P/CVE/DeRad have so far primarily focused their efforts on strengthening the field's evidence base through the promotion of (high-quality) empirical research and the development of meta-evaluations (particularly systematic reviews). However, as shown in other disciplines, **the effective promotion of EBP requires not only a strong evidence base but also effective measures to strengthen the dissemination of evidence, practitioner engagement and the creation of strong links between research and practice.**

Some, if limited, advances, in the development of measures to strengthen the dissemination and utilization of evidence have already been made in the field of P/CVE/DeRad, including at the European level. The **EU's Radicalisation Awareness Network** (RAN), for instance, shares its own series of publications directly with a strong network of working groups and frontline practitioners while **Impact Europe**, another EU initiative, offers an online platform which provides practitioners with information and guidance about evaluation designs and methods (Impact Europe, 2022a). Organizations like the **Campbell Collaboration**, moreover, have implemented measures to reach a wider audience which include "plain language summaries" of its systematic reviews (for an example, see Campbell Collaboration, 2020). Together, these initiatives point to increasing and diversifying efforts in P/CVE/DeRad to provide channels (if somewhat dispersed ones) through which practitioners can retrieve information about the field's evidence base.

Some notable efforts, moreover, have been made in the development of stakeholder networks which can serve as platforms to facilitate the implementation of EBP. Once again, such efforts are visible at the European level where actors like RAN and the **European Crime Prevention Network** (EUCPN) facilitate a sustained dialogue between communities of research and (frontline) practice. Extending such networks further, including to the national and local level, where they often remain underdeveloped, constitutes one of the key challenges for promoting the wider usage of EBP in the field of P/CVE/DeRad. Local and national networks, multi-stakeholder platforms and working groups, and communities of practice all have the potential to create additional important channels through which evidence (and the lack thereof) can be effectively identified and communicated from academia to practitioners and vice versa. They, moreover, may serve as frameworks for EBP trainings, as well as the joint development of tools which facilitate the integration of evidence in daily decision making.





3 PART II: EVALUATION DESIGNS AND THEIR USAGE IN PVE/CVE AND DE-RADICALISATION

The second part of this report provides an overview of evaluation¹ designs, including of their usage in the field of P/CVE/DeRad. It is divided into three sections. The first section offers a review of meta-evaluation designs aimed at the synthesis of existing studies. This section delineates, compares, and discusses the strengths and weaknesses of systematic reviews, meta-analyses, narrative reviews, realist reviews, scoping reviews and EMMIE-based reviews. The second section provides a comparative overview of primary evaluation designs. Specifically, it outlines the strengths and weaknesses of randomized controlled trials, natural experiments and quasi-experimental designs, longitudinal and cross-sectional studies, case series and case-control studies, theory-based designs, participatory designs as well as economic evaluation designs. Finally, the third section discusses the extent to which the reviewed evaluation designs have been applied in the field of P/CVE/DeRad.

3.1 META-EVALUATION DESIGNS

The first section of part II of the report will review different methods for evaluating and synthesizing the “state of the art”. It will start by discussing narrative and scoping reviews before examining the methodology for systematic reviews and meta-analyses. Finally, this section will discuss the methodology used in realist and EMMIE-based reviews which have been developed more recently to complement and expand the scope of traditional systematic reviews focused on determining the effect size of an intervention. The overview provided below is non-exhaustive and based on a subjective selection of relevant designs. A more comprehensive overview of meta-evaluation designs can be found in the work of Sutton et al. (2019) who have recently identified and systematized 48 different types of reviews.

3.1.1 NARRATIVE AND SCOPING REVIEW

In many disciplines, particularly in the social sciences, **narrative reviews** are the main method by which scholars evaluate the state of the art on a particular topic or research question. Indeed, this report itself takes the form of a narrative review. In narrative reviews, the author informs the reader about the state of the scientific literature by presenting and discussing, in narrative form, the most relevant studies on a particular topic or set of research questions (Collins and Fauser, 2005; Ferrari, 2015). Studies are typically selected and included into a narrative review based on an implicit (and more or less systematic) process in which the authors draw on their professional knowledge and judgement (Garg et al., 2018: 253).

Narrative reviews have several **advantages**. They enable authors to bring to bear their professional expertise, and they allow for in-depth discussions and critiques of relevant academic arguments and debates (Ferrari, 2015). Narrative reviews, moreover, allow for the presentation of relevant information on a specific topic in a format and language which readers can easily access, follow and engage with². They furthermore provide an often less time-consuming alternative to systematic reviews. At the same time, however, narrative reviews also have notable **disadvantages**. Most importantly, their selection and inclusion of studies follows an often subjective and opaque process. Readers may find it difficult to ascertain if the overview

¹ An evaluation describes the assessment of an initiative (or set of initiatives) in a limited period of time which may focus, for instance, on the initiative’s impact, sustainability or relevance. Evaluations can be distinguished from monitoring exercises which can take the form of a continuous data collection process.

² While narrative reviews can be more accessible than alternative meta-evaluations, they, in practice, however, often remain difficult to access for readers, particularly practitioners unfamiliar with the scientific discourse surrounding the reviewed subject.





provided by a narrative review is sufficiently extensive and balanced. As noted by Garg et al. (2018: 253) it is often uncertain "whether the author of a narrative review selectively cited reports that reinforced his or her preconceived ideas or promoted specific views on a topic". Narrative reviews, thus, can lack transparency and include biases which are difficult to detect for readers without extensive knowledge on the subject.

Compared to narrative reviews, **scoping reviews** promise a more rigorous and transparent overview of the existing literature on a specific topic. A scoping review aims to identify and present all available studies within a particular field or for a specific topic irrespective of their research design. It generally entails five analytical steps. After 1) formulating the research question and 2) identifying studies based on transparent search criteria, a scoping review 3) selects all studies with (and eliminates those without) relevance for the research question, 4) charts the collected data, and 5) collates, summarizes and reports the results (Arksey and O'Malley, 2005: 27-28).

Conducting a scoping review can serve multiple purposes. On the one hand, a scoping review can prepare the ground for a full systematic review by indicating the scope, costs, feasibility, and potential utility of conducting a systematic review (Arksey and O'Malley, 2005: 25-26). On the other hand, it can provide a useful overview of the extent, range and nature of existing studies, and point out relevant gaps in the literature which can serve the development of a needs-oriented research agenda (Pistone et al., 2019). Compared to systematic reviews, scoping reviews have the **advantage** of capturing all available studies irrespective of their designs which allows for a complete picture of the research landscape for a particular topic. They, moreover, are generally less time-consuming and easier to implement.

Scoping reviews, however, are subject to **limitations** as well. Notably, scoping reviews, unlike systematic reviews, do not engage in a synthesis of the available literature. Instead, they are focused on and limited to a presentation of the landscape of existing studies. As they do not weigh and compare the findings of individual studies, they cannot provide an indication of the cumulative evidence which exists, for instance, for the effectiveness of a specific intervention (Arksey and O'Malley, 2005: 42).

3.1.2 SYSTEMATIC REVIEWS AND META-ANALYSES

Systematic reviews are a hallmark of the evidence-based movement. Their aim is to systematically synthesize available evidence, most notably on the effectiveness of specific treatments or initiatives. In contrast to narrative reviews, systematic reviews follow an explicit and transparent methodology by which relevant evidence is identified and weighed (Garg et al., 2018; Collins and Fauser, 2005). In comparison to scoping reviews, they typically only include primary studies which meet specified quality standards. Often their focus is on synthesizing the findings of experimental and quasi-experimental evaluations (Arksey and O'Malley, 2005). As such, systematic reviews are generally regarded by proponents of the evidence-based movement as one of the methods which can produce the most robust and reliable evidence (e.g. for the effectiveness of a certain treatment) (see, for instance, Burns et al., 2011).

As noted above, the promotion of systematic reviews, which has its origins in the field of medicine and the EBM movement, has increasingly taken hold across various academic disciplines. This process has notably been driven by the establishment of Cochrane and the Campbell Collaboration, two international networks, which have been central to the production and codification of systematic reviews. In 2014, the Campbell Collaboration, based on reporting standards created by Cochrane, published the "Methodological expectations of Campbell Collaboration intervention reviews" (MECCIR), updated in 2019, which provide a framework for conducting systematic reviews in fields such as P/CVE/DeRad (Campbell Collaboration, 2022c)³.

³ For an alternative reporting standard, see the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Arya et al., 2021).





Broadly speaking, the development of systematic reviews, as promoted by Campbell, follows five steps. After the 1) formulation of the main research objectives and 2) the transparent identification of search methods and selection criteria, the study 3) collects, appraises and synthesizes available studies, 4) presents the results and 5) discusses the implications, which may include research and policy recommendations.

For the appraisal and synthesis of available studies, **systematic reviews draw on a variety of tools** which notably include the "Grading of Recommendations Assessment, Development and Evaluation" (**GRADE**) system. GRADE, developed in 2000 in the field of medicine, provides a set of widely used standards designed to enable researchers to classify the quality of evidence from "high quality (top trustworthiness) to low quality (the bottom) with a category of moderate in between" (Djulberg and Guyatt, 2020: 165). By applying GRADE, evidence from RCTs (starting out as "high quality") and observational data (starting out as 'low quality') can be rated up or down (e.g. based on a study's execution) to provide a clearer indication of the quality and certainty of research findings, on the basis of which recommendations (strong or weak) can be formulated in favor or against a particular intervention (Siemieniuk and Guyatt, 2022).

Another important method used in systematic reviews for the synthesis of available studies are **meta-analyses**. Meta-analyses refer to mathematical methods for combining ("pooling") the results of available studies (Walker et al., 2008). Such pooling, as noted by Garg et al. (2008: 254) "may provide a more precise estimate of the underlying "true effect" [of a certain intervention] than any other study" by increasing the size of the overall sample and statistical power of the analysis. In some cases, only meta-analyses, by pooling results and increasing the sample size, can establish the benefits of an intervention while smaller trials are more likely to show non-significant effects (Garg et al., 2008).

Systematic reviews and meta-analyses have several notable **advantages**. They allow for presenting readers with a transparent overview of the most significant studies in a particular field, and they enable a systematic and replicable calculation of the overall effect size of a particular intervention. Moreover, they serve to establish whether effect sizes of a particular intervention are consistent and generalizable across different settings and populations, and whether they vary by particular subgroups (Garg et al., 2008: 254). Systematic reviews and meta-analyses, however, are also subject to several **limitations**. Notably, the strength of a systematic review or meta-analysis ultimately depends on the availability of well-conducted primary studies. Systematic reviews and meta-analyses, as noted by Garg et al. (2008: 255), indeed are "only as reliable as the methods used to estimate the effect in each of the primary studies". Moreover, the robustness of systematic reviews and meta-analyses can be undermined by publication biases – the biases resulting from studies having a higher chance of being published if they demonstrate significant effect sizes – if these cannot be effectively mitigated (Walker et al., 2008). Where systematic reviews mix well-conducted and poorer studies they, moreover, may end up suggesting effect sizes which are less reliable than those demonstrated by a single well-conducted evaluation (Garg et al., 2008).

Scholars highlighting the limitations of systematic reviews, moreover, frequently stress their oftentimes narrow focus on effect sizes and strict inclusion criteria (see, for instance, Greenhalgh et al., 2007). This focus, on the hand, may limit their ability to sufficiently address the mechanisms or circumstances of an intervention's (lack of) effectiveness which, in turn, reduces their practical utility (Pawson et al., 2005). On the other, their often-strict inclusion criteria can create challenges in fields where experimental and quasi-experimental designs, a focal point of systematic reviews, are, for ethical and practical reasons, scarce or difficult to implement. In such research fields, systematic reviews, drawing on a very small evidence base, provide only limited guidance on the effectiveness of individual initiatives.





3.1.3 REALIST AND EMMIE-BASED REVIEWS

Realist reviews (or realist syntheses) refer to a relatively new type of meta-evaluation first comprehensively described by Pawson et al. (2005) in the field of public health. By asking how and why – rather than whether and to what extent – an initiative is effective, they aim to complement and transcend the scope of traditional systematic reviews. Unlike traditional systematic reviews which primarily focus on understanding the effect size of particular initiatives, realist reviews take an interest in explaining the mechanism by and context in which initiatives fail or succeed. As noted by Pawson et al. (2005: 21), they aim to explain “what works for whom, in what circumstances, in what respects and how”.

At a practical level, realist reviews aim to collect, evaluate and (where relevant) adjudicate between different [theories of change](#) which existing studies on a particular initiative (implicitly or explicitly) suggest (Greenhalgh et al., 2007). Their development follows eight basic steps: After 1) identifying the research question, and 2) clarifying the purpose of the review, realist reviews, 3) collect and articulate available programme theories, 4) search for evidence, 5) appraise the evidence, 6) extract results, 7) synthesize findings and 8) draw conclusions and make recommendations (Berg and Nanavati, 2016: 2).

This review process has several **advantages**. Realist reviews are well-equipped to provide a transparent overview and comparative assessment of different theories of change noted in the literature. This, in turn, can generate important insights into why a particular intervention has shown to be (in)effective (by a systematic review), and thus offer highly relevant and practical information for policy makers, including about when, how and for whom a particular intervention should best be implemented (Greenhalgh et al., 2007). Realist reviews, however, also are subject to **limitations**. While aiming for high implementation standards (Berg and Nanavati, 2016: 3; Wong et al., 2013) they are generally less standardizable and reproducible than systematic reviews (Pawson et al., 2004: 32). Their utility, moreover, relies heavily on the existence of well-conducted primary studies which (in the best case) clearly identify, discuss and rigorously assess a variety of theories of change. Finally, due to their complexity, they require a high level of reviewer expertise “in both [the] academic (critical appraisal of empirical studies) and service (programme implementation) domain” (Pawson et al., 2004: 32).

A related and even more complex type of meta-evaluation can be found in the **EMMIE-based review**. This type of review, which has recently been developed and promoted in the field of criminology by the What Works Centre for Crime Reduction, follows a framework which was initially created as a scoring system for systematic reviews (Johnson et al., 2015). Specifically, EMMIE was launched as a coding framework aimed at establishing the extent to which existing reviews provide indications not only of (E) effect sizes but also the (M) mechanisms and mediators, the (M) moderators and contexts, the (I) implementation factors and the (E) economic costs (and benefits) of an initiative (Johnson et al., 2015: 463-469).

This framework has since served to guide a new type of meta-evaluation which combines a traditional systematic review (focused on calculating the effect size of an intervention) with elements of a realist review (highlighting the mechanisms and context which enable an initiative’s effectiveness) and a synthesis of information on challenges to the implementation and economic costs of an initiative (see, for instance, Belur et al., 2020; Sidebottom et al., 2017). The key **advantage** of an EMMIE-based review is that it offers a comprehensive discussion of the utility of a particular intervention. This discussion goes beyond the simple assessment of an initiative’s effect size and offers information about the conditions under which an intervention is best implemented, as well as a discussion of potential challenges to a successful implementation, including associated costs (Johnson et al., 2015). EMMIE-based reviews, in this regard, promise to speak to all central concerns which a practitioner faces when assessing the (dis)advantages and feasibility of potential (alternative) initiatives.

While EMMIE-based reviews constitute a promising alternative to traditional systematic and realist reviews, they, like other types of reviews, however, also face some **constraints**. Notably,





EMMIE-based reviews are of a high complexity and require considerable expertise, time and resources. They, moreover, rely on the existence of well-conducted primary studies which address the different aspects covered by EMMIE. Existing EMMIE reviews have pointed out that primary studies covering the MMIE part of the EMMIE framework, and particularly the latter two elements, are often in short supply (Belur et al., 2020). While such scarcities have placed limitations on EMMIE-based reviews, a positive side effect, however, has been that these reviews have been able to point out gaps in the literature which may have otherwise remained unaddressed.

3.1.4 SUMMARY

Based on the preceding discussions, the table below provides a summary of the key strengths and weaknesses, as well as the associated methods and tools, of different meta-evaluation designs.

Table 1: Overview of meta-evaluation designs

Evaluation design	Strengths	Weaknesses	Associated methods and tools
Narrative review	<ul style="list-style-type: none"> Provides space for in-depth discussions and critiques of the existing literature Accessible and engaging format Less-time consuming than other review types Not limited to the inclusion of specific study designs (unlike many systematic reviews) 	<ul style="list-style-type: none"> The inclusion of publications is subjective and prone to selection biases Heavy reliance on the author's knowledge of the field – the overview of the literature is not necessarily complete or accurate 	<p>Methods:</p> <ul style="list-style-type: none"> Desk research
Scoping review	<ul style="list-style-type: none"> More rigorous and transparent than narrative reviews Not limited to the inclusion of specific study designs (unlike systematic reviews) Less time-consuming and easier to implement than systematic reviews Provides an overview of the scope of the existing literature; points out relevant gaps Can prepare the ground for a systematic review 	<ul style="list-style-type: none"> Includes no synthesis of research findings Offers no in-depth discussion of the arguments/ data presented in the literature Provides no indication of cumulative evidence (e.g., for the effectiveness of an intervention) 	<p>Methods:</p> <ul style="list-style-type: none"> Desk research
Systematic review	<ul style="list-style-type: none"> Presents a transparent overview of the literature (especially of (quasi-)experimental research) in a particular field; points out relevant research gaps Provides a systematic and replicable calculation of the overall effect size of a particular intervention or group of interventions Can offer indications for whether the effect sizes of a particular intervention are consistent and generalizable across different settings and populations, and whether they vary by particular subgroups 	<ul style="list-style-type: none"> More time-consuming than narrative or scoping reviews Heavy reliance on available (high-quality) primary studies Often focused narrowly on (quasi-)experimental study designs and the study of effect sizes which can lower its applicability Does not provide an in-depth discussion of the available literature 	<p>Methods:</p> <ul style="list-style-type: none"> Desk research (to collect data) Scoping review Meta-analysis (to analyse data) <p>Tools:</p> <ul style="list-style-type: none"> GRADE (quality appraisal tool) MECCIR, PRISMA (reporting tools) JBI Checklist for Systematic





			Reviews (appraisal tool)
Meta-analysis	<ul style="list-style-type: none"> Provides a systematic and replicable calculation of the overall effect size of a particular intervention or group of interventions Can offer indications for whether the effect sizes of a particular intervention are consistent and generalizable across different settings and populations, and whether they vary by particular subgroups 	<ul style="list-style-type: none"> Heavy reliance on available (quantitative) primary studies Suggested effect sizes can be distorted by publication biases – the biases resulting from studies having a higher chance of being published if they demonstrate significant effect sizes Does not provide an in-depth discussion of the available literature 	<p>Methods:</p> <ul style="list-style-type: none"> Funnel plot (a method to mitigate publication biases) <p>Tools:</p> <ul style="list-style-type: none"> PRISMA (reporting tool)
Realist review	<ul style="list-style-type: none"> Provides a transparent overview and comparative assessment of different theories of change noted in the literature Can generate insights into why a particular intervention has shown to be (in)effective Can offer recommendations for when, how, and for whom a particular intervention should best be implemented 	<ul style="list-style-type: none"> Heavy reliance on the existence of well-conducted primary studies which identify and assess theories of change Require a high level of reviewer expertise Can be time consuming (depending on the complexity of the intervention and scope of the existing literature) 	<p>Methods:</p> <ul style="list-style-type: none"> Desk research Interviews (to ascertain the completeness of the dataset) <p>Tools:</p> <ul style="list-style-type: none"> RAMESES I (reporting tool)
EMMIE-based review	<ul style="list-style-type: none"> Presents a comprehensive discussion of the utility of a particular intervention Provides information about the conditions under which an intervention is best implemented Provides information about potential challenges to the successful implementation of an initiative, including associated costs Promises to speak to all central concerns faced by a practitioner when assessing the (dis)advantages and feasibility of (alternative) interventions Can outline gaps in the literature which other reviews may not address (e.g. lack of cost-effectiveness analyses) 	<ul style="list-style-type: none"> Heavy reliance on the availability of a diverse set of primary studies Requires considerable expertise, time and resources 	<p>Methods:</p> <ul style="list-style-type: none"> Desk research Interviews Cost-effectiveness analysis <p>Tools:</p> <ul style="list-style-type: none"> EMMIE (assessment tool)

3.2 EVALUATION DESIGNS FOR PRIMARY STUDIES

The second section of Part II of this report provides an overview of evaluation designs for primary studies. It discusses randomized controlled trials, quasi-experimental designs, longitudinal designs and cross-sectional studies, case series and case-control designs, theory-based designs, stakeholder-oriented designs as well as economic evaluation designs. This overview is non-exhaustive and based on a subjective selection of relevant evaluation designs for primary studies. An alternative overview of designs can be found in the work of Hofman and Sutherland (2018) who have recently identified and systematized 24 different evaluation designs and methods.





3.2.1 RANDOMIZED CONTROLLED TRIALS

Randomized controlled trials (RCTs), as discussed in part I of this report, are widely seen as a cornerstone of the evidence-based movement. Not uncontroversially, they are often presented as a “gold standard” for – and the most rigorous type of – evaluation design (for a critical discussion of this label, see Bickman and Reich, 2009).

The primary feature of RCTs is the randomization process. In an RCT “the study population [e.g. people, schools, hospitals] is divided into an experimental group, members of which receive the treatment under test, and a control group, members of which receive something else [...] and that division is made by some random process” (Worrall, 2007: 982). This randomization process can follow several different paths. In a stratified randomization process, for instance, individual units are first grouped based on specific characteristics (e.g. financial situation) before they are randomly assigned to either the control or experimental group, which can help to ensure significant results for different sub-groups of the studied population (for an overview of randomization types useful for evaluations in P/CVE/DeRad, see Impact Europe, 2022b).

RCTs, while primarily considered a key method for understanding the effect size of a certain treatment or intervention, can serve different purposes. RCTs, for instance, if well-conceived, can gather information about and provide important insights into how the implementation of an initiative shapes observed effects. Consequently, RCTs are not necessarily wedded to impact evaluations, but can also provide a useful method for conducting process evaluations (Oakley et al., 2004). RCTs, moreover, do not necessarily draw on quantitative types of data collection. Instead, they may draw on both quantitative and qualitative methods to test the effectiveness of a treatment, or a particular theory of change (White, 2013). RCTs moreover do not necessarily involve just one experimental and one control groups (so-called ‘two-armed’ designs) but may use multiple experimental and control groups to test several treatments or conditions at once (so-called ‘multi-arm’ trials). RCTs, thus, should be viewed as a diverse set of experimental designs which are more versatile than is sometimes assumed.

The primary appeal of RCTs lies in their potential to reduce selection biases which are considered the main threat to the internal validity of an evaluation. While RCTs, in this regard, are considered a particularly rigorous type of evaluation, this, importantly, is not to say, however, that RCTs produce unbiased results. Indeed, RCTs, while mitigating selection bias, remain prone to several other biases (e.g. ascertainment bias, choice-of-question bias) which can reduce the internal validity of an evaluation (for an overview, see Jadad and Enkin, 2007; Bickerman and Reich, 2009). RCTs, moreover, face several other **limitations**. They can be expensive to conduct which, in the field of medicine, has led to a situation where RCTs are often funded by and showing results biased in favor of vested interests (Every-Palmer and Howick, 2014). They also often require more time, resources and planning than alternative evaluation methods, and they may appear unethical when randomization implies the denial of potentially highly beneficial treatments to particular units. RCTs, finally, while addressing the problem of internal validity may be less effective than other methods in addressing the external validity (the generalizability) of evaluation results.

RCTs, thus, are no panacea which provide the best evaluation method for each situation. At the same time, its limitations, as shown by White (2013), however, should not be overstated either. RCTs, if planned and conducted well can be an effective and efficient way to find out if and how an intervention works. Ethical concerns may be mitigated by careful randomization techniques, such as waiting list designs (Impact Europe, 2022b). Biases other than selection biases can be mitigated through an adherence to high implementation and reporting standards, such as those outlined in the Consolidated Standards of Reporting Trials (CONSORT) statement (Plint et al., 2006). RCTs, thus, while not without constraints or weaknesses, are often more applicable than their critics may suggest, and should thus be carefully considered for the evaluation of any type of initiative.





3.2.2 NATURAL EXPERIMENTAL STUDIES AND QUASI-EXPERIMENTAL STUDIES

Natural experimental studies (NESs) and **quasi-experimental studies** (QESs) are widely viewed as a valuable alternative to RCTs, particularly in situations where randomization appears unethical, impractical or very costly. NESs and QESs refer to investigations which approximate experimental conditions by creating treatment and control groups through means other than random assignment. These studies, which encompass a diverse set of evaluation designs, can, if well-implemented, reach a high level of internal validity and effectively control against confounding variables⁴ (Macejewski, 2020).

While NESs and QESs are sometimes used synonymously, NESs traditionally refer to evaluations which investigate the effects of naturally occurring 'quasi-random' assignments of an intervention (De Vocht et al., 2021: 2). Natural randomization, for instance, may take place if state interventions, as during the 2008 expansion of Medicaid in the US state of Oregon, are rolled out through a lottery-based system, which allows researchers to study the effects of an intervention as if it had been randomly assigned to an experimental and control group. QESs, by contrast, traditionally refer to evaluation designs in which the evaluator has some control or influence over the assignment of the intervention to individual units (De Vocht et al., 2021: 2). NESs and QESs, while referring to different ideal types of evaluation designs, are, in practice, however, often difficult to distinguish.

Researchers often, though not exclusively, employ NESs and QESs to establish the effect size of a particular intervention. NESs and QESs, thus, like RCTs, are most often associated with impact evaluations. To establish the effect of a particular initiative, QESs and NESs draw on and combine different techniques which are highly diverse and difficult to capture in a consolidated typology (Shadish and Luellen, 2006). Two of the most notable techniques include regression discontinuity designs and matched group designs which establish conditions close to those of an experimental design by creating a control and an experimental group. In regression discontinuity designs, the researcher investigates the effects of an intervention by comparing individuals who are positioned just above and below the threshold for receiving the intervention (e.g. a threshold based on income), whereby the former is taken as the experimental and the latter as the control group (White and Sabarwal, 2014: 7-9). In matched group designs, which can take the form of case-control and cohort studies, the evaluator examines an intervention's impact by comparing a group which received the intervention (experimental group) to a 'matched' control group (established ex post or ex ante) with similar characteristics (e.g. similar age group, income group) (White and Sabarwal, 2014: 2-6).

Such designs can have several **advantages**. Notably, they often present a more ethical, practically feasible and cost-effective alternative to RCTs (White and Sabarwal, 2013). QESs and NESs, moreover, can provide ex-post alternatives to RCTs (which necessarily must be developed ex-ante). Finally, QESs and NESs often provide a higher level of external validity than RCTs insofar as they are typically used to analyse interventions under "real world" rather than laboratory conditions. At the same time, NESs and QESs, however, also have several **limitations**. White and Sabarwal (2013: 11), for instance, highlight that, "as [QESs and NESs] "are based on certain assumptions, conclusion made about causality based on these studies designs are less definitive than those elicited by a well conducted randomized controlled trial (RCT)". Moreover, QESs and NESs, like RCTs, often demand careful planning, significant resources, as well the availability of data (e.g. for the creation of matching control groups). QESs and NESs, thus, while often more practical than RCTs, are not necessarily a less expensive or less time-consuming alternative.

⁴ Confounding variables refer to variables which produce distorted associations between the variables under study.





3.2.3 LONGITUDINAL AND CROSS-SECTIONAL DESIGNS

Longitudinal and cross-sectional designs can provide useful alternatives to RCTs and (the above-discussed types of) QESs where the creation of control groups (e.g. for financial or logistical reasons) appears impractical or unfeasible (Schmidt and Teti, 2004).

In longitudinal evaluations, the evaluator studies the process or effects of an intervention by following an individual or group of people over time. Two notable types of longitudinal designs, which are also considered variants of a QES, are one-group pretest-posttest designs and interrupted time-series designs. In a one-group pretest-posttest design, the evaluator tests the effectiveness of an intervention by analysing a group before and after the intervention takes place (Shadish and Luellen, 2006: 544-545). In an interrupted time-series design, the effects of an intervention are measured by testing the same group or individual regularly and repeatedly before, during and/or after the intervention (Shadish and Luellen, 2006: 546-547).

If implemented well (drawing on a sufficient sample size, guided by a clear research question, etc.) longitudinal designs can provide valuable insights into the functioning, effects and unintended consequences of an initiative. At the same time, their explanatory power, however, can be compromised by several biases, including selection biases, biases resulting from sample attrition or biases emerging from the Hawthorne effect which denotes an individual's change of behaviour in response to its awareness of being observed (for an overview, see Schmidt and Teti, 2004: 8-11). Their ability to convincingly show causal relationships, in this regard, is limited if compared to RCTs and (other types of) QESs. Longitudinal designs, however, may serve to identify and suggest causal links which can be further studied and verified by more robust evaluation designs.

Where the implementation of longitudinal designs, for logistical or financial reasons is not feasible, cross-sectional studies can be a useful alternative (Schmidt and Teti, 2004: 5-7). Cross-sectional designs, which are most often implemented through a survey, collect data on a particular group at one particular point in time, creating a "snapshot" (Wang and Cheng, 2020). This "snapshot" may provide important information about an intervention, including its operation, reception and potential effects (on a specific sub-population) (Kesmodel, 2018). While ill-equipped to establish causal links between initiatives and their effects, cross-sectional studies can identify correlations between the initiative and potential outcomes (as well as the prevalence of such outcomes) which can be investigated further by RCTs or QESs. The primary appeal of cross-sectional studies lies in their relatively quick and inexpensive implementation (Wang and Cheng, 2020: 67).

Cross-sectional studies offer a useful low-threshold design which can also be employed to prepare the ground for (or to follow up on) more elaborate studies. They, for instance, can be incorporated into the planning phase of an RCT, where they can serve to identify relevant research questions and hypotheses. A cross-sectional design, finally, can be turned into a longitudinal design if it is regularly and repeatedly applied to the same intervention (Lee and Niemeier, 1996). Such repeated cross-sectional (RCS) designs, also referred to as pseudo-longitudinal designs, can be a cost-effective alternative to longitudinal surveys with the disadvantage that they do not follow the same study population, and thus do not allow for the evaluator to draw conclusions from the trajectory of individual units (ibid.).

3.2.4 CASE SERIES, CASE REPORTS AND CASE-CONTROL DESIGNS

Next to cross-sectional studies, **case series** offer another low-threshold evaluation design which can present an alternative to or usefully support (quasi-)experimental studies. Case series typically refer to a set of **case reports** which collect in-depth information (e.g. via semi-structured interviews) on individual units who have either been exposed to the same intervention or who show similar outcomes (Parente et al., 2010). Both case series and individual case reports share similar **limitations**, including a small sample size and the lack of a control group which





constrain their ability to establish the effects of an intervention (Kooistra et al, 2009). They, moreover, are prone to several biases, including selection, information or publication bias (Garcia-Doval et al., 2018). Consequently, they are often placed at or near the bottom of the evidence pyramid (Ansaloni et al., 2007).

Despite their limitations, however, case series and case reports have important functions and **advantages**. Most notably, case series and case reports, due to their relative cost-effectiveness and ability to deliver in-depth information, are uniquely suited to detect and signal unexpected (including rare) effects of an intervention (Vandenbroucke, 2001). While case reports and case series cannot easily infer causality, they, as "the first line of evidence" can serve to generate hypotheses which can be further tested through (quasi-)experimental studies such as case-control studies (Jenicek, 1999: 117). In a case-control design, a group of cases is sampled based on a shared outcome and compared to a matched control group. The collection and comparison of data for both the case and control group can serve to explain differences in outcomes between the two groups (Coggon et al., 2009).

Case-control studies have several **advantages**. Notably, they can provide a cost-effective and less time-consuming ex-post alternative to other (quasi-)experimental and longitudinal designs which is particularly efficient in situations where the rate of a particular outcome is low, or where the time lapse between an intervention and its effects is long (Coggon et al., 2009; Schulz and Grimes, 2002). Case-control studies, however, also have disadvantages. Notably, as stressed by Schulz and Grimes (2002: 432), they are faced with methodological challenges pertaining to the identification of a suitable control group and the collection of data about the exposure history of members of the case and control group which can make them difficult to implement and vulnerable to biases.

3.2.5 THEORY-BASED DESIGNS

Theory-based designs describe a wide range of evaluation designs which focus on the identification and/or testing of theories which (implicitly) underpin an intervention. Theory-based designs generally emphasize a multi-method approach to evaluation which may draw on the above-noted designs (e.g. quasi-experimental or cross-sectional studies) for the purpose of empirical theory-testing.

Notable theory-based designs include logic models, theories of change, contribution analyses, policy-scientific evaluations, and realist evaluations. **Logic models** and theories of change are both participatory approaches which aim to create a stakeholder consensus about the functioning and objective of an intervention. While they are often seen as synonymous, they are best understood as separate (and complementary) evaluation designs. Logic models typically define and graphically illustrate the key components of an intervention – displayed as inputs, activities, outputs and outcomes – and can be a powerful tool for stakeholders to identify and articulate their intervention's objectives (Clark and Anderson, 2004; Dhillon and Vaca, 2018).

A 'theory of change' approach goes one step further and aims to create a consensus among stakeholders about the expected causal mechanisms of an intervention (Kabongo et al., 2020: 282). Following Dhillon and Vaca (2018: 66-70), the **strength** of a theory of change lies in its ability 1) to get stakeholders to reflect on the purpose of an intervention at the design stage, 2) to facilitate the collection of accurate and timely data about an intervention during the implementation phase, and 3) to create a foundation for organizational learning and the evaluation of an intervention's impact. Altogether, logic models and theories of change have multiple advantages, which besides the above-noted strengths also include their relatively inexpensive implementation.

Both, at the same time, however, also face several **challenges**. Different groups of stakeholders may hold different views of the objectives and change mechanisms of an intervention. A consensus, thus, may sometimes be difficult to create (Hansen and Vedung, 2010). Developing





a logic model or theory of change, moreover, risks creating an overly simplistic or mechanistic understanding of an intervention which can be shaped by stakeholders' preconceived ideas of or their concerns about the requirements set by funding agencies. Finally, theories of change only provide hypotheses for (but no verification of) change mechanisms. Other designs are thus needed to establish whether identified theories of change correspond with the functioning of an intervention in practice.

Designs which promise to address these challenges include contribution analyses, policy-scientific evaluations and realist evaluations. **Contribution analyses and policy-scientific analyses** both aim to identify, verify and strengthen an intervention's theory of change, but follow different paths in pursuing their objective. Contribution analyses follow a six-step process in which the evaluator 1) sets out the causal problem, 2) develops a theory of change (including rival explanations), 3) gathers existing evidence on the identified theory of change, 4) assesses the validity of the identified theory of change, 5) seeks additional evidence and 6) revises or strengthens the identified theory of change, including its contribution story (Mayne, 2012). Policy-scientific analyses, by contrast, aim to 1) (re)construct an intervention's (implicit) theory of change through an analysis of policy documents and stakeholder interviews and 2) to verify the (re)constructed theory of change through a process which may include the empirical test of the theory's validity (Sim and Van Gorp, 2019: 118-119).

Realist evaluations aim to go beyond (and complement) both contribution analyses and policy-scientific analyses by aiming to provide a detailed understanding of 'how' and 'why' different aspects of an intervention lead to observed or intended outcomes (Kabongo et al., 2020: 282). Importantly, realist evaluations highlight the importance of context and aim to identify and verify the "active ingredients" of an intervention by conceptualizing the relationship between the context (C) within which the program is implemented, the generative mechanism of change (M) and the observed outcome (O) (Westhorp et al., 2011). To verify this relationship – a so-called context-mechanism-outcome configuration (CMOs) – realist evaluations draw on a host of both qualitative and quantitative data, which may also involve the results of RCTs (on the usage of the latter, see Warren et al., 2022).

All three of these evaluation designs have notable **advantages**. Contribution analyses and policy-scientific evaluations can serve to validate, refine or rethink theories of change. Realist evaluations, moreover, can help to specify the "active ingredient" of an intervention, to understand confusing outcome patterns, and to prepare the replication of an intervention in a different context (Westhorp et al., 2011: 11-12). These designs, however, also have notable **limitations**. They can be time-consuming to implement and may require significant resources. Contribution analyses, policy-scientific evaluations and realist evaluations, moreover, are all of a high complexity and require skilled and experienced evaluators.

3.2.6 STAKEHOLDER-ORIENTED DESIGNS

Stakeholder-oriented designs refer to a wide range of evaluation designs which can be used in combination with the designs outlined above. The primary **concern** of these designs is the appropriate inclusion of stakeholders into the evaluation process.

Stakeholder-oriented designs can be broadly divided into three categories: collaborative designs, participatory designs and empowerment designs (Fetterman et al, 2014). In a collaborative evaluation, the evaluator is in full control of all proceedings but consults or works closely together with specific stakeholders at each stage of the evaluation process. In participatory evaluations, the evaluator shares control with stakeholders and actively encourages stakeholders to take part in defining the evaluation's agenda and objectives, in collecting and analysing data, or in reporting and dissemination of evaluation findings. Finally, in an empowerment evaluation, evaluators give full control to stakeholders while acting as an advisor who offers technical expertise and experience, and who helps keep the evaluation process "on track, rigorous,





responsive, and relevant" (Fetterman et al. 2014: 145; see also Fetterman and Wandersman, 2007).

Stakeholder-oriented designs can have multiple **advantages**. The systematic and substantial involvement of stakeholders in the evaluation process can increase organizational buy-in and strengthen trust between evaluators and stakeholders which in turn can improve the quality and utility of evaluation results (particularly if evaluated data are highly sensitive) (Patton, 2008; Williams and Kleinman, 2014). Stakeholder-oriented designs, moreover, can strengthen capacity building, organizational learning and deep reflection, and foster stakeholders' "emotional and logistical sense of connection to an intervention" (Odera, 2021: 213).

At the same time, stakeholder-oriented designs, however, also have notable **disadvantages**. Most importantly, they often lack clear (methodological) guidelines for their implementation which, in turn, risks creating a mismatch between postulated theoretical principles and their implementation in practice (Miller and Campbell, 2006). While some guidelines have been developed to mitigate this problem, such as Patton's 17-step guide for the implementation of a utilization-focused evaluation (Patton, 2012), stakeholder-oriented designs generally remain less standardized than most of the designs discussed above. Furthermore, it should be noted that stakeholder-oriented designs provide no clear indication for how an evaluation (beyond the inclusion of stakeholders) should be performed. In principle, thus, they can be flexibly combined with the designs outlined above of which some, notably logic models and theories of change, naturally demand a high level of stakeholder engagement.

3.2.7 ECONOMIC EVALUATION DESIGNS

The final category of evaluation designs presented in this overview are **economic evaluation designs** which aim to understand an intervention's impact in relation to its costs. These designs, which notably include cost-effectiveness and cost-benefit analyses, can complement and be combined with other primary evaluation designs discussed above, such as RCTs or cross-sectional studies, to establish the (comparative) utility of an initiative.

Cost-effectiveness analyses, specifically, aim to identify and place a monetary value on the costs of an intervention which are then related to specified (quantifiable) measures of an intervention's effectiveness (Cellini and Kee, 2015: 493). Cost-benefit analyses go one step further by weighing an intervention's costs against the monetary value of its benefits. By subtracting costs from benefits, a cost-benefit analysis ultimately aims to establish the net benefits (or net costs) of an intervention, and by extension its (comparative) financial utility (Cellini and Kee, 2015: 493).

The implementation of both designs can be summarized as a ten-step process (Boardman et al., 2006). In performing a cost-benefit or cost-effectiveness analysis, the evaluator needs to 1) define the analytical objectives, 2) decide on whose costs and benefits should be recognized (e.g. societal costs/benefits, organizational costs/benefits), 3) identify and categorize costs and benefits, 4) project costs and benefits over the life of a programme, 5) monetize costs, 6) quantify benefits in terms of units of effectiveness (for a cost-effectiveness analysis) or monetize benefits (for a cost-benefit analysis), 7) discount costs and benefits to obtain present values, 8) compute a cost-effectiveness ratio or a net present value (for a cost-benefit analysis), 9) perform a sensitivity analysis and 10) make recommendations if appropriate (for a detailed overview, see Cellini and Kee, 2015).

Undertaking such a cost-effectiveness analysis or cost-benefit analysis can have several **advantages**. Notably, their results can allow for decision-makers to justify the costs of an intervention to their colleagues and constituents (Guerin, Martina and van Gorp, 2018). Cost-effectiveness and cost-benefit analyses, moreover, can be relatively inexpensive to implement if they can draw on a strong quantitative evidence base for an intervention. At the same time, they, however, also have notable limitations. Critically, the costs and benefits of an intervention can often neither be easily quantified nor monetized. Where the quantification of an intervention





is difficult (e.g. due to challenges related to the data collection process) cost-benefit and cost-effectiveness analyses, moreover, can be prohibitively expensive and time-consuming. Finally, cost-benefit or cost-effectiveness analyses risk misrepresenting the utility of an intervention if their focus is overly narrow (e.g. restricted to one quantifiable out of many possible indicators for an intervention's impact).

3.2.8 SUMMARY

Based on the preceding discussions, the table below provides a summary of the key strengths and weaknesses, as well as the associated methods and tools, of different evaluation designs for primary studies.

Table 2: Overview of evaluation designs for primary studies

Evaluation design	Strengths	Weaknesses	Associated methods and tools
Randomized controlled trial (RCT)	<ul style="list-style-type: none"> Often considered the most rigorous evaluation design High internal validity (if well-implemented) Results of multiple RCTs can be effectively synthesized in systematic reviews/ meta-analyses 	<ul style="list-style-type: none"> Requires careful planning (ex-ante) Can require significant time and resources Can create ethical challenges 	<p>Methods:</p> <ul style="list-style-type: none"> Surveys, data analysis (for measuring outcomes) <p>Tools:</p> <ul style="list-style-type: none"> CONSORT statement (reporting tool) JB1 Checklist for RCTs (appraisal tool)
Quasi-experimental designs	<ul style="list-style-type: none"> Can present a more ethical, practically feasible and cost-effective alternative to RCTs Can provide ex-post alternatives to RCTs High internal validity (though not as high as that of RCTs) 	<ul style="list-style-type: none"> Lower internal validity than RCTs, claims about causality are less definite Can be expensive and time-consuming 	<p>Methods:</p> <ul style="list-style-type: none"> Data analysis (for the creation of control groups) Surveys, data analysis (for measuring outcomes) <p>Tools:</p> <ul style="list-style-type: none"> TREND guidelines (reporting tool) JB1 Checklist for QESs (appraisal tool)
Longitudinal designs	<ul style="list-style-type: none"> Considered the best alternative to RCTs in situations where the creation of control groups is not feasible Can be particularly useful for studying the effects of an intervention on a particular subgroup 	<ul style="list-style-type: none"> Explanatory power can be compromised by several biases (selection bias, sample attrition biases, Hawthorne effect) 	<p>Methods:</p> <ul style="list-style-type: none"> Surveys, interviews, focus groups <p>Tools:</p> <ul style="list-style-type: none"> TREND guidelines (reporting tool) STROBE checklist (reporting tool)
Cross-sectional study	<ul style="list-style-type: none"> Inexpensive and fast implementation Can provide important information about the reception and potential effects (and their prevalence) of an intervention 	<ul style="list-style-type: none"> Cannot establish causal links between interventions and their effects 	<p>Methods:</p> <ul style="list-style-type: none"> Surveys <p>Tools:</p> <ul style="list-style-type: none"> TREND guidelines (reporting tool)





	<ul style="list-style-type: none"> Can be used to establish the utility of more elaborate designs, can support the development of an RCT in the planning phase 		<ul style="list-style-type: none"> STROBE checklist (reporting tool) JBI Checklist for Analytical Cross Sectional Studies (appraisal tool)
Case-control designs	<ul style="list-style-type: none"> Can provide a cost and time efficient ex-post alternative to other (quasi-)experimental and longitudinal designs Can be an effective means to investigate rare outcomes as well as outcomes which occur after a significant time lapse 	<ul style="list-style-type: none"> Face several methodological challenges related to the selection of control groups and gathering of data which can lead to biases 	<p>Methods:</p> <ul style="list-style-type: none"> Interviews Questionnaires Data analysis <p>Tools:</p> <ul style="list-style-type: none"> SIGN Case-Control Studies Checklist JBI Checklist for Case Control Studies (appraisal tool) STROBE checklist (reporting tool)
Case series/ case reports	<ul style="list-style-type: none"> Relatively inexpensive and fast implementation Can detect and signal unexpected (including rare) effects of an intervention Can serve the formulation of hypotheses which can be further tested through RCTs and QESs 	<ul style="list-style-type: none"> Small sample size and lack of a control group Not well-placed to infer causality between intervention and outcomes Prone to several biases, including selection, information, and publication bias 	<p>Methods:</p> <ul style="list-style-type: none"> Interviews Questionnaires Data analysis <p>Tools:</p> <ul style="list-style-type: none"> JBI Checklist for Case Series (appraisal tool)
Logic model & theory of change	<ul style="list-style-type: none"> Can serve to identify, articulate and create a stakeholder consensus on an intervention's key components, objectives and causal mechanisms Encourage stakeholder and engagement and critical reflection, can create the foundation for organizational learning Can facilitate the collection of useful and accurate data during an intervention's implementation 	<ul style="list-style-type: none"> Can only be used to create (but not to verify) hypotheses about the operation and effects of an intervention Risk of creating and an overly simplistic or mechanistic understanding of an intervention, risk of reflecting biases (e.g. resulting from stakeholders' preconceived ideas) 	<p>Tools</p> <ul style="list-style-type: none"> UNDAF Companion (quality assurance checklist)
Contribution analysis & policy-scientific analysis	<ul style="list-style-type: none"> Foster organizational learning Can serve to identify, verify, refine and rethink theories of change Can be combined with strong effect evaluation designs (e.g. RCTs, QESs) for a rigorous evaluation of theories of change 	<ul style="list-style-type: none"> Highly complex, requires a skilled and experienced evaluators Can be time-consuming and require significant resources 	<p>Methods:</p> <ul style="list-style-type: none"> Document analysis (esp. policy-scientific analysis) Interviews Theory of change
Realist evaluation	<ul style="list-style-type: none"> Can provide a deep understanding of the "active ingredients" of an intervention which can explain certain outcomes and unintended consequences 	<ul style="list-style-type: none"> Highly complex, requires a skilled and experienced evaluators Can be time-consuming and require significant resources 	<p>Methods:</p> <ul style="list-style-type: none"> Interviews Document analysis Theory of change <p>Tools:</p>





	<ul style="list-style-type: none"> Can help to explain confusing outcome patterns Can help to prepare the replication of an intervention in a different context 		<ul style="list-style-type: none"> RAMESES II (reporting tool)
Stakeholder-oriented designs	<ul style="list-style-type: none"> Can improve the results and quality of an evaluation by increasing organizational buy-in and strengthening trust between stakeholders and evaluators Can strengthen capacity building, as well as deep reflection and organizational learning 	<ul style="list-style-type: none"> Driven by a set of principles rather than clear methodological guidelines Need to be combined with other designs to generate insights into the process and impact of an intervention 	
Economic evaluation designs	<ul style="list-style-type: none"> Can help to identify the comparative utility of an intervention Can help decision-makers to justify the costs of an intervention Relatively easy and quick to implement if a strong evidence base exists which indicates the benefits or effectiveness of an intervention 	<ul style="list-style-type: none"> Not feasible if the effectiveness, benefits and costs of an intervention cannot be easily quantified/monetized Can be prohibitively expensive and time-consuming where the data collection process is not supported by an existing evidence base A narrow focus on a few (quantifiable) indicators risks misrepresenting an intervention's utility 	<p>Tools:</p> <ul style="list-style-type: none"> CHEERS (reporting tool) JBI Checklist for Economic Evaluations (appraisal tool)

3.3 EVALUATION DESIGNS IN THE FIELD OF PVE/CVE AND DE-RADICALISATION

The final section of part II of this report discusses **the uptake of the above-outlined evaluation designs in the field of P/CVE/DeRad**. The section first reviews the usage of meta-evaluation designs in P/CVE/DeRad before it analyses the prevalence of different evaluation designs for primary studies. For the latter task, the section draws on two recent reviews of evaluations in P/CVE/DeRad by Feddes and Galluci (2015/6) and Bellasio et al. (2018).

3.3.1 META-EVALUATION DESIGNS USED IN THE FIELD OF PVE/CVE AND DE-RADICALISATION

This section aims to provide a brief overview of the current landscape of meta-evaluations in the field of P/CVE/DeRad. As noted in several recent reviews, the number of empirical studies in P/CVE/DeRad has risen sharply over the past two decades, which in turn has created demand for updated reviews and syntheses of the existing state of the art (see, for instance, Lum et al., 2006; Gielen, 2017; Pistone et al., 2019). The growing number of empirical studies in P/CVE/DeRad has in particular generated attempts, promoted notably by the Campbell Collaboration's CVE programme, to capture and synthesize the state of the art through systematic reviews and meta-analyses. The table below provides an overview of systematic reviews in this field published by the Campbell Collaboration.





Table 3: Overview of systematic reviews published by the Campbell Collaboration in the field of PVE/CVE and De-radicalisation

Systematic review	Objective	Included designs	Included studies
Lum et al. (2006)	Review the effectiveness of counter-terrorism strategies	Evaluations of two or more units of analysis, comparing some with and without the counterterrorism intervention Evaluations which made some attempt to provide for controls within a statistical analysis Evaluations which conducted an interrupted time series or intervention analysis to indicate some temporal ordering of effects.	Total: 7 <ul style="list-style-type: none">Longitudinal interrupted time series design: 6
Mazerolle et al. (2020)	Review the effectiveness of police programmes that seek to promote community connectedness	Quantitative impact evaluations that utilize a randomized or quasi-experimental design with a comparison group that either did not receive the intervention, or that received "business-as-usual" policing, no intervention or an alternative intervention.	Total: 1 <ul style="list-style-type: none">QES (matched control group): 1
Carthy et al. (2020)	Review the effectiveness of counter-narratives	Experimental and quasi-experimental designs where at least one of the independent variables involved compares a counter-narrative to a control (or comparison) exposure	Total: 19 <ul style="list-style-type: none">RCT: 12Longitudinal pre-/posttest design: 5Longitudinal interrupted time series design: 2
Mazerolle et al. (2021)	O1: Review the effectiveness of multiagency programs with police as a partner O1: Review potential mechanisms, moderators, implementation factors, and economic considerations related to the intervention	O1: Impact evaluations using experimental or robust quasi-experimental designs O2 Empirical studies (reporting on primary quantitative or qualitative data or conducting secondary analysis of primary quantitative or qualitative data)	O1 Total: 5 <ul style="list-style-type: none">QES (matched control group or cross-sectional): 5 O2: Total: 26 <ul style="list-style-type: none">Mix (including cross-sectional studies, cost-benefit analyses)
Wolfowicz et al. (2021)	Identify what the putative risk and protective factors for different radicalisation outcomes are	Experimental, longitudinal, case-control and cross-sectional designs	Total: 127 (206 samples) <ul style="list-style-type: none">Cross-sectional study: 186 samplesLongitudinal designs: 9Case-control designs: 10
Sarma et al. (2022)	O1: Synthesize the prevalence rate of mental health difficulties in terrorist samples and O2: prevalence of mental health disorders pre-	Cross-sectional, cohort or case-control designs	Total: 73





	<p>dating involvement in terrorism</p> <p>O3: Synthesize the extent to which mental health difficulties are associated with terrorist involvement compared to non-terrorist samples</p>		
Windisch et al. (2022)	Assess the effects of online interventions to reduce online hate speech/ cyberhate	Randomized and rigorous quasi-experimental studies	Total: 2 <ul style="list-style-type: none"> RCTs: 2
Wolfowicz et al. (2022)	<p>O1: identify and synthesize the effects of different media-related risk factors at the individual level</p> <p>O2: identify the relative magnitudes of the effect sizes for the different risk factors</p> <p>O3: compare the effects between outcomes of cognitive and behavioural radicalisation</p>	Experimental and observational studies (cross-sectional, longitudinal and case-control designs)	Total: 53 <ul style="list-style-type: none"> RCTs: 4 Longitudinal designs: 2 Cross-sectional studies: 45
Zych and Nasaescu (2022)	<p>O1: identify family-related risk and protective factors for radicalisation</p> <p>O2: assess the impact of radicalisation on families</p> <p>O3: assess the effectiveness of family-based interventions against radicalisation</p>	<p>O1, O2: cross-sectional and preferably longitudinal studies</p> <p>O3: randomized controlled trials or robust quasi-experimental designs</p>	Total: 33 <ul style="list-style-type: none"> Longitudinal designs: 3 Cross-sectional studies: 30

As illustrated in table 3, the number of systematic reviews and meta-analyses published by the Campbell Collaboration in the field of P/CVE/DeRad has grown significantly since 2020. These reviews now cover a wide range of topics from the effectiveness of different interventions to the identification of risk and protective factors. While they promise to provide useful syntheses of existing research, a closer look at the published reviews reveals that they are often based on a very thin evidence base. Mazerolle et al.'s (2020) review of the effectiveness of police programmes to promote community connectedness, for instance, found just one study which met its inclusion criteria. A lack of RCTs and QESs, moreover, prevented Zych and Nasaescu (2022) from assessing the effectiveness of family-based interventions against radicalisation. The quality and potential of systematic reviews, as these examples illustrate, are thus often constrained by a lack of high-quality impact evaluations.

An important function of systematic reviews has thus far been the identification of research gaps in the field of P/CVE/DeRad – a function which has also been performed by scoping reviews (Pistone et al., 2019) and realist reviews (Gielen, 2017). Realist reviews, which have only recently been introduced to the field, have not yet been widely applied in P/CVE/DeRad. Like EMMMIE-based reviews, they have the potential to supplement and provide a suitable alternative to systematic reviews, at least insofar as they tend to draw on a wider set of evaluation designs to address a specific research question (see Gielen, 2017). The expansion of alternative types of reviews, in this regard, constitutes research avenue which promises to usefully complement the rising number of systematic reviews and meta-analyses in the field of P/CVE/DeRad.





3.3.2 EVALUATION DESIGNS FOR PRIMARY STUDIES USED IN THE FIELD OF PVE/CVE AND DE-RADICALISATION

This section aims to discuss the prevalence of the above-outlined evaluation designs for primary studies in P/CVE/DeRad. It draws on two reviews by Feddes and Galluci (2015/6) and Bellasio et al. (2018) which have systematically investigated the usage of different evaluation designs in this field. The study by Feddes and Galluci (2015/6) reviewed evaluations of programmes aimed at preventing radicalisation or de-radicalisation between 1990 and July 2014. Bellasio et al. (2018), in turn, analyzed evaluations of counterterrorism and P/CVE policies in the Netherlands and abroad between 2013 and 2017. Taken together, both studies provide useful indications of the development of evaluations between 1990 and 2017. Selected results of both studies are presented below in tables 4 and 5.

Table 4: Prevalence of evaluation designs in PVE/CVE and De-radicalisation between 1990 and June 2014 (based on Feddes and Galluci, 2015/6)

Evaluation design	Number of samples
Cross-sectional design	74
Longitudinal design	5
Quasi-experimental	3
Not specified	50
Mixed design	4
Total	135
Theory-based design	Number of samples
Theory of change included	16
Policy-scientific approach	34
Contribution analysis	1
Realist evaluation	1
No theory-based evaluation	81
Total	133

Table 5: Prevalence of evaluation designs in PVE/CVE and De-radicalisation between 2013 and 2017 (based on Bellasio et al., 2018)

Evaluation design	Number of samples
Additionality (one-time ex-post evaluation)	35
Quasi-experimental designs	8
Longitudinal designs	5
Unclear	3
Total	51
Theory-based design	Number of samples





Theory-based	10
No clear approach	33
Realist evaluation	3
Participatory evaluation	1
Transboundary evaluation	1
Total	48

The studies by Feddes and Galluci (2015/6) and Bellasio et al. (2018) together indicate interesting trends in the P/CVE/DeRad evaluation literature. They notably affirm the above-indicated scarcity of RCTs, QESs and longitudinal designs in the field of P/CVE/DeRad. Feddes and Galluci (2015/6) only identified three out of 135 reviewed evaluation samples as underpinned by a quasi-experimental design while five samples were considered longitudinal. Similarly, Bellasio et al. (2018) found that just eight out of 51 reviewed evaluation samples implemented a quasi-experimental design while five samples were described as following a longitudinal design. Neither review identified any existing RCTs. Both, at the same time, found that non-experimental evaluations conducted at one particular moment in an intervention's life cycle – coded as cross-sectional designs by Feddes and Galluci (2015/6) and additionality by Bellasio et al. (2018) – accounted for the majority of evaluation designs.

In terms of the evolving popularity of these designs, the studies suggest that the number of quasi-experimental and longitudinal designs, relative to cross-sectional studies, has slightly increased between the periods of 1990-2014 and 2013-2017 (see figure 4).

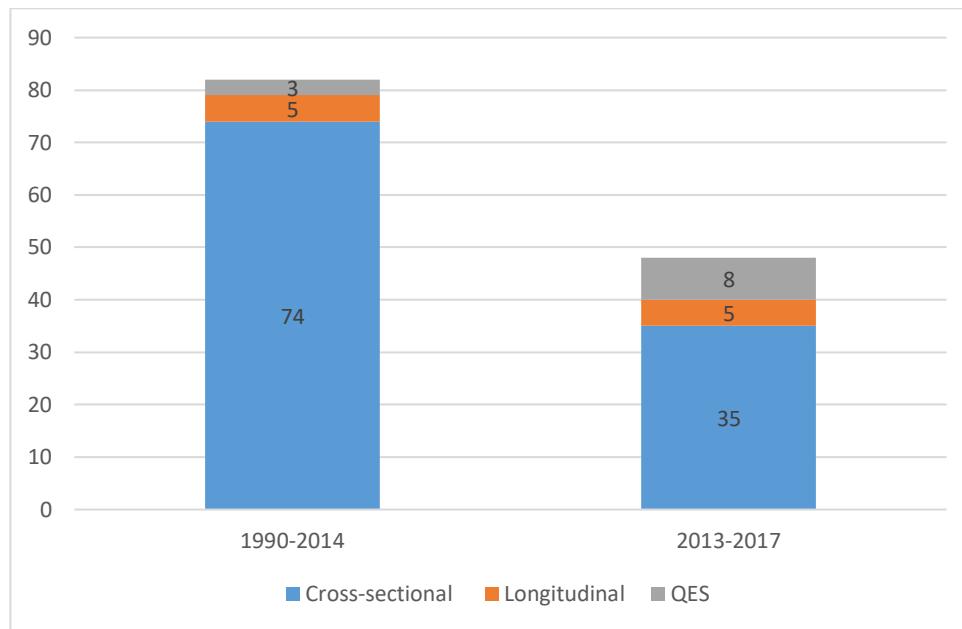


Figure 3: Prevalence of evaluation designs, 1990-2014 (based on Feddes and Galluci, 2015/6) and 2013-2017 (based on Bellasio et al., 2018)⁵

At the same time, a closer look at the coded list of evaluations provided by Bellasio et al. (2018), shows that the popularity of quasi-experimental and longitudinal designs did not notably increase

⁵ In figure 2, studies coded by Bellasio et al. (2018) as "additionality" are treated as cross-sectional studies





in the period of 2013-2017. As shown in figure 5, quasi-experimental and longitudinal studies conducted between 2013-2017 appear evenly distributed across the studied period, while the variation in the overall number of studies can primarily be attributed to an increase in one-time ex-post evaluations (additionality). The data provided by the two studies, thus, do not indicate a significant shift towards high-quality (impact) evaluation designs in the field of P/CVE/DeRad.

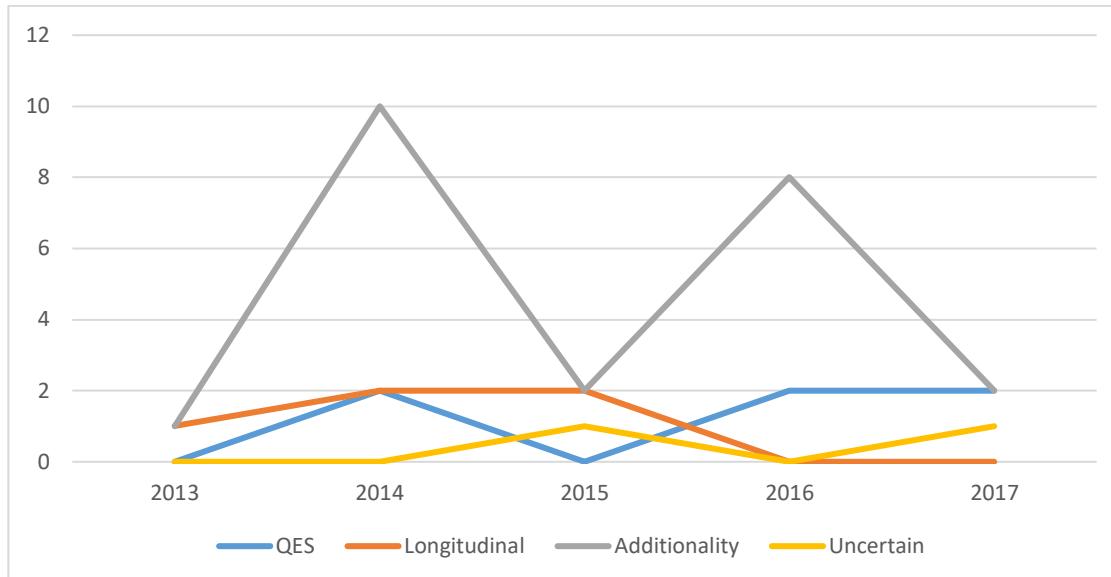


Figure 4: Number of evaluations by design, 2013-2017 (based on Bellasio et al., 2018)

The two studies, next to providing information on the prevalence of QESs, longitudinal designs and cross-sectional studies, moreover, offer an overview of the prevalence and trajectory of theory-based designs (see tables 4 and 5). Both studies suggest that only a minority of evaluations (though a sizable one) has followed a theory-based approach between 1990 and 2017. While Feddes and Galluci (2015/6) found that 52 out of 133 reviewed evaluation samples (39%) were theory-based, Bellasio et al. (2018) coded 13 out of 48 samples (27%) as following a theory-based approach or realist evaluation. These figures suggest that the usage of theory-based approaches did not markedly increase (and even slightly declined) from the period of 1990-2014 to the period of 2013-2017.

While the two reviewed studies provide important insights into the usage of evaluation designs in the field of P/CVE/DeRad, they also have notable limitations. Importantly, the two studies do not systematically address the extent to which evaluations in P/CVE/DeRad follow stakeholder-oriented designs or engage economic evaluations. They consequently do not provide information about all the design categories outlined above. Moreover, they do not cover developments since 2017 and thus cannot give insights into how the P/CVE/DeRad evaluation landscape has evolved in the past five years. Finally, the two studies reviewed here, while extensive, do not cover the entire spectrum of evaluations related to P/CVE/DeRad initiatives, as reflected by their non-inclusion of RCTs covered in the systematic reviews noted above (see table 3). Consequently, as also noted by Bellasio et al. (2018: 81) further research in this area is needed, including regular mapping and stocktaking exercises.



4 PART III: TOWARDS AN EVIDENCE-BASED APPROACH TO EVALUATION IN PVE/CVE AND DE-RADICALISATION

The final section of this report briefly summarizes the key findings of parts I and II before outlining an evidence-based approach to evaluation in P/CVE/DeRad. After outlining the key principles of EBE, this part of the report will discuss key challenges to the implementation of EBE and highlight the value of an EBE tool which will be developed in INDEED's WP3.

4.1 TAKING STOCK AND MOVING AHEAD

This report introduced the notion of EBP as a decision-making process which integrates 1) available external evidence, 2) professional expertise and 3) client values, preferences and circumstances⁶. **The report highlighted that a key challenge** for the realization of EBP lies in the creation of a strong evidence base which addresses the needs and concerns of (frontline) practitioners. This challenge is particularly acute in the field of P/CVE/DeRad where the implementation of EBP is hampered by the scarce application of robust (including theory-based and stakeholder-oriented) evaluation designs which prevent the creation of a strong evidence base.

Most evaluations in P/CVE/DeRad, as highlighted in part II of this report, have taken the form of a cross-sectional study in which the evaluator collects data from recipients (and other stakeholders) of an intervention at one specific point in time without the use of a control group. While such studies have their benefits, an over-reliance on cross-sectional designs, as discussed in the report, has left the field ill-equipped to establish the effects, efficient implementation and transferability of P/CVE/DeRad interventions. Strengthening the evidence base in P/CVE/DeRad, the reported notes in this regard, requires the expansion of alternative evaluation designs, notably longitudinal and (quasi-)experimental designs, as well as theory-driven and stakeholder-oriented designs, which currently remain in short supply as demonstrated in several recent (systematic) reviews.

To address this challenge, this report, in its final part, will **outline an evidence-based approach to evaluation** which aims to improve evaluation practices by encouraging the application of EBP to the evaluation process itself.

4.2 EVIDENCE-BASED EVALUATION: DEFINITION AND KEY PRINCIPLES

This report suggests that the principles of EBP be can usefully applied to the field of evaluation. Drawing on the concept of EBP, third report defines an evidence-based approach to evaluation as "a process of planning and implementing evaluations which integrates available external evidence, professional expertise and stakeholder values, preferences and circumstances". Following this definition, **an evidence-based approach to evaluation, akin to EBP, is based on three key principles:**

First, EBE commits evaluating practitioners to seek, appraise, engage with, and take into consideration available external evidence in the planning and implementation of an evaluation which may include research on the subject of the evaluation as well as on evaluation designs,

⁶ This definition advances the tentative conceptualization of EBP presented in the INDEED deliverable D1.1.



methods and tools. Important enablers of such a commitment are the effective presentation and communication of evidence within and between communities of research and practice.

Second, EBE commits evaluating practitioners to take client (and more broadly stakeholder) values, preferences and circumstances in the planning and implementation of an evaluation seriously. It, in this regard, requires evaluating practitioners to adopt a context-sensitive and participatory approach in which stakeholders can voice their preferences and concerns, and in which evaluation procedures and results are transparently presented and discussed.

Third, EBE commits evaluating practitioners to develop their professional expertise, and to build on their skills and professional judgement during the evaluation process to decide on a course of action while considering and integrating available external evidence and stakeholder values, preferences and circumstances.

Taken together, these three principles constitute the foundation for EBE as an ideal-typical approach to evaluation. This approach, on the one hand, can be contrasted with an opinion-based evaluation process which is driven by convention or intuition rather than a thorough engagement with the relevant research (e.g. on evaluation designs or the subject under investigation). On the other, it stands in contrast to a rigid evaluation process whose objectives and methods are pre-determined rather than developed in a participatory process in which stakeholder preferences, needs a values, and the context and characteristics of the intervention under investigation, are carefully taken into consideration. EBE, in this regard, can be graphically situated in a matrix of four ideal types (figure 6).

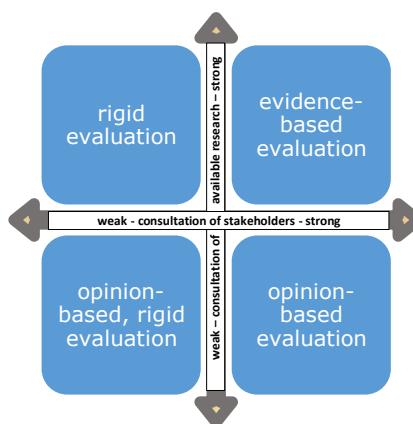


Figure 5: Evidence-based evaluation

4.3 EVIDENCE-BASED EVALUATION: WHAT IT IS AND WHAT IT ISN'T

As outlined in figure 6, EBE can be described as an ideal-typical approach to evaluation which stands in contrast to opinion-based and rigid evaluation processes. While EBE proposes a novel framework for the practice of evaluation, this framework, importantly should not be seen as necessarily opposing or standing in contrast to existing evaluation practices insofar as many evaluators may already apply (at least in part) EBE in their daily work. Nor should EBE be viewed as contradicting the insights of the evaluation literature which also, for instance, highlights the importance of stakeholder voices in the evaluation process.

Evaluation research, indeed, has long "demonstrate[d] that attention to and the involvement of key stakeholders enhances the design and implementation of evaluations and the use of evaluation results in decision making" (Bryson and Patton, 2015: 37). It has also cautioned that stakeholder involvement, if poorly implemented, can produce negative results, such as emerging biases and reduced equity (Brandon and Fukunaga, 2013: 39). EBE echoes these insights and calls on evaluators not to practice stakeholder involvement as an end in itself, but rather to





reflect and decide on the appropriate approach to stakeholder engagement through an evidence-based process which involves the integration of professional expertise, available external evidence (e.g. insights generated by previous/comparable evaluations and the academic literature on stakeholder inclusion) and the (initial) consideration and consultation of stakeholders relevant to the evaluated project. .

EBE's key contribution, in this regard, lies in its promotion of a decision-making process which integrates the three key principles noted above, rather than in promoting these principles individually. It is by promoting such decision-making that EBE, in turn, aims to address key challenges to project evaluation, such as gaps between evaluation research and practice (see, for instance, Khakee, 2003), the need for an ethical and inclusive evaluation processes (see, for instance, Barnett and Camfield, 2016), and the implementation of evaluations which are reflective of and useful for organizational learning (see Torres and Preskill, 2001). As such, EBE, constitutes an approach to evaluation with wide applicability which can be useful to disciplines well beyond the field of P/CVE and De-radicalisation. EBE, within the field of P/CVE/De-Rad, moreover can be widely applied to evaluations of different types of initiatives, including policies and strategies, long-term and short-term programmes as well as ad-hoc actions. It, naturally, is also applicable to initiatives dealing with different forms of extremism and paths to radicalisation.

It should further be highlighted, in this regard, that EBE, while promoting a robust planning and implementation of evaluation, is not wedded to any particular evaluation method or design. EBE neither promotes nor is especially suited for specific evaluation types, such as impact, process, or formative evaluations. Rather, it should be seen as an approach which can usefully guide the preparation, planning, implementation and utilization of each of these types. EBE, moreover, should not be viewed as a framework which uncritically promotes the use of (quasi)experimental evaluation designs but rather as a framework which emphasizes the need for evaluators to reflect on the (dis)advantages of different evaluation designs, through the integration of professional expertise, external evidence (e.g. relevant research on the utility of individual designs) and the consideration of stakeholder values, needs and circumstances.

4.4 THE RELATIONSHIP BETWEEN EVIDENCE-BASED PRACTICE AND EVIDENCE-BASED EVALUATION

As noted above, the concept of EBE is closely tied to the notion of EBP. Their relationship, however, is not unidirectional. Rather, **the relationship between EBE and EBP can be viewed as interdependent.**

On the one hand, EBE draws on the principles of EBP, and applies them to the field of evaluation. On the other hand, the practice of EBE serves to encourage the uptake of EBP through the promotion of robust evaluation practices. EBE, specifically, seeks to promote EBP by encouraging practitioners to carefully reflect on, and to draw on available research about suitable evaluation designs which, in turn, facilitates the implementation of robust evaluation designs. EBE, moreover, aims to encourage stakeholder engagement throughout the evaluation process which promises to not only create an evidence base which speaks more clearly to practitioner needs, but also to generate a process which gets stakeholders interested in and used to the application of evidence-based principles. EBE, in this regard, can function as an entry point for practitioners to learn about and to gain experience in applying (the underlying principles of) EBP. EBE, against this backdrop, can be understood as both an application of and a potential catalyst for EBP.





4.5 CHALLENGES TO THE IMPLEMENTATION OF EBE

While EBE promises several benefits, including a more rigorous and inclusive approach to evaluation, its implementation, especially in the field of P/CVE/De-Rad, is not without challenges.

Perhaps most significantly, the practice of EBE in P/CVE/De-Rad is challenged by the field's underdeveloped evidence base which can compromise the ability of evaluators to conduct relevant background research (see section 2.3.2 of this report). Relatedly, the practice of EBE is further challenged by the scarcity of (updated) systematic reviews and primary studies in evaluation research which poses a challenge for evaluators who seek to conduct background research on the efficacy of specific evaluation methods and designs. Brandon and Fukunaga (2014: 39), in this regard, have long noted "the lack of empirical research on evaluation in general – and more accurately, perhaps, the lack of funding for such research". For EBE to fulfill its full potential, greater investment is thus needed into both studies of P/CVE/De-Rad initiatives and research on evaluation practices.

This, however, is not to say, that an underdeveloped evidence base precludes the practice of EBE. Existing studies indeed demonstrate that rigorous background research is both possible and beneficial for the evaluation of P/CVE/De-Rad initiatives. Cops, Pauwels and Van Alstein (2020), for instance, systematically evaluated the adequacy of the Flemish Action Plan to Prevent Violent Radicalisation and Polarisation by assessing the extent to which its individual components are supported by existing research. The evaluation, drawing on the research literature, has formulated several recommendations for improving both the Action Plan's itself and its future assessment (Cops, Pauwels and Van Alstein, 2020: 340-1). The existing literature on P/CVE/De-Rad and related fields, as this evaluation shows, can thus usefully support and be systematically integrated into the process of evaluating P/CVE/De-Rad initiatives.

Next to an often-slim evidence base, another notable challenge to the implementation of EBE lies in the often-poor planning and sensitivity of evaluations in the field of P/CVE/De-Rad. As discussed in section 2.3.3., the time and resources allocated to evaluation processes is often scarce. This presents a severe challenge to the implementation of EBE which – due to its promotion of rigorous background research and the careful consideration of stakeholder needs, values and circumstances – can be more time and resource-consuming than its alternatives. Additionally, the implementation of EBE is further challenged by the sensitivity of data in the field of P/CVE/De-Rad which can hinder the full and transparent inclusion of key stakeholders in the evaluation process.

While these challenges are significant, they do, however, not preclude a comprehensive consideration of stakeholder needs, values and circumstances in the evaluation process, and thus the practice of EBE. A recent evaluation of the UK program "Prevent", for instance, conducted over 170 meetings and engagement events attended by over 1,200 people, bringing together and including a wide range of key stakeholders in the review process (Shawcross, 2023: 163-5). Information gathered through these engagement activities, in turn, was further integrated with data collected through a thorough review of existing research (Shawcross, 2023: 164). The review of Prevent, thus, carries the marks of EBE and shows how EBE, despite existing challenges, can be practiced in the field of P/CVE/De-Rad. To expand such practices, and to strengthen them further, will require investment into the planning and funding of evaluations to allow for rigorous reviews of external evidence and the comprehensive inclusion of relevant stakeholders in all stages of the evaluation process.

4.6 TOWARDS AN EVIDENCE-BASED EVALUATION TOOL

Drawing on the definition and discussion of EBE in this report, WP3 of INDEED will develop a more detailed guide to the implementation of EBE across different stages of the evaluation process. Specifically, WP3 will create a web-based online tool, supported by two e-guidebooks, which will outline how an EBE approach can guide the preparation, planning, implementation





and utilization of an evaluation in the field of P/CVE/De-Rad. To demonstrate the value of this task, this section will provide a brief overview of – and discuss how an EBE tool can draw on and complement – existing evaluation tools.

Many evaluation tools already exist in the field of P/CVE/De-Rad, as well as related areas such as crime prevention. The table below (table 6) presents an overview of these tools and the (primary) institution leading their development.

Table 6: Evaluation tools in the fields of P/CVE/De-Rad and Crime Prevention

Name of the evaluation tool	Created by
IMPACT Europe: An evaluation Toolkit for professionals working in the counter violent extremism field	EU FP7 Project
EMMASCAN	EU ISF Project
Qualiprev	EUCPN
PVE Toolkit	UNDP
A self-assessment online tool for evaluation diagnostics and strategizing	Independent Evaluation Office of UNDP
DEval Evaluation Standards	German Institute for Development Evaluation (DEval)
Countering Violent Extremism Evaluation Tool	New South Wales Government, Australia
Evaluate your CVE Results	Hedayah
Measuring up: Evaluating the Impact of P/CVE Programmes	United States Institute of Peace
Toolbox 3: Evaluation of Crime Prevention Initiatives	EUCPN
Crime Prevention Evaluation Toolkit	Australian Institute of Criminology and Community
The Policing Evaluation Toolkit	College of Policing, UK
Monitoring, Evaluation and Learning Toolkit To Support Action Plans to Prevent and Counter Violent Extremism	United Nations Office of Counter-terrorism

While many tools exist in the field of P/CVE/De-Rad, they significantly differ in their purpose, form and function. Some tools, such as EMMASCAN or Qualiprev, primarily aim to facilitate a quick institutional self-assessment. Others, in turn, revolve around a repository of evaluation indicators (Countering Violent Extremism Evaluation Tool) or the setting of basic evaluation standards (DEval Evaluation Standards). Most tools, however, focus on giving guidance and recommendations for the planning and implementation of an evaluation in P/CVE/De-Rad, both through web-based applications and detailed downloadable guidebooks (see, for instance, UNDP PVE Toolkit, the EUCPN Toolbox, the Policing Evaluation Toolkit or the UNCCT Monitoring Evaluation and Learning Toolkit).





Such guidelines often highlight the significance of elements central to EBE. Most evaluation tools stress, at least in passing, the significance of stakeholder engagement or background research (see, for instance, the Policing Evaluation Toolkit, the UNCCT Monitoring, Evaluation and Learning Toolkit, or Hedayah's "Evaluate your CVE Results"). Some tools, such as the EUCPN toolbox, further describe the importance of stakeholder involvement in different evaluation stages, including during the reporting and communication of evaluation findings. Going a step further, the UNDP PVE Toolkit, moreover, presents what can be considered an evidence-based approach to the development of a theory of change (as an early step in the evaluation process) which highlights the need to integrate external evidence and stakeholder inclusion.

While elements of EBE are reflected in existing evaluation tools in P/CVE/De-Rad, an EBE approach, however, has yet to be systematically developed across all stages of the evaluation process. The INDEED tool aims to address this gap by detailing how an EBE approach can be applied across the planning, preparation, implementation, and utilization stage of an evaluation. By systematically outlining an EBE approach across these stages, the INDEED tool can both complement and draw on existing tools. In particular, it can build on and create links to insights provided by other tools on stakeholder involvement as well as the repositories (e.g. of evaluation indicators or evaluation methods) offered by existing tools which can serve as relevant resources for the implementation of EBE.





5 CONCLUSION

This report has provided reviews of the literature on evidence-based practice and the literature on evaluation designs and methods. Based on these reviews, the report introduced and discussed the concept of evidence-based evaluation which lies at the heart of the INDEED project.

The first part of the report discussed the notion of EBP which was defined as a decision-making process which integrates 1) available external evidence, 2) professional expertise and 3) client values, preferences and circumstances⁷. **The report highlighted that a key challenge** for the realization of EBP lies in the creation of a strong evidence base which addresses the needs and concerns of (frontline) practitioners. Additional challenges include the integration of EBP in education and training programmes, the creation of workplace and political environments conducive to EBP, as well as the effective presentation and communication of evidence within and between communities of research and practice. All these challenges, the report stressed, apply to the field of the P/CVE/DeRad where the implementation of EBP is particularly hampered by the scarce application of robust (including theory-based and stakeholder-oriented) evaluation designs which prevent the creation of a strong evidence base.

This challenge was further detailed in part II of the report which provided an **extensive overview of evaluation designs and their usage in the field of P/CVE/DeRad**. The report highlighted that most primary evaluations in this field take the form of a cross-sectional study in which the evaluator collects data from recipients (and other stakeholders) of an intervention at one specific point in time without the use of a control group. While such studies have their benefits, an over-reliance on cross-sectional designs, as discussed in the report, leaves the field ill-equipped to establish the effects, efficient implementation and transferability of P/CVE/DeRad interventions. Strengthening the evidence base in P/CVE/DeRad, the reported notes in this regard, requires the expansion of alternative evaluation designs, notably longitudinal and (quasi-)experimental designs, as well as theory-driven and stakeholder-oriented designs, which currently remain in short supply as demonstrated in several recent (systematic) reviews.

The report altogether suggests that the realization of EBP in the field of P/CVE/DeRad above all requires measures to facilitate the usage of robust evaluation practices. To this end, the report, in its final part, introduced the concept of EBE which was defined as "a process of planning and implementing evaluations which integrates available external evidence, professional expertise and stakeholder values, preferences and circumstances". The report has highlighted challenges to the implementation of EBE which include an often-underdeveloped evidence base, the frequent lack of sufficient time and resources for conducting evaluations as well as the sensitivity of evaluations which create challenges for the implementation of participatory evaluation practices. The report, finally, stressed the added value of an EBE tool which can guide the planning and implementation of EBE in the field of P/CVE/De-Rad. This tool will be developed as part of WP3 in the next phase of the INDEED project.

⁷ This definition advances the tentative conceptualization of EBP presented in the INDEED deliverable D1.1.



6 REFERENCE LIST

Ahmadi, S., Baradaran, H. R., Ahmadi, E. (2015) 'Effectiveness of teaching evidence-based medicine to undergraduate medical students: a BEME systematic review'. *Medical Teacher*, 37 (1): 21-30.

Aiello, E., Puigvert, L. and Schubert, T. (2018) 'Preventing violent radicalization of youth through dialogic evidence-based policies'. *International Sociology*, 33 (4): 435-453.

Ansaloni, L., Catena, F. and Moore, E. E. (2007) 'WJES and case reports/ case series'. *World Journal of Emergency Surgery*, 2(1): 1-4.

Arksey, H. and O'Malley, L. (2005) 'Scoping studies: towards a methodological framework'. *International Journal of Social Research Methodology*, 8 (1): 19-32.

Arya, S., Kaji, A. H., and Boermeester, M. A. (2021). 'PRISMA Reporting Guidelines for Meta-analyses and Systematic Reviews'. *JAMA Surgery*, 156 (8): 789-790.

Barnett, C. and Camfield, L. (2016) Ethics in evaluation. *Journal of Development Effectiveness*, 8 (4): 528-534.

Baruch, B., Ling, T., Warnes, R., Hofman, J. (2018) 'Evaluation in an emerging field: Developing a measurement framework for the field of counter-violent-extremism'. *Evaluation*, 24 (4): 475-495.

Beerkens, M. (2018) 'Evidence-based policy and higher education quality assurance: progress, pitfalls and promise'. *European Journal of Higher Education*, 8 (3): 272-287.

Bellasio, J., Hofman, J., Ward, A., Nederveen, F., Knack, A., Meranto, A. S., and Hoorens, S. (2018) 'Counterterrorism evaluation: Taking stock and look ahead'. *RAND Europe research report*. Available at: https://www.rand.org/pubs/research_reports/RR2628.html.

Belur, J., Thornton, A., Tompson, L., Manning, M., Sidebottom, A. and Bowers, K. (2020) 'A systematic review of the effectiveness of the electronic monitoring of offenders'. *Journal of Criminal Justice*, 68: 1-18.

Berg, R. C. and Nanavati, J. (2016) 'Realist review: Current practice and future prospects'. *Journal of Research Practice*, 12 (1): 1-28.

Bickman, L. and Reich, S. M. (2009) 'Randomized Controlled Trials: A Gold Standard With Feet of Clay?' In Donaldson, S. I., Christie, C. A. and Mark, M. M. (eds) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* (Thousand Oaks, CA: SAGE), 51-77.

Blaustein, J. (2016) 'Exporting criminological innovation abroad: Discursive representation, 'evidence-based crime prevention' and the post-neoliberal development agenda in Latin America'. *Theoretical Criminology*, 20 (2): 165-184.

Boardman, A. A., Greenberg, D. H., Vining, A. R., and Weiner, D. L. (2006) *Cost-Benefit Analysis: Concepts and Practice*, 3rd edition (Upper Saddle River, N. J.: Prentice Hall).

Brandon, P. R. and Fukunaga, L. L. (2014) 'The State of the Empirical Research Literature on Stakeholder Involvement in Program Evaluation' *American Journal of Evaluation*, 35(1): 26-44.

Brownson, R. C., Fielding, J. E., Maylahn, C. M. (2009) 'Evidence-Based Public Health: A Fundamental Concept for Public Health Practice'. *Annual Review of Public Health*, 30: 175-201.

Bryson, J. M. and Quinn Patton, M. (2015) "Analyzing and Engaging Stakeholders" In: Newcomer, K. E., Hatry, H. P. and Wholey, J. S. (eds) *Handbook of practical program evaluation* (Wiley: Hoboken, N.J.), pp. 36-61.

Brownson, R. C., Fielding, J. E., Green, L. W. (2017) 'Reconciling the Pulls of Practice and the Push of Research'. *Annual Review of Public Health*, 39: 3.1-3.27.

Burns, P. B., Rohrich, R. J. and Chung, K. C. (2011) 'The Levels of Evidence and Their Role in Evidence-Based Medicine. *Plastic and reconstructive surgery*, 128(1): 305-310.

Campbell Collaboration (2020). "Counter-narrative interventions may affect some risk factors related to violent radicalization, but there is no effect on intent to act violently". Available at: https://www.campbellcollaboration.org/media/k2/attachments/0362_CJCG_Carthy_Counter-narratives_radicalisation_PLS_EN.pdf.





Campbell Collaboration (2022a) 'About the CVE Programme'. Available at: <https://www.campbellcollaboration.org/about-the-cve-programme.html>.

Campbell Collaboration (2022b) 'Campbell CVE Studies'. Available at: <https://www.campbellcollaboration.org/research-on-cve.html>.

Campbell Collaboration (2022c) 'Methodological Expectations of Campbell Collaboration Intervention Reviews (MECCIR)'. Available at: <https://www.campbellcollaboration.org/meccir.html>.

Carthy, S. L., Doody, C. B., Cox, K., O'Hora, D. and Sarma, K. M. (2020) 'Counter-narratives for the prevention of violent radicalization: A systematic review of targeted interventions'. *Campbell Systematic Reviews*, 16: e1106.

Cellini, S. R. and Kee, J. E. (2015) 'Cost-effectiveness and cost-benefit analysis'. In Newcomer, K. E., Hatry, H. P. and Wholey, J. S. (eds) *Handbook of Practical Program Evaluation, Fourth* (Hoboken, N.J.: John Wiley & Sons, Inc.), pp. 636-672.

Claridge, J. A. and Fabian, T. C. (2005) 'History and Development of Evidence-based Medicine'. *World Journal of Surgery*, 29 (5): 547-553.

Clark, H. and Anderson, A. A. (2004) 'Theories of change and logic models: Telling them apart'. *American Evaluation Association Conference*. Available at: https://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/TOCs_and_Logic_Models_forAEA.pdf.

Clear, T. R. (2009) 'Policy and evidence: the challenge to the American society of criminology: 2009 Presidential Address to the American Society of Criminology'. *Criminology*, 48 (1): 1-25.

Clément, P-A., Madriaza, P. and Morin, D. (2021). "Constraints and opportunities in evaluating programs for prevention of violent extremism: how the practitioners see it". UNESCO Chair in Prevention of Radicalization and Violent Extremism, available at: https://chaireunesco-prev.ca/wp-content/uploads/2021/07/FINAL_UncescoPrev_Rapport_PREV-IMPACT_EN.pdf.

Coggon, D., Barker, D. and Rose, G. (2009) *Epidemiology for the Uninitiated* (John Wiley & Sons), 46-50.

Collins, J. A. and Fauser, C. J. M. (2005) 'Balancing the strengths of systematic and narrative reviews'. *Human Reproduction Update*, 11 (2): 103-104.

Cops, D., Pauwels, A. and Van Alstein, M. (eds) (2020) *Gewelddadige radicalisering & polarisering. beleid & preventie in Vlaanderen: evaluatie en uitdagingen*. Vlaams Vredesinstituut. Available at: <https://vlaamsvredesinstituut.eu/rapport/gewelddadige-radicalisering-en-polarisering-beleid-preventie-in-vlaanderen-evaluatie-en-uitdagingen/>

Cowen, N. (2019) 'For whom does 'what works' work? The political economy of evidence-based education'. *Educational Research and Evaluation*, 25 (1): 81-98.

Davies, P. (1999) 'What is evidence-based education?'. *British Journal of Educational Studies*, 47 (2): 108-122.

Dekkers, M., Egger, M., Altman, D. G., Vandenbroucke, J. P. (2012) 'Distinguishing case series from cohort studies'. *Annals of Internal Medicine*, 156 (1): 37-40.

De Vocht, F., Katikireddi, S. V., McQuire, C., Tilling, K., Hickman, M. and Craig, P. (2021) 'Conceptualising natural and quasi experiments in public health'. *BMC Medical Research Methodology*, 21 (1): 1-8.

Dhillon, L. and Vaca, S. (2018) 'Refining theories of change'. *Evaluation*, 14(30): 64-87.

Djulbegovic, B. and Guyatt, G. (2020) 'Evidence-based medicine in times of crisis'. *Journal of clinical epidemiology*, 126: 164-166.

Every-Palmer, S. and Howick, J. (2014) 'How evidence-based medicine is failing due to biased trials and selective publication' *Journal of Evaluation in Clinical Practice*, 20 (6): 908-914.

Feddes, A. R. and Galluci, M. (2015/6) 'A Literature Review on Methodology used in Evaluating Effects of Preventive De-radicalization Interventions'. *Journal for Deradicalization*, 5: 1-27.

Ferrari, R. (2015) 'Writing narrative style literature reviews'. *Medical Writing*, 24 (4): 230-235.

Fetterman, D. and Wandersman, A. (2007) 'Empowerment evaluation: Yesterday, today and tomorrow' *American Journal of Evaluation*, 28 (2): 179-198.

Fetterman, D., Rodriguez-Campos, L, Wandersman, A., O'Sullivan, R.G. (2014) 'Collaborative, participant and empowerment evaluation: Building a strong conceptual foundation for





stakeholder involvement approaches to evaluation (A reponse to Cousins, Whitmore and Shulha, 2013).' *American Journal of Evaluation*, 35 (1): 144-148.

Fielding, J. E. and Briss, P. A. (2005) 'Promoting Evidence-Based Public Health Policy: Can We Have Better Evidence And More Action?' *Health Affairs*, 25 (4): 969-978.

Freese, R. (2014) 'Evidence-based Counterterrorism or Flying Blind? How to Understand and Achieve What Works'. *Perspectives on Terrorism*, 8 (1): 37-56.

Gambrill, E. (1999) 'Evidence-based practice: an alternative to authority-based practice'. *Families in Society*, 80 (4): 341-350.

Garcia-Doval, I., Albrecht, J., Flohr, C., Batchelor, J. and Ingram, J. R. (2018) 'Optimizing case reports and case series: guidance on how to improve quality'. *British journal of Dermatology*, 178(6): 1257-1262.

Garg, A. X., Hackam, D. and Tonelli, M. (2008) 'Systematic reviews and meta-analysis: when one study is just not enough'. *Clinical Journal of the American Society of Nephrology*, 3 (1): 253-260.

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*, 32 (6): 1149-1167.

Gill, P., Cemmw, C., Hetzel, F., Rottweiler, B., Salman, N., Van der Vegt, I., Marchment, Z., Schumann, S., Zolghadriha, S., Schulten, N., Taylor, H. and Corner, E. (2020) 'Systematic Review of Mental Health on Violent Extremism'. *The Journal of Forensic Psychiatry & Psychology*, 1-28.

Greenhalgh, T., Kristjansson, E. and Robinson, V. (2007) 'Realist review to understand the efficacy of school feeding programmes'. *BMJ*, 335 (7625): 858-861.

Greenhalgh, T., Howick, J. and Maskrey, N. (2014) 'Evidence-based medicine: a movement in crisis? *BMJ*, 348: g3725.

Greenhalgh, T. (2020) 'Will Covid-19 be evidence-based medicine's nemesis?' *PLOS Medicine*, 17(6): 1-4.

Guerin, B., Martin, A. and van Gorp, A. (2018) 'Cost-benefit analysis'. In Hofman, J. and Sutherland, A. (eds) *Evaluating interventions that prevent or counter violent extremism* (Santa Monica, CA and Cambridge, UK: RAND Corporation), 37-42.

Guyatt, G. et al. (1992) 'Evidence-based medicine: a new approach to teaching the practice of medicine'. *JAMA*, 268(17): 2420-2425.

Guyatt, G. and Busse, J. W. (2006) 'The Philosophy of Evidence-based Medicine'. In Montori, V. M. (ed) *Evidence-based Endocrinology. Contemporary Endocrinology* (Totowa, N.J: Humana Press), pp. 25-33.

Hassan, G., Rabah, J. Madriaza, P., Brouillette-Alarie, S., Borokhovski, E., Pickup, D., Varela, W., Girard, M., Durocher-Costa, L. and Danis, E. (2022) 'PROTOCOL: Hate online and in traditional media: A systematic review of the evidence for associations or impacts on individuals, audiences and communities. *Campbell Systematic Reviews*, 18: e1245.

Helmus, T. C., Matthews, M., Ramchand, R., Beagley, S., Stebbins, D., Kadlec, A., Brosn, M. A., Kofner, A. and Acosta, J. D. (2017) *RAND Program Evaluation Toolkit: Countering Violent Extremism* (Santa Monica, Cal.: RAND Corporation).

Heneghan, C., Mahtani, K. R., Goldacre, B., Godlee, Fl, Macdonald, H. and Jarvies, D. (2017) 'Evidence based medicine manifest for better healthcare: a response to systematic bias, wastage, error and fraud in research underpinning patient care'. *Evidence Based Medicine*, 22: 120-122.

Ioannidis, J. P. A. (2016) 'Evidence-based medicine has been hijacked: a report to David Sackett' *Journal of Clinical Epidemiology*, 73: 82-86.

Impact Europe (2022a) 'Interventions database'. Available at: <http://www.impact.itti.com.pl/index#/inspire/search>.

Impact Europe (2022b) 'Randomized control trials'. Available at: <http://www.impact.itti.com.pl/index#/repository/44/whatisit>.

Impact Europe (2022c) 'Quasi-experimental designs'. Available at: <http://www.impact.itti.com.pl/index#/repository/39/whatisit>.

Jadad, A. J. and Enkin, M. W. (2007) *Randomized Controlled Trials: Questions, Answers and Musings, 2nd edition* (Backwell Publishing), 29-46.





Jenicek, M. (1997) 'Epidemiology, evidence-based medicine and evidence-based public health'. *Journal of Epidemiology*, 7(4): 187-197.

Jenicek, M. (1999) *Clinical Case Reporting in Evidence-Based Medicine* (Oxford: Butterworth-Heinemann).

Kabongo, E. M., Mukumbang, F. C., Delobelle, P. and Nicol, E. (2020) 'Combining the theory of change and realist evaluation approaches to elicit an initial program theory of the MomConnect program in South Africa'. *BMC Medical research methodology*, 20(1): 1-11.

Kemm, J. (2006) 'The limitations of 'evidence-based' public health' *Journal of Evaluation in Clinical Practice*, 12(3): 319-324.

Kesmodel, U. S. (2018) 'Cross-sectional studies: what are they good for?'. *Acta obstetrician et gynecologica Scandinavica*, 97 (4): 388-393.

Khakee, A. (2003) "The Emerging Gap between Evaluation Research and Practice". *Evaluation*, 9(3): 340-352.

Kohatsu, N. D., Robinson, J. G., and Torner, J. C. (2004) 'Evidence-Based Public Health: An Evolving Concept'. *American Journal of Preventive Medicine*, 27(5): 417-421.

Kooistra, B., Dijkman, B., Einhorn, T. A. and Bhandari, M. (2009) 'How to design a good case series'. *Journal of Bone and Joint Surgery*, 91(3): 21-26.

Kyriakoulis, K., Patelarou, A., Laliotis, A., Wan, A. C., Matalliotakis, M., Tsou, C. and Patelarou, E. (2016) 'Educational strategies for teaching evidence-based practice to undergraduate health students: Systematic review'. *Journal of Educational Evaluation for Health Professions*, 13 (34)

Lafuente-Lafuente, C., Leitao, C., Kilani, I., Kacher, Z., Engels, C., Canoui-Poitrine, F. and Belgim, J. (2019) 'Knowledge and use of evidence-based medicine in daily practice by health professionals: a cross-sectional survey'. *BMJ Open*, 9 (3): 1-6.

Lancaster, K., Rohdes, T. and Rosengarten, M. (2020) 'Making evidence and policy in public health emergencies: lessons from COVID-19 for adaptive evidence-making and intervention'. *Evidence & Policy*, 1-14.

Lösel, F., King, S., Bender, D. and Jugl, I. (2018). 'Protective Factors Against Extremism and Violent Radicalization: A Systematic Review of Research'. *International Journal of Developmental Science*, 12 (1-2): 89-102.

Lum, C., Kennedy, L. W. And Sherley, A. J. (2006). 'The Effectiveness of Counter-Terrorism Strategies'. *Campbell Systematic Reviews*, 2 (1): 1-50.

Maciejewski, M. L. (2020) 'Quasi-experimental design'. *Biostatistics & Epidemiology*, 4 (1): 38-47.

Marsden, S. (2020) 'Radicalization: Trajectories in Research, Policy and Practice'. *International Journal of Conflict and Violence*, 14 (2): 1-6.

Mazerolle, L., Eggins, E., Cherney, A., Hine, L., Higginson, A. and Belton, E. (2020). "Police programmes that seek to increase community connectedness for reducing violent extremism behaviour, attitudes and beliefs". *Campbell Systematic Reviews*, 16: e1111.

Mayne, J. (2012) 'Contribution analysis: Coming of age?'. *Evaluation*, 18(3): 270-280.

Mazerolle, L., Cherney, A., Eggins, E., Hine, L. and Higginson, A. (2021). "Multiagency programs with police as a partner for reducing radicalization to violence". *Campbell Systematic Reviews*, 17: e1162.

Miller, R. L. and Campbell, R. (2006) 'Taking Stock of Empowerment Evaluation: An Empirical Review'. *American Journal of Evaluation*, 27 (3): 296-319.

Molenkamp, M., Wouterse, L. And Gielen, A. (2018). 'Ex Post Paper: Guideline Evaluation of PCVE Programmes and Interventions'. Available at: https://ec.europa.eu/home-affairs/system/files/2020-09/ms_workshops_guidelines_evaluation_of_pcve_programmes_and_interventions_july_2018_en.pdf

Montori, V. M. and Guyatt, G. H. (2008) 'Progress in Evidence-Based Medicine'. *JAMA*, 300 (15): 1814-1816.

Mullen, E. J., Bledsoe, S. E. and Bellamy, J. L. (2007) 'Implementing Evidence-Based Social Work Practice'. *Research on Social Work Practice*, 18 (4): 325-338.

Nehls, I., Biene, J., Coester, M., Greuel, F., Milbradt, B., Armbrost, A. (2020). 'Evident and Effective? The Challenges, Potentials and Limitations of Evaluation Research on





Preventing Violent Extremism'. *International Journal of Conflict and Violence*, 14 (2): 1-20.

Oakely, A., Strange, V., Bonell, C., Allen, E. And Stephenson, J. (2006) 'Process evaluation in randomized controlled trials of complex interventions'. *BMJ*, 332 (7538): 413-416.

Odera, E. L. (2021) 'Capturing the Added Value of Participatory Evaluation'. *American Journal of Evaluation*, 42 (2): 201-220.

Parente, R. C. M., de Oliveira, M. A. and Celeste, R. K. (2010) 'Case reports and case series in the era of evidence-based medicine'. *Brazilian Journal of Videoendoscopic Surgery*, 3 (2): 63-66.

Parrish, D. A. (2018) 'Evidence-Based Practice: A Common Definition Matters'. *Journal of Social Work Education*, 54 (3): 407-411.

Patton, M. Q. (2008) *Utilization-focused evaluation, 4th edition* (Thousand Oaks, CA: Sage).

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. (2005) 'Realist review – a new method of systematic review designed for complex policy interventions'. *Journal of Health Services Research & Policy*, 10 (1): 21-34.

Pistone, I., Eriksson, E., Beckman, U., Mattson, C. And Sager, M. (2019). 'A scoping review of interventions for preventing and countering violent extremism: Current status and implications for future research'. *Journal for Deradicalization*, 19: 1-84.

Plint, A. C., Moher, D., Morrison, A., Schulz, K., Altman, D. G., Hill, C., Gaboury, I. (2006). 'Does the CONSORT checklist improve the quality of reports of randomized controlled trials? A systematic review'. *Medical Journal of Australia*, 185 (5): 263-267.

Rosen, A. (2003) 'Evidence-based social work practice: Challenges and promise'. *Social Work Research*, 27 (4): 197-208.

Rychetnik, L., Frommer, M., Hawe, P. and Shiell, A. (2002) 'Criteria for evaluating evidence on public health interventions'. *Journal of Epidemiological Community Health*, 56: 119-127.

Rychetnik, L., Hawe, P., Waters, E., Barratt, A. and Frommer, M. (2004) 'A glossary for evidence based public health'. *Journal of Epidemiology & Community Health*, 58 (7): 538-545.

Sackett, D. L., Rosenberg, W. M., Gray, J., Haynes, R. B., Richardson, W. S. (1996) 'Evidence-Based Medicine: What It Is and What It Isn't. *BMJ*, 312: 71-72.

Sackett, D. L. (1997) 'Evidence-based medicine'. *Seminars in Perinatology*, 21 (1): 3-5.

Sadeghi-Bazargani, H., Tabrizi, J. S. and Azami-Aghdash, S. (2014) 'Barriers to evidence-based medicine: a systematic review'. *Journal of Evaluation in Clinical Practice*, 20 (6) :793-802.

Sarma, K. M., Carthy, S. L., and Cox, K. M. (2022) 'Mental disorder, psychological problems and terrorist behaviour: A systematic review and meta-analysis'. *Campbell Systematic Reviews*, 18: e1268.

Schmidt, K. R. T. and Teti, D. M. (2004) 'Issues in the Use of Longitudinal Cross-Sectional Designs'. In: Teti, D. M. (ed) *Handbook of Research Methods in Developmental Science* (Malden, MA: Blackwell), 3-20.

Schulz, K. F. and Grimes, D. A. (2002) 'Case-control studies: research in reverse'. *The Lancet*, 359(9304): 431-434.

Shadish, W. R. and Luellen, J. K. (2006) In: Green, J. L., Camilli, G. and Elmore, P. B. (eds) *Handbook of Complementary Methods in Education Research* (Washington D. C.: Lawrence Erlbaum Associates Inc.), 539-550.

Shahjahan, R. A. (2011) 'Decolonizing the evidence-based education and policy movement: revealing the colonial vestiges in education policy, research, and neoliberal reform'. *Journal of Education Policy*, 26 (2): 181-206.

Shapira, Y., Enosh, G. and Havron, N. (2017) 'What Makes Social Work Students Implement Evidence-Based Practice Behaviours?'. *Journal of Social Work Education*, 53 (2): 187-200.

Shawcross, W. (2023) "Independent Review of Prevent". Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1134986/Independent%20Review%20of%20Prevent.pdf

Sherman, L. W. (1998) *Evidence-based policing* (Washington DC: Police Foundation).

Sidebottom, A., Thornton, A., Thompson, L., Belur, J., Tilley, N. and Bowers, K. (2017) 'A systematic review of tagging as a method to reduce theft in retail environments'. *Crime Science*, 6 (1): 1-17.





Siemieniuk, R. and Guyatt, G. (2022) 'What is GRADE?' *BMJ Best Practice*. Available at: <https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/>.

Sim, M. and van Gorp, A. (2018) 'Policy scientific approach'. In Hofman, J. and Sutherland, A. (eds) *Evaluating interventions that prevent or counter violent extremism* (Santa Monica, CA and Cambridge, UK: RAND Corporation), pp. 117-119.

Spensberger, F., Kollar, I., Gambrill, E., Ghanem, C. and Pankofer, S. (2019) 'How to teach evidence-based practice in social work: A systematic review'. *Research on Social Work Practice*, 30 (1): 19-39.

Strauss, S. E., Glasziou, P., Richardson, W. S., Haynes, R. B. (2019) *Evidence-based Medicine: How to Practice and Teach EBM* (fifth edition) (London: Elsevier).

Sutton, A., Clowes, M., Preston, L. and Booth, A. (2019) 'Meeting the review family: exploring review types and associated information retrieval requirement'. *Health Information and Libraries Journal*, 36 (6): 202-222.

Torres, R. T. and Preskill, H. (2001) "Evaluation and organizational learning: Past, present and future". *The American Journal of Evaluation*, 22 (3): 387-395.

Vandenbroucke, J. P. (2001) 'In Defense of Case Reports and Case Series'. *Annals of internal medicine*, 134(4): 330-334.

Van der Zwart, R. (2018) *Implementing evidence-based practice in social work: a shared responsibility*. Tilburg University.

Walker, E., Hernandez, A. V. and Kattan, M. W. (2008) 'Meta-analysis: Its strengths and limitations'. *Cleveland Clinic journal of medicine*, 75 (6): 431-439.

Wang, X. and Cheng, Z. (2020) 'Cross-sectional studies: strengths, weaknesses, and recommendations'. *Chest*, 158 (1): 65-71.

Warren, E. A., Melendez-Torres, G. J. and Bonell, C. (2022) 'Are realist randomized controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention'. *Trials*, 23 (1): 1-10.

Welsh, B. C. and Farrington, D. P. (2001) 'Toward an Evidence-Based Approach to Preventing Crime. *The Annals of the American Academy of Political and Social Science*, 578: 158-173.

Westhorp, G., Prins, E., Kusters, C., Guijt, I. and Brouwers, J. (2011) 'Realist Evaluation: an overview'. *Seminar Report, Wageningen UR Centre for Development Innovation (CDI)*. Available at: <https://library.wur.nl/WebQuery/wurpubs/fulltext/173918>.

White, H. (2013) 'An introduction to the use of randomized control trials to evaluate development interventions'. *Journal of Development Effectiveness*, 5 (1): 30-49.

White, H. and Sabarwal, S. (2014) 'Quasi-Experimental Design and Methods'. UNICEF Methodological Briefs, Impact Evaluation No. 8. Available at: [https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental Design and Methods ENG.pdf](https://www.unicef-irc.org/KM/IE/img/downloads/Quasi-Experimental%20Design%20and%20Methods%20ENG.pdf).

Wike, T. L., Grady, M., Massey, M., Bledsoe, S. E., Bellamy, H. and Putzu, C. (2019) 'Newly Educated MSW Social Workers' Use of Evidence-Based Practice and Evidence-Supported Interventions: Results From an Online Survey'. *Journal of Social Work Education*, 55 (3): 504-518.

Williams, M. J. and Kleinman, S. M. (2014) 'Utilization-focused guide for conducting terrorism risk reduction program evaluations'. *Behavioural Sciences of Terrorism and Political Aggression*, 6 (2): 102-146.

Windisch, S., Wiedlitzka, S., Olaghene, A. and Jenaway, E. (2022). 'Online interventions for reducing hate speech and cyberhate: A systematic review'. *Campbell Systematic Reviews*, 18: e1243.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D. and Hasisi, B. (2021). "Cognitive and behavioural radicalization : A systematic review of the putative risk and protective factors". *Campbell Systematic Reviews*, 17: e1174.

Wolfowicz, M., Hasisi, B. and Weisburd, D. (2022) 'What are the effects of different elements of media on radicalization outcomes? A systematic review'. *Campbell Systematic Reviews*, 18: e1244.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. and Pawson, R. (2013) 'RAMESES publication standards: meta-narrative reviews'. *BMC Medicine*, 11 (1): 1-15.

Worrall, J. (2007) 'Evidence in Medicine and Evidence-Based Medicine'. *Philosophy Compass*, 2(6): 981-1022.





Yee, J. L. and Niemeier, D. (1996) 'Advantages and disadvantages: Longitudinal vs. repeated cross-section surveys'. Discussion paper, Project Batelle. Available at: <https://rosap.ntl.bts.gov/view/dot/13793>

Zych, I. and Nasaescu, E. (2012) 'Is radicalization a family issue? A systematic review of family-related risk and protective factors, consequences, and interventions against radicalization'. *Campbell Systematic Reviews*, 18: e1266.

