



The Global
SDG Synthesis
Coalition



Advancing Evidence.
Improving Lives.



16 PEACE, JUSTICE
AND STRONG
INSTITUTIONS



The impacts of violence prevention programmes on homicides and conflict-related deaths and their implementation

**A PROTOCOL FOR A SYNTHESIS OF EVALUATIVE
EVIDENCE RELATED TO THE PEACE PILLAR OF THE
SUSTAINABLE DEVELOPMENT GOALS**

Thomas de Hoop, Andi Coombes, Anna Warren, Chinmaya Holla, Emily Mutea, Suchi Kapoor Malhotra, Eve Namisango, Swati Mantri, Shalu Jain, Bhumika TV, Howard White

July, 2024

The impacts of violence prevention and peace-building programmes on homicides and conflict-related deaths and their implementation

A protocol for a synthesis of evaluative evidence related to the peace pillar of the Sustainable Development Goals

Copyright © The Global SDG Synthesis Coalition July 2024

This publication, produced by American Institutes for Research with inputs from the Global SDG Synthesis Coalition partners does not necessarily reflect the views of the individual organizations, entities and bodies partnering with the Global SDG Synthesis Coalition. Individual partners of the Global SDG synthesis Coalition (including staff and personnel) cannot ensure the accuracy, completeness or currency of the data included in this work and they do not assume responsibility for any errors, omissions, or discrepancies in the information, or liability with respect to the use of or failure to use the information, methods, processes or conclusions set forth.

Citation: de Hoop, T., Coombes, A., Warren, A., Holla, C., Mutea, E., Malhotra, S.K., Namisango, E., Mantri, S., Jain, S., Bhumika T.V., White, H. (2024). "The impacts of violence prevention and peace-building programmes on homicides and conflict-related deaths and their implementation: A protocol for a synthesis of evaluative evidence related to the peace pillar of the Sustainable Development Goals". New York: The Global SDG Synthesis Coalition.

<https://www.sdgsynthesiscoalition.org/pillar/peace-pillar>

Acknowledgements

The Global SDG Synthesis Coalition would like to thank all those who have contributed to this synthesis protocol.

Synthesis Team (American Institutes for Research):

Thomas de Hoop, Andi Coombes, Anna Warren, Chinmaya Holla, Emily Mutea, Suchi Kapoor Malhotra, Eve Namisango, Swati Mantri, Shalu Jain, Bhumika TV, Howard White

Management Group Co-Leads:

UNDP Independent Evaluation Office: Shivit Bakrania (Senior Evaluation Synthesis Specialist), Isabelle Mercier (Director), Taeyoung Kim (Evaluation Synthesis Analyst)

UNFPA Independent Evaluation Office: Deborah McWhinney (Evaluation Advisor)

UNODC Independent Evaluation Section: Katharina Kayser (Chief), Katherine Aston (Deputy Chief), Moritz Schuberth (Evaluation Officer), Emanuel Lohninger (Evaluation Officer)

UN Office of Internal Oversight Services Inspection and Evaluation Division: Eddie Yee Woo Guo (Director), Srilata Rao (Chief of Peacekeeping Evaluation Section), Hanife Cakici (Evaluation Officer)

UN Peacebuilding Support Office: Bushra Hassan (Senior Monitoring and Evaluation Adviser), Tim Heine (Monitoring and Evaluation Manager)

Management Group Members:

IOM Evaluation Office: Andres Botero (Senior Evaluation Officer), Christophe Franzetti (Evaluation Officer)

Ireland: Rachel Ingersoll (Evaluation Manager, Department of Foreign Affairs)

OHCHR: Kristof Smits (Policy Officer)

Spain: Daniel Jacobo Orea (Head of Division, Evaluation of Sustainable Development Policies Division, Ministry of Foreign Affairs)

UNESCWA: Julien Letellier (Economic Affairs Officer), Ian Gately (Associate Evaluation Officer), Youssef Chaitani (Project Coordinator)

UNHCR: Joel Kinahan (Evaluation Officer)

UNICEF: Michele Tarsilla (Chief of Humanitarian Evaluation)

UNOCT: Josephine Mwenda (Programme Management-Evaluation and Compliance Officer)

Technical Advisory Group Members:

Angi Yoder-Maina (Executive Director, Green String Network), Cate Buchanan (Conflict Advisor, Livelihoods and Food Security Fund Myanmar), Cedric de Coning (Research Professor, Norwegian Institute of International Affairs), Christoph Zuercher (Professor, University of Ottawa), Elizabeth Drew (Head of Cross Government Strategy, UK Foreign, Commonwealth and Development Office), Emmanuel Kwesi Aning (Director of Faculty of Academic Affairs & Research, Kofi Annan International Peacekeeping Training Centre), Ieva Giedraityte (Researcher, Vilnius University), Nadim Houry (Executive Director, Arab Reform Initiative), Paul Jackson (Professor, University of Birmingham), Robert Muggah (Principal, SecDev Group), Sarah Bolger (Principal Consultant, IOD PARC), Sharon Bhagwan Rolls (Project Manager, Global Partnership for the Prevention of Armed Conflict)

Publication and Communication:

UNDP Independent Evaluation Office: Anna Guerraggio (Senior Evaluation Adviser), Flora Jimenez (Communications Analyst), Iben Hjorth (Communications Analyst)

Donors:

Ireland, Spain, UNDP, UNFPA

Contents

- Introduction 6
- Synthesis Scope 7
- Theory of Change 10
 - Synthesis Questions..... 15
- Synthesis Protocol 17
 - Approach to Evidence Identification and Inclusion 18
 - Impact Evaluation Synthesis 32
 - Performance and Process Evaluation Synthesis 35
 - Development of Evidence Gap Maps 39
 - Limitations..... 40
- Work Plan..... 41
 - Quality Assurance 42
 - Coordination with Management Group and Technical Advisory Panel..... 42
- References 44
- Annex A. Search Terms 50
- Annex B. Critical Appraisal Tools 54

Exhibits

- Exhibit 1. Priority SDG-16 Targets and Outcome Indicators..... 7
- Exhibit 2. Included Categories and Example Initiatives 9
- Exhibit 3. Working Theory of Change..... 11
- Exhibit 4. Synthesis Matrix 16
- Exhibit 5. Methods Overview..... 18
- Exhibit 6. Key Criteria for Study Inclusion..... 19
- Exhibit 7. Inclusion and Exclusion Criteria for Study Types 23
- Exhibit 8. Databases and Search Approaches 26
- Exhibit 9. Summary of Final Search String by Sources..... 30
- Exhibit 10. Preliminary Coding Framework..... 31
- Exhibit 11. Estimation of Odds Ratios 34
- Exhibit 12. Quality Appraisal Categories for Qualitative Studies 36
- Exhibit 13. Draft Deductive Synthesis Framework 38
- Exhibit 14. Limitations and Mitigation Strategies 40
- Exhibit 15. Work Plan 41
- Exhibit B-1: Risk of Bias Tool for Experimental and Quasi-Experimental Studies..... 54
- Exhibit B-2: Qualitative Review Protocol..... 56
- Exhibit B-3: Quality Appraisal Tool for Small-n Impact Evaluations)..... 61

Introduction

1. The most recent Sustainable Development Report indicates that, according to current trends, “not a single Sustainable Development Goal (SDG) is projected to be met by 2030, with the poorest countries struggling the most” (Sachs et al., 2023, p. 2). Cascading global crises, such as the increased number of conflicts, threaten the ability of countries to achieve the SDGs, including the objectives defined under the peace pillar or SDG-16 (Sachs et al., 2023). The number of homicides reached a peak in 2021 at nearly a half-million worldwide, and the number of civilians killed in war operations in 2022 was 16,988 – an increase of 53 per cent over the previous year (UNODC, OHCHR, and UNDP OGC, 2023).

2. As discussed in the evidence synthesis of the partnership pillar of the SDGs (de Hoop et al., 2023), developing evidence-based and practical lessons can contribute towards accelerating progress on the SDGs and making progress on the United Nations’ New Agenda for Peace. This requires assessing which programmes, policies, and interventions are most successful in achieving progress, why those initiatives are (or are not) effective, and under which conditions.

3. To contribute to this objective, the Global SDG Synthesis Coalition contracted American Institutes for Research (AIR) and its partner Campbell South Asia to design and implement an evidence synthesis on violence and peace-building in the context of SDG-16, the Peace Pillar of the SDGs. The synthesis will serve to generate lessons for UN organizations and UN Member States on what works to reduce violence. These UN organizations and UN Member States will also be the main audience of the synthesis.

4. As part of a scoping exercise prior to contracting AIR, the Global SDG Synthesis Coalition determined that the review will cover the effectiveness of initiatives that aim to improve progress towards target 16.1 (significantly reduce all forms of violence and related death rates everywhere) and 16.4 (significantly reduce illicit financial and arms flows, strengthen the recovery and return of stolen assets, and combat all forms of organized crime). The global SDG Synthesis Coalition then determined in partnership with the AIR team, the Management Group (MG), and the Technical

Advisory Group (TAP) that the review will focus on the following broad categories of initiatives (Sonnenfeld et al., 2020):

- 1) Social inclusion initiatives¹
- 2) Initiatives to stimulate peace processes
- 3) Initiatives to generate safe environments

5. For each of these initiatives, AIR will examine what works to reduce homicides (e.g., by focusing on initiatives that aim to prevent organized crime) and conflict-related deaths. The intended users of the synthesis include UN agencies, all UN Member States, researchers and evaluators, and other stakeholders focused on achieving the SDG-16 objectives.

6. This report presents the methodological protocol for this evidence synthesis. We first detail the scope, synthesis questions and the theory of change guiding the synthesis, including key decisions from the inception phase. We then present the synthesis protocol, which includes the approach to searching, identifying, and including evaluations and synthesis; the approach to the synthesis of impact evaluations; and the approach to the synthesis of performance and process evaluations.

Synthesis Scope

7. To manage the scope of the synthesis and ensure utilization-focused lessons, the Global SDG Synthesis Coalition carried out an initial scoping of existing evidence that ultimately prioritized SDG-16 targets 16.1 and 16.4 and their associated outcome indicators (Exhibit 1). Prioritizing these targets will allow for more in-depth analysis on the effects of specific initiatives on homicides, conflict-related deaths, and intermediate outcomes. Exhibit 1 presents the priority indicators and targets.

Exhibit 1. Priority SDG-16 Targets and Outcome Indicators

Target N	Target	Indicator N	Outcome Indicator
16.1	Significantly reduce all forms of violence and	16.1.1	Number of victims of intentional homicide per 100,000 population, by sex and age

¹ Initiatives can include interventions, policies, programmes, etc.

	related death rates everywhere	16.1.2	Conflict-related deaths per 100,000 population, by sex, age, and cause
16.4	Significantly reduce illicit financial and arms flows, strengthen the recovery and return of stolen assets, and combat all forms of organized crime	16.4.1	Total value of inward and outward illicit financial flows (in current United States dollars)
		16.4.2	Proportion of seized, found, or surrendered arms whose illicit origin or context has been traced or established by a competent authority in line with international instruments

8. To further narrow the synthesis scope, the Global SDG Synthesis Coalition recommended using the WHO definitions of violence (Krug et al., 2002; Sardinha et al., 2022) to distinguish among self-directed, interpersonal, community, and collective violence. AIR will only *include* evaluations of initiatives that aim to reduce community violence (that is, violence in public or institutional settings, such as gang violence) and collective violence (that is, social, political, and economic violence among larger groups or states, such as civil war or ethnic cleansing).² The current review will *exclude* interventions that aim to reduce self-directed (e.g., suicidal behaviour and self-abuse) or interpersonal violence at the family or interpersonal level (e.g., gender-based violence in the household).

9. Finally, the Global SDG Synthesis Coalition, MG, and TAP reached consensus to include initiatives that aim to reduce conflict-related deaths or homicides through violence *prevention*³ rather than *protection* within the following categories: (1) social inclusion, (2) peace processes, and (3) safe environments (Sonnenfeld et al., 2020). These categorizations – developed in the 3ie Building Peaceful Societies Evidence Gap Map conducted by Sonnenfeld et al. (2020) – guided scoping discussions and search strategy testing. Exhibit 2 shows these categories and example initiatives (according to Sonnenfeld et al., 2020) that may be included in this synthesis. In the social inclusion category, we consider both social inclusion and social cohesion initiatives. These two concepts play crucial yet distinct roles in peace-building. While social inclusion emphasizes people’s equal access to opportunities to participate fully in society regardless of their background, social cohesion focuses on

² This means that we will not include evaluations of initiatives that aim to improve mental health except when such initiatives focus on the mental health of former combatants or criminals.

³ “Violence prevention involves taking direct actions to reduce violent victimization or perpetration by addressing the underlying causes. These causes encompass structural inequalities, social and cultural norms, traditional power dynamics, and risk factors at individual, family, and community levels” (Institute for Security Studies, 2021, p. 6).

connectedness, trust, shared values, and a sense of belonging for all communities (IMO, 2019).

Exhibit 2. Included Categories and Example Initiatives

Activity Category	Example Initiatives
Social inclusion (strengthening social cohesion and conflict resolution as well as inter-group perceptions and relations)	Peace education
	Peace messaging and media
	Dispute resolution
	Social inclusion, social cohesion, and reintegration*
	Intergroup dialogue and interaction
Safe environments (ending violence and building a safe and secure environment)	Peacekeeping missions
	Disarmament, demobilization, reintegration of gang members/ex-combatants
	Conflict-focused early warning systems
	Demining
	Civilian police reform
	Preventative protection measures
	Organized crime-focused (conventional, transnational), including crime related to smuggling migrants*
Peace processes and conflict prevention (supporting peace processes, oversight, and post-conflict justice)	Support for peace processes and negotiation (including peace missions regardless of the specific components of the mission)
	Support for peace agreement implementation and oversight
	Transitional justice processes
	Peace policy influencing

Note: Types of initiatives listed are examples based on the 3ie Building Peaceful societies evidence gap map; final list of included initiatives will be based on search results. Such initiatives should fall under the activity categories.*

10. We will focus on initiatives that have violence reduction as a specific objective. We will *exclude* evaluations of interventions that focus on other objectives that may indirectly reduce violence, but that do not explicitly aim to reduce violence (e.g., cash transfers or vocational and business training as well as social inclusion activities that do not explicitly aim to reduce violence). This decision is aligned with decisions made in the evidence gap map (EGM) on atrocity prevention strategies conducted by Kapoor-Malhotra et al. (2023). This EGM also only focused on initiatives that explicitly aimed to reduce violence, which is an approach similar to ours. The synthesis will

only *include* initiatives that explicitly aim to reduce violence. We will also exclude evaluations of government sanctions against radical groups. The protocol elaborates the inclusion and exclusion criteria in the sections that follow.

Theory of Change

11. Over the past two decades there has been little improvement in global homicide rates (United Nations Office on Drugs and Crime [UNODC], 2023; UNODC et al., 2023). Conflict-related deaths remained relatively low in the first decade of the 21st century, but experienced a sharp increase from 2011 to 2014, showed some improvement, then increased sharply again in 2020. In 2021 the rate of intentional homicides was at its highest in more than 20 years, which may have partly resulted from the COVID-19 pandemic, in addition to increases in gang-related and sociopolitical violence in some places (General Assembly Economic and Social Council, 2023). In 2021 nearly a half-million people died from homicide (UNODC et al., 2023), more than 120,000 died in armed conflicts (Uppsala Conflict Data Program, 2023), and more than 23,000 were killed in terrorist attacks (Statista Research Department, 2023).

12. While previous evidence syntheses have described existing evidence as fragmented and weak (Cramer et al., 2016, p. iii), more recent evidence gap maps found existing evidence from a substantial number of interventions. Sonnenfeld et al. (2020) grouped 40 interventions into six categories. The synthesis identified 195 completed and 47 ongoing impact evaluations, in addition to 29 completed systematic reviews and 5 ongoing systematic reviews. However, the synthesis also indicated that the studies that focused on mental health and psychosocial support (MHPSS) constitute a significant portion of the overall evidence base, and that many impact evaluations focus on gender equality behaviour change communication, community-driven development and reconstruction, and cash transfer and subsidy programmes.

13. Kapoor-Malhotra et al. (2023) grouped conflict and atrocity prevention (CAP) interventions by the extent to which they adopted direct or indirect approaches to peace-building. The evidence gap map on conflict and atrocity prevention strategies found 540 eligible studies of interventions with explicit conflict prevention aims (428 studies), explicit atrocity prevention aims (21 studies), and studies with implicit mixes of both conflict and atrocity aims (91 studies) (Kapoor-Malhotra et al., 2023). While primary evaluations exist for indirect approaches in conflict prevention (CP), atrocity prevention (AP), and joint CP/AP in dormant conflict settings, there are

no evaluations of purely AP or joint AP/CP interventions using indirect prevention approaches in active or latent conflict contexts. Research tends to emphasize direct approaches across all settings.

14. Based on these previous evidence syntheses, AIR created a working Theory of Change to guide the evidence synthesis. We will adapt the theory based on the key findings of the evidence synthesis, which will help in capturing lessons and the development of recommendations for policy and practice. Exhibit 3 presents the working Theory of Change.

Exhibit 3. Working Theory of Change



15. Initial conditions. Different disciplines provide different perspectives on drivers of violence. Political science perspectives often emphasize decision-making by political leaders who decide about conflict, for example, because of their own agenda (i.e., if autocrats are unaccountable), ideological reasons, mistaken beliefs or biases, uncertainty, and commitment problems (Blattman, 2022). Each of these factors can contribute to conflict-related deaths. At the same time, psychologists, sociologists, and micro-economists often discuss structural and social determinants of violence, such as the socio-economic and political context, the socio-economic community conditions, and social and physical environments (Armsteadt et al., 2021). The synthesis will include an analysis of how different initiatives relate to these drivers of violence.

16. Risk factors vary from poverty and economic challenges to political instability, corruption, poor mental health, lack of education, social inequality, climate change and environmental degradation, and inadequate policing and justice (Global Alliance for Reporting Progress on Peaceful, Just and Inclusive Societies, 2019; UNODC, 2023; United Nations and World Bank, 2017). For example, young people without sufficient educational and economic opportunities are more likely to engage in violent behaviour (Bagby et al., 2021).⁴ Similarly, social inequality, including unequal access to education, income-generating opportunities, and health services, contributes to violence (UNODC, 2023). In addition, social scientists expect conflict over natural resources to increase as climate change increasingly affects economies and migration (Anderson & Delisi, 2011; Mares & Moffett, 2016). The synthesis will include an analysis of how different initiatives relate to these risk factors.

17. Activities (Inputs, programmes, and policies). Given the range of types of violence, as well as the range of populations affected by violence, understanding what works in preventing violence requires specific interventions, as well as principles that bridge efforts. A Review of the Evidence and a Global Strategy for Violence Prevention (Carbonari et al., 2020) points to the following “principles for effective prevention” across the different forms of violence: building state capacity; promoting inclusive and participatory governments; strengthening social cohesion and resilience; promoting peaceful, inclusive, and gender-balance norms; promoting mediation and negotiation; targeting high-risk places, people, and behaviours; and building multisectoral partnerships and coalitions. The synthesis will include an analysis of how different initiatives include principles for effective prevention.

18. Initiatives use different approaches to mitigate community and collective violence. In this review we will focus on the effectiveness of the following intervention categories:

- **19. Anti-violence programmes and policies focus on social inclusion or changing community norms to prevent conflict.** These include public awareness campaigns on the value of life and senselessness of violence (Cano & Rojido, 2016) or targeted interventions that challenge views on the instrumentality of violence (Carthy et al., 2020). Within this category, we will

⁴ While young people without sufficient educational and economic opportunities are more likely to engage in violent behaviour, education interacts with support for terrorist and militant groups in different ways. In some countries higher education levels are correlated with support for suicide bombing, for example (Shafiq and Sinno, 2010).

also include programmes that aim to deradicalize terrorists and other extremist groups through social inclusion (e.g., initiatives that aim to improve social inclusion and social cohesion of former child soldiers, terrorists etc.). However, we will not include deradicalization initiatives that do not link directly to social inclusion (e.g., initiatives focusing on deradicalization of the general population).

- **20. Interventions that promote safe environments.** These could include measures to reduce armed conflict or community-level violence, such as targeted policing and police-neighbourhood partnerships (Cano & Rojido, 2016; UNODC, 2023), youth and gang outreach (Waller, 2021), and situational preventative measures such as improvements in lighting and surveillance or reoccupying abandoned spaces (Cano & Rojido, 2016). Approaches to transitional justice processes and prosecuting offenders after human rights violations may reduce the likelihood of future outbreaks of conflict (Global Alliance for Reporting Progress on Peaceful, Just and Inclusive Societies, 2019), as with measures to improve access to security and justice, particularly for vulnerable groups, and measures to generally strengthen the rule of law (United Nations and World Bank, 2017; UNODC, 2023).
- **21. Peace processes and conflict prevention:** These include power-sharing, settlement of disputes, bargaining agreements, and inclusive access to resources and political influence (United Nations and World Bank, 2017). Other governance structures that may abate conflict include embedding inclusive arrangements into constitutions, promoting decentralized and subnational governance, supporting dialogue on access to land and resources, increasing transparency and accountability for extractive resources such as oil and minerals, and increasing transparency and fiscal management regarding access to basic services (United Nations and World Bank). We also consider peace missions as part of this category.

22. Intermediate outcomes. Most inputs, programmes, and policies are likely to lead to intermediate outcomes, which reduce the likelihood of violence over the longer term. Intermediate outcomes include effective peace processes and oversight, reduced illicit financial and arms flows (SDG outcome indicators 16.4.1 & 16.4.2); strengthened social cohesion and conflict resolution; and a safe and secure environment (Sonnenfeld et al., 2020). Other intermediate outcomes may include knowledge about and attitudes towards the use of violence. Lastly, programmes to

reduce organized crime may also aim to reach intermediate outcomes, such as illicit financial flows and seized, found, and surrendered arms (in line with SDG 16.4).

23. Impacts. The direct, long-term impacts of these programmes include decreases in community violence (that is, violence in public or institutional settings) and collective violence – that is, social, political, and economic violence among larger groups or states (Krug et al., 2002; Sardinha et al., 2022). Such impacts include reduced homicides (SDG Indicator 16.1.1), armed conflict and conflict-related deaths (SDG Indicator 16.1.2), interstate violence, mass atrocities, and violent extremism. Programmes focused on reducing conflict may have larger effects on conflict-related deaths, while programmes emphasizing the reduction of organized crime may have larger effects on reductions in illicit financial flows and homicides.

24. We also included a number of intermediate outcomes. Intermediate outcomes reflect the pathways of the Theory of Change, or the mechanisms through which the programmes, policies, and interventions can achieve their longer-term objectives. These outcomes include strengthened social cohesion and conflict resolution, safe and secure environments, reduced financial and arms flows, and effective peace processes, oversight, and transitional justice.

25. Moderators. Given the context-specificity of many underlying factors of violence, our Theory of Change highlights key moderators that may partially influence the mechanisms through which programmes achieve their objectives. Moderators are contextual, household-level, or individual-level factors that are hypothesized to determine the effectiveness of programmes, policies, and interventions. Moderators may include considerations such as the gender, age, religion, and social status of programme participants, or contextual characteristics such as geography (including acute or protracted crisis settings) and institutions (e.g., impacts may be different in contexts with different state institutions, as well as different state capacity and fragility).

26. Leave no one behind. The SDG principle of Leave No One Behind, which must consider equity, equality, and non-discrimination, play an important role in the Theory of Change. The effectiveness of interventions may, for example, depend on the income and human development status of the country, which relates to the Universality principle. In addition, interventions may have different effects for men and women. Although the current synthesis will focus on violence outside the home, the Theory of Change emphasizes a cross-cutting focus on the gendered dimensions of violence, as well as other disadvantaged groups such as youth,

refugees, and internally displaced populations. We will also code for a focus on these populations and state capacity.

27. Risks and assumptions. Some key assumptions in our Theory of Change include that politicians and other stakeholders have incentives to settle disputes and reduce violence. Such incentives may differ between authoritarian and democratically elected governments (Blattman, 2022). In addition, it may be challenging to reduce violence in conflicts caused by ideological reasons, considering the historical context of many of these conflicts. Several other risks and assumptions may influence the effectiveness of specific programmes. We will explore these risks and assumptions when interpreting the results.

Synthesis Questions

28. The synthesis of evidence related to targets 16.1 and 16.4 of the Peace Pillar will cover the following broad synthesis questions:

1. **What works? Where?** What interventions are effective in reducing homicides and conflict-related deaths? What is the impact of these interventions on homicides and conflict-related deaths? How do these impacts differ across different contexts (e.g., low-income vs. middle-income countries; acute crisis⁵ vs. protracted crisis⁶ vs. non-crisis low/middle-income countries contexts).
2. **How or why does it work?** Why are interventions to reduce conflict-related deaths and homicides effective (or not)? How do these interventions achieve their objectives (or not)? How do contextual, implementation, and design factors influence the effectiveness of these interventions?
3. **Leave no one behind.** How do gender and marginalization affect the impact of interventions to reduce conflict-related deaths and homicides? How did interventions consider non-discrimination, equity, and equality considerations in their design and implementation?

29. AIR and its partner Campbell South Asia will answer these questions using a mixed-methods synthesis approach that will include (a) an evidence synthesis of impact evaluations and specifically randomized controlled trials, quasi-experimental studies, and small-n impact evaluations (e.g., process tracing, contribution analysis,

⁵ Acute crises refer to humanitarian crisis for which there is little or no warning (UNICEF, n.d.)

⁶ In protracted crisis a “significant proportion of the population is acutely vulnerable to death, disease, and disruptions in livelihoods over a prolonged period of time” (FAO, n.d.).

most significant change, outcome mapping);⁷ and (b) an evidence synthesis of process and performance evaluations⁸ that were independently conducted or commissioned by UN entities and development partners from multilateral or bilateral organizations, civil society organizations, or the private sector. We detail the synthesis questions, their sub-questions, and the approaches we will use to answer them in the evaluation matrix depicted in Exhibit 4.

Exhibit 4. Synthesis Matrix

Synthesis questions	Impact evaluation narrative synthesis	Meta-analysis	Methodological triangulation	Thematic analysis of performance and process	Heterogeneous effects by geography	Heterogeneous effects by gender
1. What does the evidence say about the impacts of interventions on conflict-related deaths and homicides? (synthesis question 1)						
2. How does effectiveness differ across different contexts and subregions? How does effectiveness differ between acute crisis, protracted crisis, and other international development settings? (synthesis question 1)						
3. What does the evidence say about unanticipated effects of these interventions? (synthesis question 1)						
4. How well suited to the context were interventions, and what was the quality of analysis that underpinned activity? (synthesis question 2)						
5. What are the common bottlenecks in or barriers to outcomes and/or operational effectiveness? This includes bottlenecks or barriers related to social and environmental factors. (synthesis question 2)						

⁷ The rest of this protocol presents more details on some of these methods. White and Phillips (2012) provide an overview of small-n impact evaluations. We will only focus on small-n impact evaluations that were independently conducted or commissioned by UN entities and development partners from multilateral or bilateral organizations, civil society organizations, or the private sector.

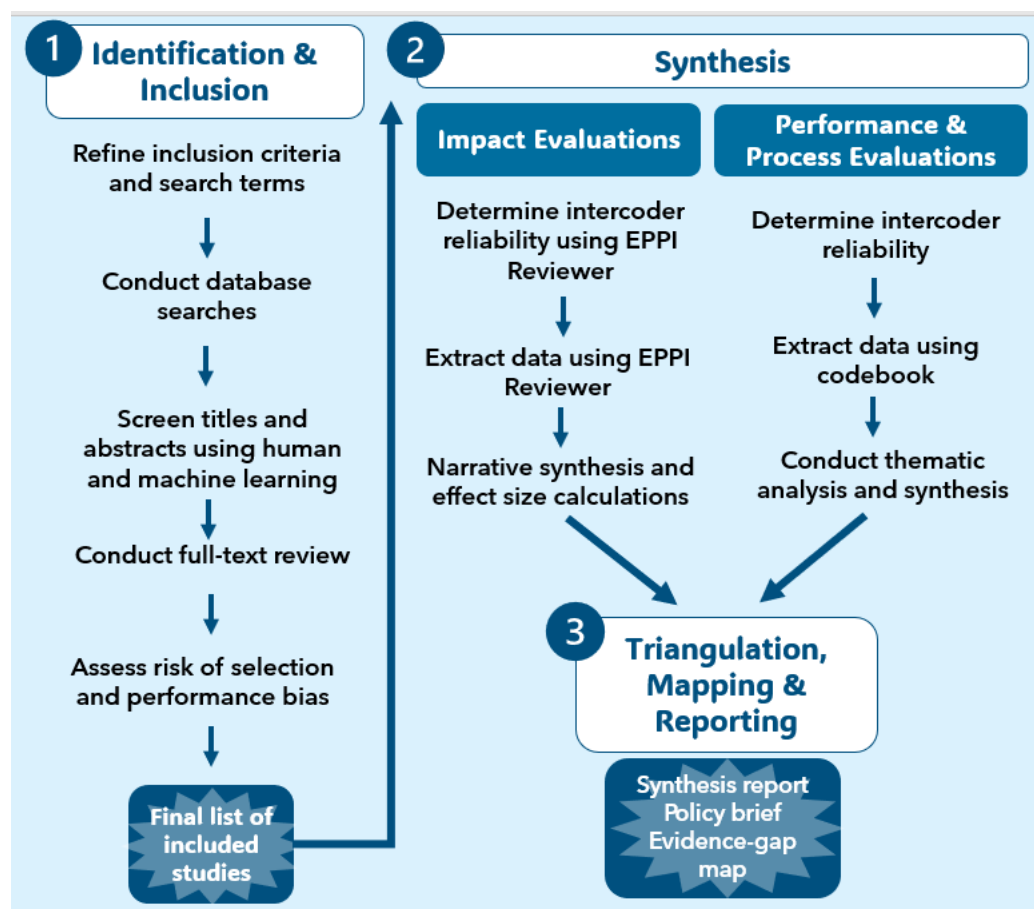
⁸ We recognize that some UN agencies use different labels for evaluations with similar objectives as performance and process evaluations. We will also include these evaluations, which may include implementation science; formative research, evaluations, or assessments; developmental evaluations; participatory evaluations; midterm evaluations; midterm reviews; or summative evaluations, among other evaluation types.

Synthesis questions	Impact evaluation narrative synthesis	Meta-analysis	Methodological triangulation	Thematic analysis of performance and process	Heterogeneous effects by geography	Heterogeneous effects by gender
6. Under what conditions have interventions been most effective and sustainable? (synthesis question 1)						
7. To what extent and in what ways were the needs of those furthest left behind addressed? (synthesis question 3)						
8. To what extent and in what ways were human rights principles incorporated? Which principles were or were not incorporated? (synthesis question 3)						

Synthesis Protocol

30. This section presents our approach to synthesizing the evidence, including refining search terms, pilot searching, inclusion criteria, and study screening. We include methodological protocols specific to the synthesis of impact evaluations and the synthesis of process and performance evaluations. Our synthesis design combines a systematic database search, critical appraisal of identified evaluations, and a synthesis of impact, performance, and process evaluations that meet the inclusion criteria. Exhibit 5 presents a summary of the synthesis design during the inception and synthesis phases.

Exhibit 5. Methods Overview



Note: Subsequent subsections include detailed descriptions of each activity highlighted in this figure.

Approach to Evidence Identification and Inclusion

31. We will use transparent inclusion and exclusion criteria to increase internal and external validity and to ensure that claims are representative of the existing high-quality evidence on SDG-16. Typical literature reviews can provide biased views of the state of the evidence without transparent inclusion and exclusion criteria and are thus not sufficient to identify the state of the current evidence and evidence gaps. Further, individual evaluations usually have a limited external validity because they are not able to provide a comprehensive overview of the literature. Their lack of external validity also may limit the ability of individual evaluations to assess how contextual considerations from different settings outside the evaluation context matter for the relevance, effectiveness, efficiency, sustainability, and impact of interventions at the global level.

Inclusion and exclusion criteria

32. This section describes the inclusion and exclusion criteria based on the scoping during the inception phase and recommendations by the Global SDG Synthesis Coalition, considering the timeline and available resources. The criteria aim to account for outcomes related to SDG 16.1 (“significantly reduce all forms of violence and related death rates everywhere”) and 16.4 (“significantly reduce illicit financial and arms flows, strengthen the recovery and return of stolen assets and combat all forms of organized crime”).

33. We developed and refined inclusion and exclusion criteria and a search strategy, relying on sample, phenomenon of interest, design, evaluation, and research type (SPIDER) criteria,⁹ as well as consultations with the Global SDG Synthesis Coalition, MG, and TAP.

34. Exhibit 6 below summarizes the key criteria for study inclusion according to sample, phenomenon of interest, design, evaluation, and research type. This is followed by detailed descriptions of each.

Exhibit 6. Key Criteria for Study Inclusion¹⁰

Domain	Inclusion criteria
Publication dates	2019–2024
Publication accessibility	Published in English, Spanish, or French Publicly available or shared with the synthesis team
Sample	Focuses on sample(s) in low- and middle-income countries. We will also include studies which link interventions in high-income countries and outcomes in low- and middle-income countries.

⁹ The SPIDER framework has many similarities with Population, Intervention, Comparison, and Outcome (PICO) criteria. The SPIDER framework refers to sample (groups of people, population, contexts, or settings that are of interest), phenomenon of interest (violence in this case), design (referring to evaluation methods), evaluation (referring to specific measures), and research type (quantitative, qualitative, or mixed-methods) (United Nations Evaluation Group, 2024).

¹⁰ **Impact evaluations** must have a credible control or comparison group and baseline measures (for quasi-experimental studies) to establish causal effects of an intervention. **Performance evaluations** must evaluate the performance of an intervention based on normative, descriptive, or cause-and-effect questions. **Process evaluations** must focus on implementation elements (such as programme design, institutional capacity, and community support) and explain why implementation failed or succeeded, identifying barriers to and facilitators of successful implementation. The performance or process evaluation needs to self-identify as an evaluation to be included (evaluations may use alternative terms such as “mid-term,” “participatory,” “strategic,” “institutional,” “programme,” “policy,” “summative,” or “global” evaluations). We will include three types of performance and process evaluations: (1) sibling studies (those linked to the experimental and quasi-experimental studies included in the synthesis), (2) high-quality evaluations commissioned or conducted by UN agencies or UN partners, and (3) high-quality studies from multilateral/bilateral agencies, civil society, or the private sector.

Phenomenon of interest	Assesses the impacts or implementation process of programmes pertaining to social inclusion, peace processes, and safe environments using impact, performance, and process evaluations
Design	Impact evaluations (randomized controlled trials/experiments or quasi-experimental studies with a comparison group), small-n impact evaluations (process tracing, contribution analysis, most significance change, outcome mapping, etc.), and performance or process evaluations (e.g., process evaluations or implementation science). Studies can be included if they cover both impact and performance/process elements (e.g., mixed-methods studies).
Gender considerations	Estimation of heterogeneous impacts by gender and examine gender considerations in performance and process evaluations
Evaluation	Outcomes related to violence prevention and peace-building at micro and macro levels. Outcomes will include terms such as "conflict," "war," "battle," "violence*," "armed clash," "extremism," "terrorism," "radicalism," "insurgency*," and "killing*." To search for illicit flows-related outcomes, we use terms such as "undeclared income," "criminal proceeds," and "evade tariff*." To search for organized crime-related outcomes, we use terms such as "criminal network*," "drug trafficking organization*," and "drug cartel*."

35. Sample: While we will include evidence related to interventions in high-income countries only when the outcomes are described for low/middle-income countries (LMICs), we will primarily select studies that cover initiatives implemented in LMICs.

36. While we recognize the universality principle of the SDGs, we do not consider it feasible, realistic, or methodologically appropriate to treat evidence from high-income countries the same as evidence from LMICs (unless the evidence from high-income countries includes LMICs as part of the scope), just as in the partnership pillar of the SDGs (de Hoop et al., 2023). The volume of research from high-income countries is much larger than the volume of evidence from LMICs. For instance, research on gun violence in the United States is likely to dominate the included studies on violence since it has been studied vastly more than violence in developing countries. Research output on a given country increases considerably with the country's wealth, as shown by Das et al. (2013) who use a database of 76,046 empirical economics papers to demonstrate that "over the 20-year span of the data, there were 4 empirical economics papers on Burundi, 9 on Cambodia and 27 on Mali. This compares to the 37,000 or so empirical economics papers published on the U.S. over the same time period." More recent evidence also shows that of the research in health economics journals only 2 percent covers LMICs (Hirvonen, 2020). Including all

evidence from high-income countries would lead to skewed results and misleading policy recommendations.

37. Including high-income countries in the “Sample” terms may surface studies that focus largely on interventions and programmes in high-income countries but that have implications for LMICs. These may include studies on legislation regarding illicit drugs in developed countries and their impact on SDG-16 indicators in developing countries, supply of arms from developed countries, or the provision of peacekeeping forces in conflict-affected regions. We will include such studies, though we will exclude studies that focus exclusively on phenomena of interest and evaluations in high-income countries.

38. Phenomena of interest: The broad categories of interventions are laid out in the Theory of Change described above (Exhibit 3), which are derived from the evidence gap map developed by 3ie on building peaceful societies (Sonnenfeld et al., 2020). Within each of these intervention categories, we identified specific interventions (e.g., “peace education” within the “social inclusion” category). We will focus on interventions and initiatives that aim to reduce community violence (i.e., violence in public or institutional settings) and collective violence (i.e., social, political, and economic violence among larger groups or states). The current review will *exclude* interventions and initiatives that aim to reduce self-directed violence (e.g., suicidal behaviour and self-abuse) as well as interpersonal violence at the family or interpersonal level (i.e., initiatives to improve mental health and gender-based violence). However, we will examine heterogeneous effects by sex and, if possible, for refugees and displaced populations.

39. Design: We will include impact evaluations that have either an experimental or a quasi-experimental design. Studies with an experimental research design use random assignment to the intervention (as part of randomized controlled trials). Quasi-experimental studies eligible to be included in this evidence synthesis include regression discontinuity designs, difference-in-differences analyses, instrumental variable analyses, and matching based on propensity scores. Quasi-experimental studies should leverage longitudinal data (baseline and endline data) and at least one comparison group to account for counterfactual trends and selection issues (systematic differences between those receiving and not receiving an intervention) to be included.

40. We will also include a selection of small-n impact evaluations (e.g., contribution analysis, process tracing, most significant change methods, outcome mapping, etc.) focusing on programmes for which it is challenging to identify a control or

comparison group because there are insufficient units of assignment to conduct statistical analysis of what difference the intervention has made. We will select these small-n impact evaluations from UN databases or databases of bilateral donors.

41. White and Phillips (2012) provide a comprehensive overview of such evaluation methods. For example, process tracing examines causal links between putative causes and outcomes by identifying intervening processes and mechanisms (Reilly, 2010). It starts with process induction to generate various hypotheses on how an intervention might connect to an outcome, and then establishes a series of causal mechanisms for evaluators to test using quantitative and/or qualitative data (White and Phillips, 2012). Contribution analysis “aims to demonstrate a plausible association between a program and observed outcomes, using weight of evidence by building a credible contribution story in which each step lying between program inputs and outcomes is clearly evidenced” (White and Phillips, 2012, pp. 11–12). Most significant change methods use participatory monitoring and evaluation to collect and select stories of significant changes, which have occurred in the field (White and Phillips, 2012). Outcome mapping uses similar participatory methods to track intervening steps between activities and outcomes. White and Phillips (2012) provide more details on these and other small-n impact evaluations.

42. Various interventions can be evaluated using such methods. Evaluations may focus on peacekeeping missions covering an entire country or diplomatic missions. They may also focus on the effectiveness of United Nations operations to reduce violence in a specific country. However, we will only include such evaluations with a transparent and clearly specified analysis method. In addition, we will likely only include a small number of such evaluations to keep the review manageable.

43. Performance and process evaluations will not require a control or comparison group. However, the performance or process evaluation needs to self-identify as an evaluation to be included (this may include evaluations using alternative terms such as mid-term, participatory, strategic, institutional, programme, policy, summative, or global evaluations). Exhibit 7 below presents the methodological inclusion/exclusion criteria for performance and process evaluations.

44. Evaluation: We will include studies that only cover intermediate outcomes as defined in the Theory of Change (Exhibit 3). During preliminary searches we found that a significant portion of the quantitative studies discussed intermediate outcomes – i.e., outcomes that are not measures of violence or violence-reduction, but that are likely to result in violence reduction or are proxy measures for violence

reduction. Similarly, because of the low likelihood of finding qualitative evaluations that directly measure these types of impacts, we will include intermediate outcomes that explicitly mention an aim to ultimately reduce violence. Examples of intermediate outcomes include social cohesion, effective peace processes, inclusive and accountable state institutions, and building a safe and secure environment. AIR will *exclude* evaluations on programmes that aim to achieve these intermediate outcomes, but fail to directly state an ultimate aim to reduce violence as part of the Theory of Change of that intervention. The Theory of Change presents our current overview of intermediate outcomes.

45. Research Type: AIR will include impact, performance, and process evaluations focused on outcome related to SDG indicators 16.1 and 16.4. Based on the re-scoping conducted by the Global SDG Synthesis Coalition and conversations with TAP and MG, the team will focus on the following phenomena of interest: social inclusion, peace processes, and safe environments. For example, AIR will include evaluations that focus on the effects of these phenomena of interest on violence reduction, but will not include evaluations that focus on the impacts of these interventions on poverty alleviation. Examples of SDG-16 outcomes in our final search string include conflict, persecution, and illicit financial flows. Further, the team will not include impacts of the selected phenomena of interest on inter-personal violence, economic empowerment, and gender empowerment.

Exhibit 7. Inclusion and Exclusion Criteria for Study Types

Evaluation Type	Definition	Inclusion	Example paper	Notes on inclusion/exclusion
Project or programme evaluation	Evaluations funded by multi-/bilateral agencies, civil society, or the private sector, even when not specifically linked to an impact evaluation.	Include	https://evaluation.iom.int/sites/g/files/tmzbd1151/files/docs/resources/RES%20Building%20Peace%20Human%20Security%20Cross%20Border%20Security%202018.pdf	
Country portfolio/	Evaluations to assess agency performance at	Include	https://erc.undp.org/evaluation/doc	We will include evaluations that explicitly mention a focus on

Evaluation Type	Definition	Inclusion	Example paper	Notes on inclusion/exclusion
programme evaluation	country-level across projects and programmes.		uments/download/15592	outcomes related to violence prevention.
Regional evaluation	Evaluations to assess agency performance for an entire regional across projects and programmes.	Include	https://erc.undp.org/evaluation/documents/detail/9585	We will include evaluations that explicitly mention a focus on outcomes related to violence prevention.
Thematic evaluation	Evaluations to assess specific themes across UN agency programmes.	Include	https://erc.undp.org/evaluation/documents/download/18508	
Strategy/policy evaluation	Evaluations to assess whether a policy or strategy is achieving its stated objectives in the most efficient way possible.	Include	https://www.wfp.org/publications/evaluation-policy-wfps-role-peacebuilding-transition-settings	
Mixed-method impact evaluation	Performance or process evaluation of or study related to the same programme evaluated through an included experimental or quasi-experimental impact evaluation	Include	TBD	

*Note: Evaluation type is a necessary but insufficient condition for inclusion. Evaluations must also meet other inclusion criteria in Exhibit 2 (e.g., methodological parameters) to be included.

Search strategy

46. This section describes the search strategy, which was designed based upon the SPIDER criteria described above. During this process, we selected the most relevant websites and databases for our review of impact, performance, and process

evaluations. For each database, we piloted and refined the search terms based on the SPIDER criteria based on the database specifications. Below are descriptive summaries of keywords for the sample, phenomenon of interest, design, evaluation, and research type.

- **47. Sample.** We will include the names of all countries categorized by the World Bank as LMICs, as well as the demonyms for these countries (e.g., “Mozambique” as well as “Mozambican”), as papers may refer to their study samples using either term. We will also include several descriptive keywords that authors may use to refer to these countries, such as “fragile state,” “failed state,” “low-income country,” “middle-income economy,” “developing country,” “low GDP,” and “Third World.” For such terms, we will use left truncation to pick up different variations: “low-income countr*” will enable us to pick up papers mentioning “low-income country” as well as “low-income countries.” We provide more details in later sections.
- **48. Phenomenon of interest.** We will use a wide range of keywords for different types of “violence prevention and peace-building” programmes, which are our topics of research, starting with general terms (e.g., “violence prevention”). We will rely on the framework provided by the “Building Peaceful Societies Evidence Gap Map” (Sonnenfeld et al., 2020), as recommended by UNDP based on the re-scoping exercise; and we will rely on the evidence gap map on atrocities and conflict prevention (Malhotra et al., 2024) to identify the types of phenomena of interest to focus on. We will generate search terms focused on the defined initiatives (e.g., “peace education”).
- **49. Design.** This part of the search string will comprise keywords on the different types of methods that this evidence synthesis will cover. To identify relevant impact evaluations, we will create a composite list of quantitative study designs, including “randomized controlled trials,” “propensity score,” “difference-in-differences,” and “regression discontinuity.”
- **50. Evaluation.** For this part of the string we will combine keywords related to (a) conflict and violence (such as “ethnic conflict,” “civil war,” “terrorism,” and “insurgency”); (2) violence prevention and peace-building (such as “reconciliation,” “conflict resolution,” and “peacekeeping”); and (3) psychosocial determinants of violence and conflict (such as “inter-group dialogue,” “intergroup dialogue,” “inter-group cooperation,” “intergroup cooperation,” and “co-existence”).

- **51. Research Type.** We will look at qualitative, quantitative, and mixed-methods studies. However, because of constraints of the databases from which we are sourcing qualitative studies (Exhibit 8), we will follow the SPIDER framework predominantly for quantitative studies. Since we are specifying the design strictly in our search term to yield only quantitative studies, we will not use the “Research Type” element in our search string.

52. We list the included databases and corresponding search approach in Exhibit 8. AIR will also write management and technical advisory group members to suggest evaluations that meet the inclusion criteria.

Exhibit 8. Databases and Search Approaches

Database	Type	Primary Evaluation Types	Approach
3ie Development evidence portal	Impact evaluations	Impact evaluations	<ul style="list-style-type: none"> • Use full search string except for sample and design terms (the database only includes experimental and quasi-experimental studies, and studies from low- and middle-income countries) • Filter by language: English • Filter by date, eliminating dates before 2019
DEReC database	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> • Search the following list of key words: “conflict” OR “violence” OR “security” OR “weapon” OR “weapons” or “peace” OR “peacebuilding” OR “peace-building” OR “disarmament” OR “guerrilla” OR “SDG 16” or “SDG-16” OR “terrorism” OR “terrorist” OR “war” OR “armed resistance” OR “organized crime” OR “police” OR “policing,” which has been tested and refined based on the relevance of preliminary results • Searched by sector: “conflict; peace and security” • Filter by date, eliminating dates before 2019 • Combine results of the two searches and eliminate duplicates
Web of Science	Academic	Impact, evaluations	<ul style="list-style-type: none"> • Use full search string • Filter by language: English • Filter by date, eliminating dates before 2019
United Nations Evaluation Group	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> • Use single keyword search for the following terms, determined by test-searching relevant articles that were tagged to SDG 16: “peace,” “conflict,” “violence,” “violent,” “organized crime” • Export results, de-duplicate, and filter by date

Database	Type	Primary Evaluation Types	Approach
World Bank Documents & Reports	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Filter by document date: 1 Jan 2019 to 30 April 2024 Filter by the following document types: IEG Evaluation, Impact Evaluation Report, Project Performance Assessment Report, Journal Article, and Report Filter by language: English, French, Spanish Filter by topic: armed conflict; conflict and fragile states; national protection and security; peace & peacekeeping; education violence and social cohesion; business, peace and democracy; crime and society; social cohesion; social conflict and violence Conduct an additional search with all of the filters, except by topic; search with keywords: conflict, violence, peace, peace-building, peacebuilding, disarmament, SDG 16, SDG-16, terrorism, terrorist, war, organized crime, police, policing <ul style="list-style-type: none"> Sort by most relevant; review first 500 documents to find any that do not appear in topic search or until saturation is achieved
USAID Development Experience Clearinghouse	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Filter by language: English, French, Spanish Filter by document type: Evaluation Summary, Final Evaluation Report, Journal Article, Other USAID Evaluation, Significant Evaluation, Special Evaluation Filter by publication date range Search the following thesaurus terms, determined by reviewing the 1,000 most recent studies: citizen security, conflict prevention, conflict resolution, crime prevention, crimes, peace building, violence, war
German Agency for International Cooperation (GIZ)	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Conduct individual searches by keyword: peace, conflict, violence, violent, organized crime, SDG 16 Filter by Type: GIZ-Publikation, Report, Project Report Filter by Language: English, French, Spanish Filter by Publication year: 2019 through 2024 Export searches, combine, and de-duplicate
Swedish International Development Cooperation Agency (SIDA)	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Search keywords: peace OR conflict OR violence OR organized crime OR SDG 16 Filter by type: SIDA evaluation, joint evaluation, decentralized evaluation, centralized evaluation Filter by language: English, French, Spanish Filter by year: 2019–

Database	Type	Primary Evaluation Types	Approach
UN organizations	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Requesting UN organizations to confirm whether their evaluations are in the United Nations Evaluation Group and request they send directly to AIR if they are not May search some key databases as feasible within timeline
Active Learning Network for Accountability and Performance	Institutional	Process and performance evaluations	<ul style="list-style-type: none"> Filter by resource type: "evaluations"

53. The suggested databases provide a comprehensive overview of the available literature from academic and non-academic sources. Web of Science is generally considered one of the most comprehensive academic databases, and the 3ie database of impact evaluations has the largest collection of impact evaluations in LMICs. The UN databases and the databases of bilateral donors were selected based on AIR’s experience during the synthesis of the partnership pillar and on the suggestions of management and technical advisory group members.¹¹

54. We developed a search string (see Annex A) and conducted a series of tests in Web of Science, identifying almost 11,000 hits. In developing this search string, we extracted key terms associated with the selected SDG-16 indicators (16.1 and 16.4). We then expanded our search string to include terms found in prominent databases on conflict, such as the Armed Conflict Location and Event Data Project. Further, we used the search results as input into the LitsearchR package to discover relevant terms that we missed in our search string. We also reviewed a series of abstracts from seminal papers on related topics and added further terms previously missing. Finally, we used ChatGPT to provide further alternative terms to avoid linguistic blind spots.

55. Starting with the search string in Annex A, we piloted the search strategy in the Web of Science and 3ie Development Evidence Portal databases. We refined the search strings by adding qualifying terms to keywords (e.g., added “evaluation” to “performance” to search for “performance evaluation”); by eliminating superfluous words (e.g., “implementation,” since phrases such as “implementation science” were

¹¹ AIR may consider searching databases of additional bilateral donors depending on manageability and MG and TAP recommendations.

already in the string); and by introducing punctuation to make the string more targeted (e.g., including quotations around phrases such as “dispute resolution” to return papers that had the full phrase instead of just “dispute” or just “resolution”). We also dropped terms that the Global SDG Synthesis Coalition suggested were beyond the scope of the current evidence synthesis (e.g., “economic empowerment” and “gender-based violence”).

56. We tailored the search strings to specific database requirements. For example, as the 3ie database only covers research conducted in LMICs, we omitted the country names from the search string for this database. The United Nations Evaluation Group (UNEG) and GIZ databases only allow users to search one keyword at a time, so for those databases we tested the key terms, conducted individual searches, combined results, and de-duplicated evaluations that appeared more than once. The USAID Development Experience Clearinghouse and the World Bank databases have preferred terms, which we tested, for example, by reviewing a sample of all studies, identifying relevant evaluations, and selecting the key terms for those evaluations until we achieved saturation.

57. We identified anchor/pivotal papers to ensure our search strings returned relevant documents in all databases. The anchor papers are either systematic reviews that provide critical evidence related to SDG-16, highly cited individual studies related to SDG-16, or reports suggested by the members of the TAP and the management group. During an initial scan of academic databases, our team identified primary impact studies, performance and process evaluations, and systematic reviews of relevant work related to SDG-16, specifically 16.1 and 16.4.

58. AIR also leveraged the anchor papers to increase the comprehensiveness of our search strings. To achieve this goal, we extracted the titles, keywords, and abstract or executive summary from the articles and reports in Zotero, a reference management tool. We then analysed the bibliographic information with an R package called litsearchr (Grames et al., 2019), which uses text-mining algorithms to analyse bibliographic information, to produce a list of phrases that appear frequently in the papers. We then included additional terms that litsearchr identified via a scan of the title and abstracts of the anchor papers.

59. Subsequently, we re-ran our updated search in the impact evaluation databases (i.e., Web of Science, 3ie) and screened the first few pages of results for each SDG-16 area to ensure the search strings identified relevant articles. We further refined inclusion/exclusion criteria after experimenting with keyword inclusions and

exclusions. For example, we identified the search terms detecting irrelevant articles and excluded these terms (e.g., “assessment”). This iterative process resulted in the final search strings, which we present in Annex A. Exhibit 9 summarizes the results of this final search string for each of the impact evaluation databases by topic.

Exhibit 9. Summary of Final Search String by Sources

SDG-16 Intervention	Web of Science	3ie
Social inclusion	4,402	1,944
Peace processes	936	250
Safe environments	2,226	365

Note 1. Web of Science terms contain sample, phenomenon of interest, design, and evaluation search terms. 3ie terms contain sample, phenomenon of interest, and evaluation terms due to database limitations.

Note 2: The “Total” is not a raw sum of the count of results for each intervention category since there is overlap in the results across multiple intervention categories.

Screening the evidence

60. In the next phase of this review we will screen results from the databases listed above. We will screen all studies and exclude those that do not meet the inclusion criteria (Exhibit 6). Initially, two team members will work independently to screen a sub-sample of the abstracts. We will discuss the results and continue with a single rater after we achieve sufficient interrater reliability (>0.90). To create efficiencies, we will also employ machine learning approaches in EPPI Reviewer to screen out studies that have a low likelihood of inclusion based on Artificial Intelligence tools.

61. During the title and abstract reviews using the above criteria, reviewers will select “yes” or “no” in EPPI Reviewer. If a reviewer marks “yes” for any of the criteria, the reviewer will continue to the next criterion on the coding sheet. If the reviewer marks “yes,” the study qualifies for the review of the full text. If a reviewer marks “no,” the study does not meet the criteria for further review. If reviewers disagree, the study is tagged for reconciliation. We will share the list of included studies with the Global SDG Synthesis Coalition, MG, and TAP.

62. We plan to code each included study based on the key indicators in Annex D using a coding sheet in EPPI-Reviewer web software for review management and coding. We will code each included study by beneficiary characteristics, geography, evaluation methodology, outcome, and whether the study is a sibling study of an

impact evaluation. We will code for the inclusion of vulnerable populations (e.g., individuals with a disability, youth, women, etc.) to enable the estimation of heterogeneous effects for these subgroups for all full-text studies. We will, for example, code for gender, age, displaced status, and poverty levels. Exhibit 10 presents the preliminary coding framework.

Exhibit 10. Preliminary Coding Framework

Thematic category	Definition of thematic category
Evaluated intervention	
Topic Level 1 (“parent category”)	E.g., social inclusion, peace processes, safe environments
Intervention type	Tag relevant intervention type, as defined by 3ie (includes individual programmes under broader PK missions)
SDG target	Tag relevant SDG-16 target
SDG indicator	Tag relevant SDG-16 indicator
Region	Region where the evaluated intervention was implemented
Conflict context	Acute crisis, protracted crisis, post-conflict, no conflict
Country	Country (or countries) where the evaluated intervention was implemented
Income level	Low income, lower-middle income, upper-middle income
Focus on disadvantaged group(s) and younger populations	Indigenous, women, low-income, disabled, caste, ethnicity, youth, other age groups, none
Age	
Evaluation	
Type of evaluation*	Process evaluation, performance evaluation, impact evaluation
Language of evaluation	English, Spanish, French
Evaluation year	Enter the evaluation year

Thematic category	Definition of thematic category
Source	Enter source of the evaluation (e.g., UN entity, university, government)
Quality (<i>process and performance evaluations</i>)	High, medium, low, based on quality appraisal score
Risk of selection bias (<i>impact evaluations</i>)	High, medium, low, based on risk of bias assessment
Included in evidence synthesis	Yes or No
Reason for exclusion	Explanation for why we excluded studies after closer examination

63. We will use three assessment tools to assess the methodological quality of included studies (see Annex B for the tools):

- A risk of bias assessment tool to appraise the quality of impact evaluations.
- A qualitative review tool to assess the quality of performance and process evaluations.
- A review tool to assess the risk of bias of small-n impact evaluations (process tracing, contribution analysis, most significant change, outcome mapping, etc.).

64. We will use different tools for impact evaluations and performance and process evaluations because, although there is overlap among the tools, each tool addresses unique methodological details not covered in the other tools. We anticipate that two members will work independently to apply the critical appraisal to a sub-sample of the full-text studies. We will then discuss the results and continue with a single rater after we achieve sufficient interrater reliability (>0.90).

65. The following sections detail the approaches for impact and process and performance evaluation syntheses.

Impact Evaluation Synthesis

Data extraction

66. Team members with expertise in impact evaluations will extract information from each experimental or quasi-experimental study included in the review. We will use a data extraction checklist that will be coded in EPPI Reviewer and fill in the details. We

will start by double-coding 5 percent of the studies, where two team members will independently extract information and then cross-check to ensure that there is sufficient inter-rater reliability.

Effect size calculations

67. To synthesize the quantitative studies, we will use the extracted information from experimental or quasi-experimental studies focused on intervention-outcome combinations for which we have more than five studies. For these intervention-outcome combinations, AIR will estimate the standardized effect sizes (for continuous variables) or odds ratios (for binary variables) across studies. In addition, we will calculate standard errors and 95 percent confidence intervals, where possible. This section presents the process to calculate effect sizes and is heavily based on Brody et al. (2015).

68. We will include meta-analyses for those intervention-outcome combinations that have more than five impact evaluations available.

69. We will report two types of effect sizes. We will calculate the Hedges' g sample-size-corrected standardized mean differences (SMDs) for continuous outcome variables, which measure the effect size in units of standard deviation of the outcome variable. We will calculate odds ratios for binary outcome variables.

70. First, we will calculate SMD in Cohen's d effect sizes by dividing the mean difference with the pooled standard deviation by applying the formula in Equation 1:

$$(1) \text{ SMD} = \frac{Y_t - Y_c}{S_p}$$

71. Here SMD refers to the standardized mean differences, Y_t refers to the outcome for the treatment group, Y_c refers to the outcome for the comparison group, and S_p refers to the pooled standard deviation.

72. The pooled standard deviation S_p can be calculated by relying on the formulas in Equations 2 and 3:

$$(2) S_p = \sqrt{\frac{((SDy^2) * (nt + nc - 2)) - \left(\frac{\beta^2 * (nt * nc)}{nt + nc}\right)}{nt + nc}}$$

$$(3) S_p = \sqrt{\frac{(nt - 1) * st^2 + (nc - 1) * sc^2}{nt + nc - 2}}$$

73. We will use Equation 2 for regression studies with a continuous dependent variable. In this equation, SD_y refers to the standard deviation for the point estimate from the regression, n_t refers to the sample size for the treatment group, n_c refers to the sample size for the control group, and β refers to the point estimate. We will use Equation 3 when there is information about the standard deviation for the treatment group and the control group separately.

74. We will correct the SMD for small sample size bias by relying on Equation 4, which transforms Cohen's d to Hedges' g .

$$(4) \text{SMD}_{\text{corrected}} = \text{SMD}_{\text{uncorrected}} * \left(1 - \frac{3}{4 * (n_t + n_c - 2) - 1}\right)$$

75. We will rely on Equation 5 to estimate the standard error of the SMD:

$$(5) \text{SE} = \sqrt{\frac{n_t + n_c}{n_c * n_t} + \frac{\text{SMD}^2}{2 * (n_c + n_t)}}$$

76. Where possible, we will calculate odds ratios by relying on 2X2 contingency tables (Lipsey & Wilson, 2001; see Exhibit 11).

Exhibit 11. Estimation of Odds Ratios

Treatment or comparison group	Frequencies	
	Success	Failure
Treatment group	A	B
Comparison group	B	D

77. We will calculate the odds ratio using Equation 6, where \overline{ES} refers to the effect size:

$$(6) \overline{ES} = \frac{ad}{bc}$$

78. In the cases in which we are not able to retrieve the missing data, we will extract or impute effect sizes and associated standard errors based on commonly reported statistics, such as the t or F statistic or p - or Z -values, using David B. Wilson's practical meta-analysis effect-size calculator. In studies that do not report sample sizes for the treatment and the control or comparison group, we will assume equal sample sizes across the groups.

Risk of bias assessment

79. We will determine the rigor of the quantitative studies using an adaptation of a set of criteria to assess risk of bias in experimental and quasi-experimental studies (Hombrados & Waddington, 2012). We will assess the risk of selection-bias and confounding based on quality of identification strategy to determine causal effects and assessment of equivalence across the beneficiaries and non-beneficiaries.

80. Assessing the risk of selection-bias will enable us to examine whether studies have a low, medium, or high risk of selection-bias. Because of the ambitious timeline, we decided not to include assessments of performance, outcome, and analysis reporting bias and other biases in our risk of bias assessment.

Meta-analysis

81. We will pool the results of the quantitative studies that focus on the effects of violence reduction interventions using meta-analysis for each combination of outcome measures and intervention type that includes five or more studies. We will conduct separate meta-analyses for the different outcome measures and separate meta-analyses by intervention. We will examine the heterogeneity of the effect sizes for each outcome across studies.

Narrative synthesis

82. We will report the results following guidelines for systematic review without meta-analysis (SwIM) in cases where a combination of outcome measures and group type only results in four or fewer studies (Campbell et al., 2020). This will involve providing a rationale for grouping studies for the synthesis (based on group type and outcome measure as discussed above), describing the effects and the synthesis methods, and providing a description of the criteria used to prioritize results for the summary and synthesis as well as an investigation of the heterogeneity in the reporting of the effects. AIR will also examine linkages between the different initiatives included in the review when such linkages are discussed in the primary studies.

Performance and Process Evaluation Synthesis

83. This section describes the approach to synthesizing performance and process evaluations, including quality review, mapping, coding, and thematic analysis.

Critical appraisal of performance and process evaluations

84. UN agency evaluation offices have already quality assessed evaluation reports using their respective quality assessment tools. We will include evaluations that

were rated Highly Satisfactory or Satisfactory (or equivalent ratings in different UN evaluation office quality assessment systems).

85. We developed a **quality review protocol** for the critical appraisal of qualitative research and evaluation for evaluations that do not have an existing rating. This tool combines 16 overall questions and 35 sub-questions from quality appraisal approaches across the UN system, including UNICEF’s Global Evaluation Reports Oversight System (UNICEF, 2020), UNDP’s evaluation quality assessment (UNDP Independent Evaluation Office, 2021), United Nations Environment Programme’s evaluation criteria and ratings (UNEP, n.d.), UNFPA’s quality assurance and assessment tools (UNFPA, 2020), WFP’s evaluation quality assurance system (WFP, 2020), IOM’s evaluation quality control tool (IOM, 2022), and the UN Peace Building Support Office’s Evaluation Quality Assessment tool. Exhibit 12 shows the overarching categories included in the tool, included in full in Annex B.

Exhibit 12. Quality Appraisal Categories for Qualitative Studies

Section	Category
A	Intervention, context, and key stakeholders
B	Evaluation purpose, objectives, and scope
C	Evaluation design and methodology
D	Evaluation findings
E	Evaluation conclusions and lessons learned
F	Recommendations

86. We will rate each category on a scale of High (mentioned and well explained), Medium (mentioned, but missing at least one element), or Low (alluded to, but not described in full or explicitly). The team will determine a cut-off for inclusion in the evidence synthesis after the review of all studies is complete. Cut-off points vary according to the needs of the study (Noyes et al., 2023). For example, we may include all studies that rate “High” on the purpose, design, and findings categories, as the detail and quality of recommendations may vary by type of study.

Sampling and saturation

87. Depending on the results of the screening, we may sample the included studies to further narrow the scope. We will ensure geographic and agency variation in the included studies. However, it may be necessary to assess the extent to which various intervention categories have enough studies to ensure depth of coverage. For example, if we find only two performance and process evaluations of peace

messaging and media initiatives, we may decide to eliminate those two studies to focus efforts on synthesizing evidence related to initiatives with more evidence. We will note the lack of evidence in those areas as part of an evidence gap map.

88. Further, if we identify many performance and process evaluations, we may apply principles of saturation. Researchers have argued that qualitative synthesis should be grounded in principles of qualitative research rather than simply trying to transfer quantitative synthesis approaches (e.g., Booth, 2019). Saturation is an example of this principle, in which reviews can apply the saturation principle to limit the number of studies coded for the synthesis. Qualitative synthesis increasingly applies the saturation principle using stratification or purposeful sampling (e.g., Hennegan et al., 2019; Rohwer et al., 2021; Jain et al., 2022; UNEG, forthcoming). The number of studies will depend on the final list of included studies, their focus, and the available time for analysis.

89. Because we cannot make such sampling and saturation decisions until after understanding the evidence available, we will discuss these decisions in consultation with the Global SDG Synthesis Coalition, MG, and TAP.

Qualitative evidence synthesis

90. To analyse content we will import into NVivo the full-text PDFs of all qualitative studies that meet the inclusion criteria and pass the critical appraisal. Importing the full texts will enable reviewers to account for the context of the study and any characteristics that may have influenced the implementation of an intervention.

91. Data extraction. Members of the synthesis team will focus on extracting relevant second-order data from the sections on the theoretical framework, findings, and author's conclusions into a coding framework that uses a combination of deductive (top-down) and inductive (bottom-up) codes. Using a deductive approach, we will apply predetermined codes to the data derived from similar syntheses of process and performance evaluations, particularly using lessons from the synthesis of SDG-17 (de Hoop et al., 2023). Predetermined codes may include application of theories and logic models underpinning the interventions; contextual factors affecting the implementation; and barriers to and facilitators of delivery, participation, and engagement to account for perspectives and experiences of different stakeholders (Johansson et al., 2022). Exhibit 13 shows the draft deductive synthesis framework.

92. Intercoder reliability. We will specifically define each of the predetermined codes and provide examples whenever possible. To ensure intercoder reliability, all team

members will pilot the deductive framework by coding the same two evaluations. The team members will compare coding to discuss any inconsistencies in the understanding of codes and the type of data extracted from each of the pilot studies. During this process researchers will refine the codebook itself, the definitions of deductive codes, and the process for adding inductive codes throughout the coding process. In addition, researchers will meet regularly throughout the coding process to discuss any questions, emerging themes, or divergences from the agreed approach.

93. Thematic analysis and synthesis. Our approach to inductive coding will enable us to identify thematic patterns, emergent themes, and notable outliers from the data to answer the synthesis questions. Specifically, researchers will nest emerging findings within each of the deductive codes as patterns emerge in the data. Key to this approach is using open coding, in which researchers keep an ongoing log of observations as patterns emerge. Using these inductive codes, we will synthesize findings across studies about what works and what does not work in violence prevention and peace-building. These inductive codes serve as the basis for the study findings.

Exhibit 13. Draft Deductive Synthesis Framework

Thematic Category	Codes	Sub-codes
Inputs	Problem analysis* Theory of Change* Intervention type* Relevance of design <i>Extent to which the objectives and design respond to global, country, and institutional needs, policies, and priorities, and adapt if circumstances change</i>	Sub-codes include 3ie categories and interventions; code details on intervention inputs Direct response to participant needs Consideration and inclusion of local or disadvantaged groups* Inclusion of gender considerations in design
Outcomes*	Effectiveness <i>Extent to which intervention achieved its objectives and results, including</i>	Sub-codes include intermediate outcomes. Text will be coded on perceived achievement of outcomes.

Thematic Category	Codes	Sub-codes
	<i>differential results across groups</i>	
Perceived Impacts* <i>Extent to which the intervention generated positive or negative higher-level effects.</i>	Final outcomes	Sub-codes include final outcomes/impacts (e.g., homicides, conflict-related deaths). Text will be coded on perceived achievement of final outcomes/impacts.
Sustainability* <i>Extent to which the net benefits of the intervention are likely to continue.</i>	Planning Financial Ownership Challenges	
Moderators	Partnerships	Inhibiting Facilitating
	Outputs <i>Only code if indicator moderates the evaluated initiative</i>	Public awareness and stakeholder engagement Quality of delivery Coherence <i>Compatibility of the intervention with others in a country, sector, or institution</i>
	Institutional	Internal coordination and management (i.e., capacity) Efficiency <i>Extent to which the intervention delivered results in an economic and timely way</i> Transparency Accountability
	External and contextual	Economic Governance Socio political Culture Infrastructure

Notes: Italicized code definitions derived from OECD-DAC criteria (Organisation for Economic Co-operation and Development, 2000). Categories denoted by an asterisk will also be included in Excel indicator list.

Development of Evidence Gap Maps

94. Based on the results of the critical appraisal and evidence synthesis, we will develop evidence gap maps (EGMs) to highlight categories relevant to SDG-16 targets 16.1 and 16.4 with a limited evidence-base.

95. We will follow guidelines for 3ie EGMs adapted to this synthesis (Snilstveit et al., 2017), and we will use EPPI Reviewer to create an interactive EGM. Since we will have already coded the information in EPPI, we will first create a matrix using the phenomena of interest and evaluation concepts from the SPIDER criteria on each axis. We will base the phenomena of interest and evaluation concepts on the results of the evidence syntheses. Second, using the information from the critical appraisal, we will map each of the included reports and evaluations on the intersections of phenomena of interest–evaluation combinations. If a study has more than one evaluation focus, we will include the results in multiple intersections.

Limitations

96. Exhibit 14 presents identified limitations and mitigation strategies.

Exhibit 14. Limitations and Mitigation Strategies

Limitation	Mitigation Strategy
<p>Ambitious timeline. The team will not be able to conduct a full systematic review in the time required to generate preliminary lessons before the Summit of the Future.</p>	<p>We will address this limitation by conducting a rapid review, using AI to speed screening, and potentially employing sampling and saturation to limit the final included studies.</p>
<p>Searches. We will only conduct searches in a limited number of databases, which will limit the comprehensiveness of the review to some extent.</p>	<p>We will address this limitation by limiting the scope as discussed in previous sections and relying on existing EGMs and syntheses on related topics.</p>
<p>Lack of effect size calculations for each programme. We will only conduct effect size calculations for a limited number of intervention–outcome categories for which five or more studies are available.</p>	<p>We will address this limitation by focusing meta–analyses on those intervention–outcome combinations for which five or more studies are available</p>
<p>Language. We will primarily include evaluations in English</p>	<p>We will include a sub–sample of evaluations in French and Spanish when available and relevant.</p>
<p>Access to evaluations. Not all evaluations on the topic may be publicly available.</p>	<p>We will request MG and TAP for suggestions for evaluations, including those that are not publicly available.</p>

Work Plan

97. During the inception phase in March and April 2024, AIR conducted virtual meetings with the Global SDG Synthesis Coalition, MG, and TAP to discuss key decisions related to the scope and process of the synthesis. AIR refined the methodological protocol and scope of the synthesis, incorporating key decisions from the inception phase. With the draft methodological protocol complete, we will work on the synthesis of impact, performance, and process evaluations from May to July. We plan to finalize a draft consolidated report by 6 September and a final consolidated synthesis report by 30 September 2024.

Exhibit 15. Work Plan

Phase	Activity	Month						
		Mar	Apr	May	June	July	Aug	Sept
Inception and scoping	Contract signing	■						
	Kick-off meeting	■						
	Test and refine search strings	■						
	Prepare a document with key methodological decisions by 4 April meeting	■	■					
	Select SPIDER criteria by 19 April		■					
	Prepare and submit an abbreviated (approx. 15-page) methodological protocol by 19 April; comments by 26 April		■					
	Meeting to discuss comments and reach agreement, if needed (30 April)		■					
	Deliverable 1: Final methodological protocol (by 3 May)			■				
Screening and analysis	Conduct database searches			■				
	Screen resulting studies			■				
	Status update ¹² for the Management Group and TAP for High-Level Political Forum (HLFP) – by the first week of July				■			

¹² Status updates to include description of progress on screening and coding, illustrative important studies, database results, number of studies moved to full screening, etc.

Phase	Activity	Month						
		Mar	Apr	May	June	July	Aug	Sept
	Coding							
	Analysis							
	Deliverable 2: Status update close to the HLFP (8–17 July)							
Reporting	Report writing							
	Preparation for presentation							
	Draft presentation for Summit of the Future – mid-August							
	Deliverable 3: Final presentation of emerging findings for Summit of the Future, including draft Evidence-Gap Map – late August							
	Submit draft synthesis report (first week of September)							
	Review and commenting							
	Deliverable 4: Final synthesis report, policy brief, and evidence gap map							

Quality Assurance

98. We will submit each deliverable after quality assurance by Dr. Ozen Guven. Dr. Guven specializes in qualitative evaluation and has contributed to multiple evidence syntheses, including the synthesis of SDG-17 conducted for the Global SDG Synthesis Coalition. In addition, we will pilot each screening, coding, and quality appraisal tool with two reviewers to ensure a consistent approach.

Coordination with Management Group and Technical Advisory Panel

99. We will also continue to closely coordinate with the MG and TAP. MG and TAP consist of representatives of UN organizations and Member States who have expertise and interest in the synthesis of SDG-16, as well as the Global SDG Synthesis Coalition generally. Together with the Global SDG Synthesis Coalition and the AIR team, MG and TAP provide inputs into the scope and review each deliverable. Specifically, TAP will provide methodological and content expertise to the evidence

synthesis team and conduct quality assurance of the deliverables. They will also support the development of relevant lessons based on rigorous evidence that can be presented during the United Nations General Assembly Summit of the Future.

References

- Anderson, C. A., & Delisi, M. (2011). "Implications of global climate change for violence in developed and developing countries," in J. Forgas, A. Kruglanski, & K. Williams (eds.), *The psychology of social conflict and aggression*, pp. 249–265 (Psychology Press).
- Armsteadt T., Wilkins N., Nation M., (2019). "Structural and social determinants of inequities in violence risk: A review of indicators," *Journal of Community Psychology* 49 (4), pp. 878–906, <https://doi.org/10.1002/jcop.22232>.
- Bagby, E., Murray, N., Felix, E., Liuzzi, S., Alldredge, J. M., Ingwerson, N., Abarcar, P., & Aponte, A. (2021). "Evidence review: The effect of education programs on violence, crime, and related outcomes" (USAID), https://pdf.usaid.gov/pdf_docs/PA00XFMG.pdf.
- Blattman, C. (2023). *Why we fight: The roots of war and the paths to peace* (Penguin).
- Booth, A. (2019). "Harnessing Energies, Resolving Tensions: Acknowledging a Dual Heritage for Qualitative Evidence Synthesis," *Qualitative Health Research* 29(1), pp. 18–31, doi:10.1177/1049732318808247.
- Brody, C., de Hoop, T., Vojtkova, M., Warnock, R., Dunbar, M., Murthy, P., & Dworkin, S.L. (2015). "Economic self-help group programs for improving women's empowerment? A systematic review," *Campbell Systematic Reviews* 11(1), pp. 1–182.
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S. et al. (2020). "Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline" (The BMJ); 368 :l6890 doi:10.1136/bmj.l6890
- Cano, I., & Rojido, E. (2016). "Mapping of homicide prevention programs in Latin America and the Caribbean" (Laboratório de Análise da Violência, Universidade do Estado do Rio de Janeiro and Fórum Brasileiro de Suburbanca Publica).
- Carbonari F., Willman A., Manolio F., Reinach S., and Marques D., (2020). "A Review of the Evidence and a Global Strategy for Violence Prevention," https://www.sdg16hub.org/system/files/2020-10/6c192f_f6036b2b1ecf4fd1a3d7687ff7098a46.pdf.

- Carthy, S. L., Doody, C. B., Cox, K., O’Hora, D., & Sarma, K. M. (2020). “Counter-narratives for the prevention of violent radicalism: A systematic review of targeted interventions,” *Campbell Systematic Reviews* 16(3), <https://doi.org/10.1002/cl2.1106>.
- Cramer, C., Goodhand, J., Morris, R., Pérez-Niño, H., Petrini, B., & Rogers, J. (2016). “Evidence synthesis: What interventions have been effective in preventing or mitigating armed violence in developing and middle-income countries?” (Department for International Development).
- Das, J., Do, Q-T., Shaines, K., & Srikant, S. (2013). “U.S. and them: The geography of academic research,” *Journal of Development Economics* 105, pp. 112–130.
- de Hoop, Thomas, Andi Coombes, Averi Chakrabarti, Marlous de Milliano, Anna Warren, Chinmaya Holla, Ozen Guven, Beth Anne Card, Philomena Panagoulas, Melissa Rodgers, et al. (2023). “What works to accelerate progress on the Partnership Pillar of the Sustainable Development Goals: A synthesis of evaluative evidence” (New York: The Global SDG Synthesis Coalition), <https://www.sdgssynthesiscoalition.org/pillar/partnership-pillar>.
- General Assembly Economic and Social Council (2023). “Progress towards the Sustainable Development Goals: Towards a rescue plan for people and planet: Report of the Secretary-General (Special Edition),” <https://hlpf.un.org/sites/default/files/2023%2004/SDG%20Progress%20Report%20Special%20Edition.pdf>.
- Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019). “An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks,” *Methods in Ecology and Evolution* 10(10), pp. 1645–54.
- Global Alliance for Reporting Progress on Peaceful, Just and Inclusive Societies (2019). “Enabling the implementation of the 2030 Agenda through SDG 16+: Anchoring peace, justice and inclusion” (United Nations), <https://www.sdg16hub.org>.
- Hennegan, J., Shannon, A. K., Rubli, J., Schwab, K. J., & Melendez-Torres, G. J. (2019). “Women’s and girls’ experiences of menstruation in low- and middle-income

- countries: A systematic review and qualitative metasynthesis," *PLoS Med*16(5): e1002803. <https://doi.org/10.1371/journal>.
- Hirvonen, K. (2020). "This is US: Geography of Evidence in Top Health Economics Journals," *Health Economics* 29(10), pp. 1316–23.
- Hombrados, J., & Waddington, H. (2012). "Internal validity in social experiments and quasi-experiments: An assessment tool for reviewers" (Mimeo, 3ie).
- Institute for Security Studies (2021). "What is violence prevention" (The Violence Prevention Forum: Policy Brief), https://www.violence-prevention.org/docrepo/policybrief-161_2%20September%202021.pdf.
- International Organization for Migration (2019). "International Migration Law: Glossary on Migration," https://publications.iom.int/system/files/pdf/iml_34_glossary.pdf.
- International Organization for Migration (2022). "Guidance on Quality Management of IOM Evaluations," <https://evaluation.iom.int/sites/g/files/tmzbd1151/files/documents/Guidance%20on%20quality%20management%20of%20evaluations.pdf>.
- Jain, S., Skinner, G., Lakshminarayanan, M., Alva, D., Bond, A., Oprea, E. & White, H. (2022). "An Evidence and Gap Map of research and literature related to systems of support for children and young people who are at risk of or involved in violence," <https://youthendowmentfund.org.uk/wp-content/uploads/2022/06/Systems-EGM-Technical-report-June-2022.pdf>.
- Johansson, J. F., Lam, N., Ozer, S., Hall, J., Morton, S., English, C., Fitzsimons, C. F., Lawton, R., Forster, A., & Clarke, D. (2022). "Systematic review of process evaluations of interventions in trials investigating sedentary behaviour in adults" (The BMJ Open, 12), <https://bmjopen.bmj.com/content/12/1/e053945.abstract>.
- Krug, E., Dahlberg, L., Mercy, J., et al (2002). "World report on violence and health" (Geneva: World Health Organization), https://iris.who.int/bitstream/handle/10665/42495/9241545615_eng.pdf?sequence=1.

- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. (Thousand Oaks, CA, United States: SAGE).
- Kapoor-Malhotra, S., Moore, N., Umezawa, H., Sharma, S., Taddei, L., Young S., Gallagher, A., White, H., Sonnenfeld, A., Waddington, H. S. (2023). "Conflict and atrocity prevention strategies for violence reduction and sustainable peace: evidence and gap map."
- Mares, D. M. & Moffett, K. W. (2016). :Climate change and interpersonal violence: A 'global' estimate and regional inequities," *Climatic Change* 135, pp. 297–310, <http://dx.doi.org/10.1007/s10584-015-1566-0>.
- OECD/UNCDF (2020). *Blended Finance in the Least Developed Countries 2020: Supporting a Resilient COVID-19 Recovery* (OECD Publishing), <https://doi.org/10.1787/57620d04-en>.
- Reilly, R. (2010). "Process tracing," in A. J. Mills, G. Durepos, and E. Wiebe (eds.), *Encyclopedia of Case Study Research* (London: SAGE).
- Rohwer, A., Hendricks, L., Oliver, S., & Garner, P. (2021). "Testing for saturation in qualitative evidence syntheses: An update of HIV adherence in Africa," *Plos one* 16(10), e0258352.
- Sachs, J. D., Lafortune, G., Fuller, G., & Drumm, E. (2023). "Implementing the SDG Stimulus: Sustainable Development Report 2023" (Dublin: Dublin University Press), 10.25546/102924.
- Sardinha, L., Maheu-Giroux, M., Stöckl, H., Meyer, S. R., & García-Moreno, C. (2022). "Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018," *Lancet* Feb 6; 399(10327):803–813. doi: 10.1016/S0140-6736(21)02664-7. PMID: 35182472; PMCID: PMC8885817.
- Shafiq, M. N. & Sinno, A. H. (2010). "Education, income, and support for suicide bombings: Evidence from six Muslim countries," *Journal of Conflict Resolution* 54(1).
- Snilstveit, B., Gallagher, E., Phillips, D., Vojtkova, M., Eyers, J., Skaldiou, D., et al. (2017). "PROTOCOL: Interventions for improving learning outcomes and access to

education in low- and middle-income countries: A systematic review," *Campbell Systematic Reviews* 13(1), 1–82. <https://doi.org/10.1002/cl2.176>.

Statista Research Department (2023). "Number of fatalities due to terrorist attacks worldwide between 2007 and 2022," *Statista*, <https://www.statista.com/statistics/202871/numberof-fatalities-by-terrorist-attacks-worldwide/>.

Sonnenfeld, A., Chirgwin H., Berretta M., Longman K., Krämer M., & Snilstveit, B. (2020). "Building peaceful societies: An evidence gap map." 3ie Evidence Gap Map Report 15 (New Delhi: International Initiative for Impact Evaluation), doi: <https://doi.org/10.23846/EGM015>.

UNEG (forthcoming). "Evaluation Synthesis Guidance."

UNODC, OHCHR, and UNDP (2023). "Global progress report on Sustainable Development Goal 16 indicators: A wake-up call for action on peace, justice and inclusion," <https://www.undp.org/policy-centre/oslo/publications/globalprogress-report-sdg-16>.

UNDP Independent Evaluation Office (2021). "Evaluation Quality Assessment," <http://web.undp.org/evaluation/guideline/documents/PDF/section-6.pdf>.

UNEP (n.d.). "Methodology for SDG-indicator 17.14.1: Mechanisms in place to enhance policy coherence for sustainable development."

UNFPA (2020). "Quality Assurance and Assessment Tools and Guidance," <https://www.unfpa.org/admin-resource/evaluation-quality-assurance-and-assessment-tools-and-guidance>.

UNICEF (2020). "Global Evaluation Report Oversight System Guidance Handbook," <https://www.unicef.org/evaluation/media/2571/file/2020%20GEROS%20Handbook.pdf>.

United Nations Evaluation Group (2024). "UNEG Evaluation Synthesis Guidance."

United Nations (2023). "Our Common Agenda Policy Brief 9: A New Agenda for Peace," <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf>.

- United Nations and World Bank (2017). "Pathways for peace: Inclusive approaches to preventing violent conflict," (World Bank), <https://doi.org/10.1596/978-1-4648-1162-3>.
- UNODC (2023). "Global study on homicide 2023," https://www.unodc.org/documents/dataand-analysis/gsh/2023/Global_study_on_homicide_2023_web.pdf.
- Uppsala CONFLICT Data Program (2023). "Deaths in armed conflicts, world" (Our World in Data), <https://ourworldindata.org/grapher/deaths-in-armed-conflicts-by-type>.
- Waller, I. (2021). "Reducing violent crime by 50% before 2030: Decisive action now to achieve these SDGs," in H. Kury & S. Redo (eds.), *Crime prevention and justice in 2030* (Springer), https://doi.org/10.1007/978-3-030-56227-4_26.
- White, H. and Phillips, D. (2012). "Addressing Attribution of Cause and Effect in Small n Impact Evaluations: Towards an Integrated Framework," Working Paper 15, International Initiative for Impact Evaluation, <https://www.3ieimpact.org/file/7911/download?token=unt2c-N2>.
- World Food Programme (2020). "Evaluation Quality Assurance System," <https://www.wfp.org/publications/eqas-evaluation-quality-assurance-system-0>.

Annex A. Search Terms

Final SPIDER search strings to identify impact studies:

- **Sample:** *Humanitarian OR emergency OR conflict OR crisis OR disaster OR conflict-ridden OR conflict-affected OR crisis-affected OR "fragile state" OR "fragile country" OR "low income countr*" OR "low-income countr*" OR "low-income econom*" OR "low income econom*" OR "lower-middle-income countr*" OR "lower middle income countr*" OR "lower-middle-income econom*" OR "lower middle income econom*" OR "middle income countr*" OR "middle-income countr*" OR "middle-income econom*" OR "middle income enconm*" OR "developing countr*" OR "less developed countr*" OR "less-developed countr*" OR "underdeveloped countr*" OR "under developed countr*" OR "under-developed countr*" OR "underserved countr*" OR "under served countr*" OR "under-served countr*" OR "LMIC*" OR "low GDP" OR "low-GDP" OR "low GNP" OR "low-GNP" OR "fragile state" OR "third world" OR "transitional countr" OR "high burden countr*" OR "high-burden countr*" OR Asia* OR "South Asia*" OR "Africa*" OR "Latin America*" OR "South America*" OR "Central America*" OR "Middle East*" OR "sub-Saharan Africa*" OR "sub Saharan Africa*" OR Caribbean OR "West Indies" OR Afghanistan* OR Afghan* OR Albania* OR Algeria* OR "American Samoa*" OR Angola* OR Argentin* OR Armenia* Or Azerbaijan* OR Azeri OR Bangladesh* OR Belarus* OR Belize* OR Benin* OR Bhutan* OR Bolivia* OR Bosnia* OR "Bosnia and Herzegovina" OR Botswana OR Motswana OR Brazil* OR Bulgaria* OR "Burkina Faso" OR Burkinabè OR Burkinabe OR Burundi* OR "Cabo Verde*" OR "Cape Verde*" OR Cameroon* OR Cambodia* OR "Central African Republic" OR "Central African" OR Chad* OR China OR Chinese OR Colombia* OR Comoros OR Comorian OR "Cote d'Ivoire" OR "Ivory Coast" OR Ivorian OR Congo* OR "Costa Rica*" OR Cuba* OR "Democratic Republic of Congo" OR "Republic of Congo" OR "Democratic People's Republic of Korea" OR "North Korea*" OR Korea* OR Djibouti* OR Dominica* OR "Dominican Republic" OR Ecuador* OR Egypt* OR "Arab Republic of Egypt" OR "El Salvador" OR Salvador* OR Eritrea* OR Eswatini OR Swazi OR Ethiopia* OR "Equatorial Guinea*" OR Equatoguinean OR Fiji* OR Gabon* OR Gambia* OR Gaza* OR Palestin* OR Georgia* OR Ghana* OR Grenada OR Granad* OR Guatemala* OR Guam* OR Guinea* OR "Guinea-Bissau" OR Guyan* OR Haiti* OR Hondura* OR India* OR Indonesia* OR Iran* OR "Islamic Republic of Iran" OR Iraq* OR Jamaica* OR Jordan* OR Kazakhstan* Or Kazakh* OR Kenya* OR Kiribati OR "I-Kiribati" OR Kosovo OR Kosova* OR Kyrgyz* OR Lao* OR Lao PDR OR Lao People's Democratic Republic OR Lebanon OR Leban* OR Lesotho OR Mosotho OR Basotho OR Liberia* OR Libya* OR Madagascar OR Malagasy OR Malawi* OR Malaysia* OR Maldives OR Maldivian OR Mali* OR "Marshall Islands" OR Marshallese OR Mauritius OR Mauritian OR Mauritania* OR Mexic* OR Micronesia* OR "Federated States of Micronesia" OR Moldova* OR Mongolia* OR Montenegr* OR Morocc* OR Mozambique OR Mozambican OR Burma OR Burmese OR Myanmar OR Myanma* OR Namibia* OR Nepal* OR Nicaragua* OR Niger* OR Nigeria* OR "North Macedonia" OR Macedonian OR Palau* OR Pakistan* OR Paraguay* OR Peru OR Philippines OR Philipines OR Phillipines OR Phillipines OR Filipino OR "Papua New*

Guinea* OR "Republic of Congo" OR "Republic of Korea" OR "South Korea*" OR Rwanda OR Rwand* OR "Russian Federation" OR Russia* OR Samoa* OR "Sao Tome and Principe" OR "São Tomé*" OR "Sao Tome*" OR Santomean OR "SãoToméan" OR Senegal* or Serbia* OR "Sierra Leone*" OR "Sri Lanka*" OR "Solomon Island*" OR Somalia* OR "South Africa*" OR "South Sudan*" OR Sudan* OR "St. Lucia" OR "Saint Lucia*" OR "St. Vincent" OR "Saint Vincent and the Grenadines" OR "St. Vincent and the Grenadines" OR "Vincentian and Grenadian" OR Vincy OR Swaziland OR Emaswati OR Liswati OR Suriname* OR Syria* OR "Syrian Arab Republic" OR Tajikistan* Or Tajik OR Tanzania* OR Thailand OR Thai OR "Timor-Leste" OR "Timor Leste" OR "East Timor*" OR Timorese OR Maubere OR Tokelau* OR Togo* OR Tonga* OR Tunisia* OR Turkey OR Turkish OR Turkiye OR Turk OR Turkmenistan* Or Turkmen* OR Tuvalu* OR Uganda* OR Ukraine OR Ukrainian OR Uzbekistan OR Uzbek OR Vanuatu* OR "Ni-vanuatu" OR Vietnam* OR "Viet Nam" OR "West Bank" OR Gaza* OR Yemen* OR "Republic of Yemen*" OR Zambia* OR Zimbabwe* OR Zimbo

- **Phenomenon of Interest:** "peace educat*" OR "peace messag*" OR "peace medi*" OR "dispute resol*" OR "mental health" OR "psychosocial" OR "psycho-social" OR "behavioral therapy" OR "behavioural therapy" OR "cognitive behavioral" OR "cognitive behavioural" OR "mental health service" OR "mental health treatment" OR "mental healthcare" OR "social inclusion" OR "reintegrat*" OR "re-integrat*" OR "intergroup dialo*" OR "inter-group dialo*" OR "peace process*" OR "peace negotiat*" OR "peace agree*" OR "peace implement*" OR "transitional justice process*" OR "peace polic*" OR "peacekeeping" OR "peace-keeping" OR "disarmament*" OR "demobili*" OR "gang dropout" OR "gang drop-out" OR "violen* extrem*" OR "demin*" OR "policing" OR "police" OR "prevent* protect*"
- **Design:** evaluation OR "impact evaluation" OR "impact analysis" OR "random* control* trial" OR RCT OR experiment* OR "quasi-experiment*" OR "regression discontinuity" OR "difference-in-difference*" OR "difference in difference*" OR "propensity score" OR "evidence synthesis" OR "quasi random" OR "quasi-random" OR "instrumental variable*" OR "random* eval*" OR "random* assign*" OR "interrupted time series" OR "ITS"
- **Evaluation:** conflict OR war OR battle OR violen* OR "armed clash" OR insurgen* OR killing* OR paramilitarism OR guerrilla OR kidnapping OR "war crime" OR abuse OR torture OR exploitation OR trafficking OR refugee* OR displace* OR IDP OR exile* OR "asylum seeker*" OR "forced migration" OR homicid* OR "use of force" OR brutality OR crackdown OR persecution OR vigilantism OR "atrocit*" OR genocide OR "ethnic cleansing" OR shelling OR "bomb*" OR explosion OR IED OR casualties OR "child soldier" OR combatant OR "rebel*" OR uprising OR riot "enforced disappearance" OR "arbitrary detention" OR "arbitrarily detain*" OR "physical punishment" "psychological aggression against children" OR "unsentenced detention" OR "unsentenced detain*" OR "illicit financial flow*" OR "illicit arms flow*" OR "arms proliferation" OR "organized crime" OR peace OR "Paris principl*" OR "conflict resolution" OR "dispute resolution" OR amnesty OR disarmament OR DDR OR ceasefire OR security OR "rights violation" OR "social cohesion" OR "lawlessness" OR "rebellion" OR "property right*" OR "toleran*" OR "criminal organisation" OR "criminal organization" OR "criminal association" OR "organized crime" OR "organised crime" OR mafia OR "crim* network*" OR dto* OR "drug trafficking organ*" OR "drug cartel*" OR "crim*

group OR "crim* cartel" OR "undeclared capital" OR "undeclared income" OR "undeclared profit*" OR "evade tarif*" OR "criminal proceeds" OR "corrupt payment*" OR "drug law enforcement" OR "drug crime*" OR "drug gang*" OR "smuggl*" OR "traffick*" OR "black market*" OR "peace" OR "rule of law" OR "stabili*" OR "solidar*"*

3ie Database

We conducted 3ie searches on 26 April 2024 for all hits that returned from publication date of 1 January 2019, through 1 April 2024. The following list of search terms should be searched for Title, and Abstract fields only (using the search string below specific for the 3ie database).

Phenomenon of Interest/Design terms in all searches:

(title:(("peace educat*" OR "peace messag*" OR "peace medi*" OR "dispute resol*" OR "mental health" OR "psychosocial" OR "psycho-social" OR "behavioral therapy" OR "behavioural therapy" OR "cognitive behavioral" OR "cognitive behavioural" OR "mental health service" OR "mental health treatment" OR "mental healthcare" OR "social inclusion" OR "reintegrat*" OR "re-integrat*" OR "intergroup dialo*" OR "inter-group dialo*" OR "peace process*" OR "peace negotiat*" OR "peace agree*" OR "peace implement*" OR "transitional justice process*" OR "peace polic*" OR "peacekeeping" OR "peace-keeping" OR "disarmament*" OR "demobili*" OR "gang dropout" OR "gang drop-out" OR "violen* extrem*" OR "demin*" OR "policing" OR "police" OR "prevent* protect*") AND (conflict OR war OR battle OR violen* OR "armed clash" OR insurgen* OR killing* OR paramilitarism OR guerrilla OR kidnapping OR "war crime" OR abuse OR torture OR exploitation OR trafficking OR refugee* OR displace* OR IDP OR exile* OR "asylum seeker*" OR "forced migration" OR homicid* OR "use of force" OR brutality OR crackdown OR persecution OR vigilantism OR "atrocit*" OR genocide OR "ethnic cleansing" OR shelling OR "bomb*" OR explosion OR IED OR casualties OR "child soldier" OR combatant OR "rebel*" OR uprising OR riot "enforced disappearance" OR "arbitrary detention" OR "arbitrarily detain*" OR "physical punishment" "psychological aggression against children" OR "unsentenced detention" OR "unsentenced detain*" OR "illicit financial flow*" OR "illicit arms flow*" OR "arms proliferation" OR "organized crime" OR corruption OR bribery OR "pay bribe*" OR "paid bribe*" OR "solicit bribe*" OR "solicited bribe*" OR peace OR "Paris principl*" OR "conflict resolution" OR "dispute resolution" OR amnesty OR disarmament OR DDR OR ceasefire OR security OR "rights violation" OR "social cohesion" OR "lawlessness" OR "rebellion" OR "property right*" OR "toleran*" OR "criminal organisation" OR "criminal organization" OR "criminal association" OR "organized crime" OR "organised crime" OR mafia OR "crim* network*" OR dto* OR "drug trafficking organ*" OR "drug cartel*" OR "crim* group*" OR "crim* cartel" OR "undeclared capital" OR "undeclared income" OR "undeclared profit*" OR "evade tarif*" OR "criminal proceeds" OR "corrupt payment*" OR "drug law enforcement" OR "drug crime*" OR "drug gang*" OR "smuggl*" OR "traffick*" OR "black market*" OR "peace" OR "rule of law" OR "stabili*" OR "solidar*")) OR abstract:(("peace educat*" OR "peace messag*" OR "peace medi*" OR "dispute resol*" OR "mental health" OR "psychosocial" OR "psycho-social" OR "behavioral therapy" OR "behavioural therapy" OR "cognitive behavioral" OR

"cognitive behavioural" OR "mental health service" OR "mental health treatment" OR "mental healthcare" OR "social inclusion" OR "reintegrat*" OR "re-integrat*" OR "intergroup dialo*" OR "inter-group dialo*" OR "peace process*" OR "peace negotiat*" OR "peace agree*" OR "peace implement*" OR "transitional justice process*" OR "peace polic*" OR "peacekeeping" OR "peace-keeping" OR "disarmament*" OR "demobili*" OR "gang dropout" OR "gang drop-out" OR "violen* extrem*" OR "demin*" OR "policing" OR "police" OR "prevent* protect*") AND (conflict OR war OR battle OR violen* OR "armed clash" OR insurgen* OR killing* OR paramilitarism OR guerrilla OR kidnapping OR "war crime" OR abuse OR torture OR exploitation OR trafficking OR refugee* OR displace* OR IDP OR exile* OR "asylum seeker*" OR "forced migration" OR homicid* OR "use of force" OR brutality OR crackdown OR persecution OR vigilantism OR "atrocit*" OR genocide OR "ethnic cleansing" OR shelling OR "bomb*" OR explosion OR IED OR casualties OR "child soldier" OR combatant OR "rebel*" OR uprising OR riot OR "enforced disappearance" OR "arbitrary detention" OR "arbitrarily detain*" OR "physical punishment" OR "psychological aggression against children" OR "unsentenced detention" OR "unsentenced detain*" OR "illicit financial flow*" OR "illicit arms flow*" OR "arms proliferation" OR "organized crime" OR corruption OR bribery OR "pay bribe*" OR "paid bribe*" OR "solicit bribe*" OR "solicited bribe*" OR peace OR "Paris principl*" OR "conflict resolution" OR "dispute resolution" OR amnesty OR disarmament OR DDR OR ceasefire OR security OR "rights violation" OR "social cohesion" OR "lawlessness" OR "rebellion" OR "property right*" OR "toleran*" OR "criminal organisation" OR "criminal organization" OR "criminal association" OR "organized crime" OR "organised crime" OR mafia OR "crim* network*" OR dto* OR "drug trafficking organ*" OR "drug cartel*" OR "crim* group*" OR "crim* cartel" OR "undeclared capital" OR "undeclared income" OR "undeclared profit*" OR "evade tarif*" OR "criminal proceeds" OR "corrupt payment*" OR "drug law enforcement" OR "drug crime*" OR "drug gang*" OR "smuggl*" OR "traffick*" OR "black market*" OR "peace" OR "rule of law" OR "stabili*" OR "solidar*"))

Annex B. Critical Appraisal Tools

Exhibit B-1: Risk of Bias Tool for Experimental and Quasi-Experimental Studies

Ask these questions for all quantitative studies

Are the mean values or the distributions of the covariates at baseline statistically different for the control or comparison group ($p < 0.05$)?

Are these differences controlled for using covariate analysis in the impact evaluation?

Is difference-in-difference estimation used?

If the study is quasi-experimental and uses difference-in-difference estimation, is it showing that the parallel trends assumption is valid?

If the study does not use difference-in-difference, does the study control for baseline values of the outcome of interest (ANCOVA)?

Attrition

Is the attrition rate from the study below 10%?

Is the attrition rate statistically significantly different between the treatment and comparison group?

Sample size

Does the study account for lack of independence between observations within assignment clusters if the outcome variables are clustered?

Is the sample size likely to be sufficient to find significant effects of the intervention?

Ask questions below only for studies that apply randomization

Does the study apply randomized assignment?

Ask questions below only for studies that apply regression discontinuity designs

Is the allocation of the programme based on a pre-determined continuity on a continuous variable and blinded to the beneficiaries or, if not blinded, individuals cannot reasonably affect the assignment variable in response to knowledge of the participation rule?

Ask questions below only for studies that apply matching

Are the characteristics of the treatment and comparison group similar (based on statistical significance tests) after matching?

Ask questions below only for studies that apply instrumental variable estimation

Does the study describe clearly the instrumental variable(s)/identifier used and why it is exogenous?

Are the instruments jointly significant at the level of $F \geq 10$? If an F test is not reported, does the author report and assess whether the R-squared of the instrumenting equation is large enough for appropriate identification (R-sq > 0.5)?

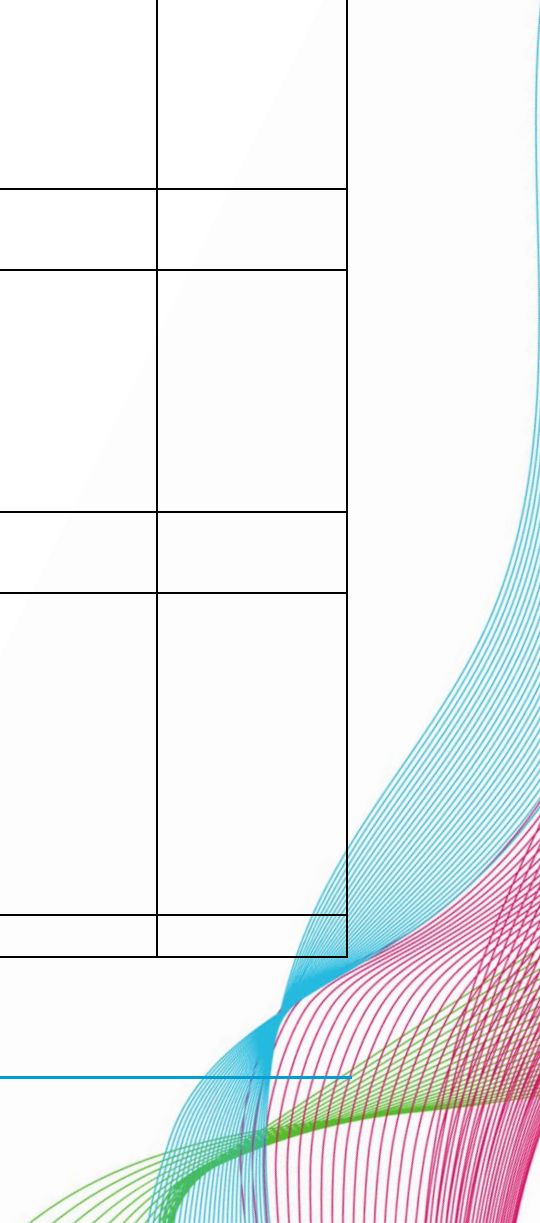
Exhibit B-2: Qualitative Review Protocol

Number	Question	High: <i>mentioned & well explained</i>	Med: <i>Mentioned, but missing at least one element</i>	Low: <i>Alluded to, but not described in full or explicitly</i>	Not Applicable	Not Mentioned	Reason for assessment of H/M/L and supporting text, where necessary
SECTION A:	EVALUATION PURPOSE, OBJECTIVES, AND SCOPE						
Question 1.	Is the purpose of the evaluation clearly described?						
i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.						
ii	Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as, if applicable, the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention).						

Question 2.	Is the Theory of Change, results chain, or logic well articulated?						
i	Clear description of the intervention's intended results, or of the parts of implementation that are applicable to, or are being assessed by, the evaluation.						
SECTION B: EVALUATION DESIGN AND METHODOLOGY							
Question 3.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation?						
i	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.						
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions, and/or other processes from which the analysis can be based and conclusions drawn.						
Question 4.	Does the report specify adequate methods for data collection, analysis, and sampling?						
i	Evaluation design and set of methods is relevant and adequately robust for the evaluation's purpose, objectives, and scope, and are fully and clearly described.						



ii	Qualitative and quantitative data sources are appropriate and are clearly described.						
iii	Sampling strategy is provided. It should include a description of how diverse perspectives are captured (or if not, provide reasons for this), with articulated consideration and/or inclusion of vulnerable/marginalized groups, equity, and intersectionality						
iv	Clear and complete description of the methods of data analysis.						
v	Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).						
Question 5.	Are ethical issues and considerations described?						
ii	Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm).						
SECTION C:	EVALUATION FINDINGS						



Question 6.	Do the findings clearly address all evaluation objectives and scope?						
i	Findings marshal sufficient levels of evidence to systematically address all of the evaluation’s questions, sub-questions, and criteria.						
Question 7.	Are evaluation findings derived from the conscientious, explicit, and judicious use of the best available, objective, reliable, and valid data and by accurate quantitative and qualitative analysis of evidence.						
i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident through the use of multiple data sources.						
ii	Findings are clearly supported by, and respond to, the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question.						
SECTION D:	EVALUATION CONCLUSIONS & LESSONS LEARNED						



Question 8.	Do the conclusions clearly present an objective overall assessment of the intervention?						
i	Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).						
ii	Conclusions are derived appropriately from findings, and present a picture of the strengths and limitations of the intervention that adds insight and analysis beyond the findings.						

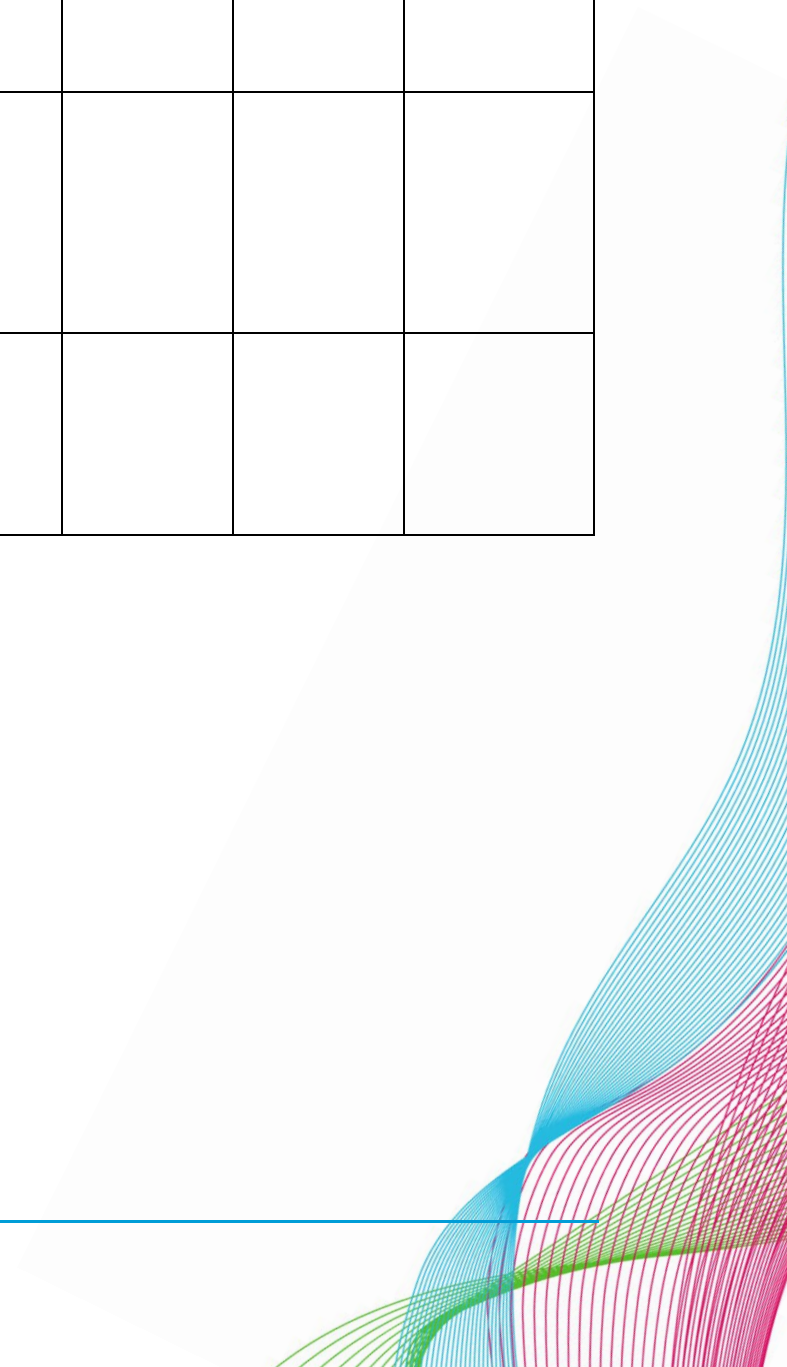


Exhibit B-3: Quality Appraisal Tool for Small-n Impact Evaluations (process tracing, contribution analysis, most significant change, outcome mapping, etc.)

Item	Description	Key	Notes
Design	Does the study use an established small n approach (e.g., contribution analysis, process tracing)?	<p>High: Approach is named with clear descriptions of analytical process actually taken</p> <p>Medium: Approach is named without description of analytical process</p> <p>Low: Established small n approach is not used/unclear</p>	<p>Examples of established small n approach:</p> <p>contribution analysis, process tracing, qualitative comparative analysis, realist evaluation, general elimination methodology</p>
Theory	Is the analysis based on an explicit Theory of Change or theoretical framework?	<p>High: ToC/theoretical framework presented with 3 or more elements listed</p> <p>Medium: ToC/theoretical framework presented</p>	<p>Desirable information : (1) inputs, activities, outputs, intermediate and final intense outcomes, including indicators</p>

Item	Description	Key	Notes
		with 2 or fewer elements Low: ToC/theoretical framework not presented, or refers to external sources	(2) underlying intervention logic and theoretical links (3) programme participants and project-affected people (4) timeline (5) assumptions (6) contextual factors and external influences
Data collection method	Is the data collection method clearly described (e.g., key informant interview, focus group discussion, document review)?	High: Method(s) clearly described, including when it happened, who collected data from whom, survey instrument, etc. Medium: Method(s) named without clear description Low: No description	



Item	Description	Key	Notes
Data analysis process	Are the procedural steps for data analysis clearly defined (e.g., presentation of a coding tool)?	High: Clear description Medium: Unclear description (e.g., analytical principle/strategy named without explanation) Low: No description	
Selection of cases	Is the sample/case selection strategy explained and justified?	High: Explained and justified Medium: Explained, but not justified Low: No explanation and justification	
Analysis	Was triangulation done and is it clear how it was done?	High: Triangulation was done with description of how it was done Medium: Triangulation is said to have been	



Item	Description	Key	Notes
		done but unclear how it was done Low: No triangulation	
Bias	Are potential sources of bias addressed (e.g., alternate explanations for the changes in outcomes observed, reporting bias)?	High: Sources of bias mentioned and addressed, and it is clear how it was addressed Medium: Sources of bias mentioned, but not addressed (e.g., only discussed as limitation) Low: No discussion of bias	



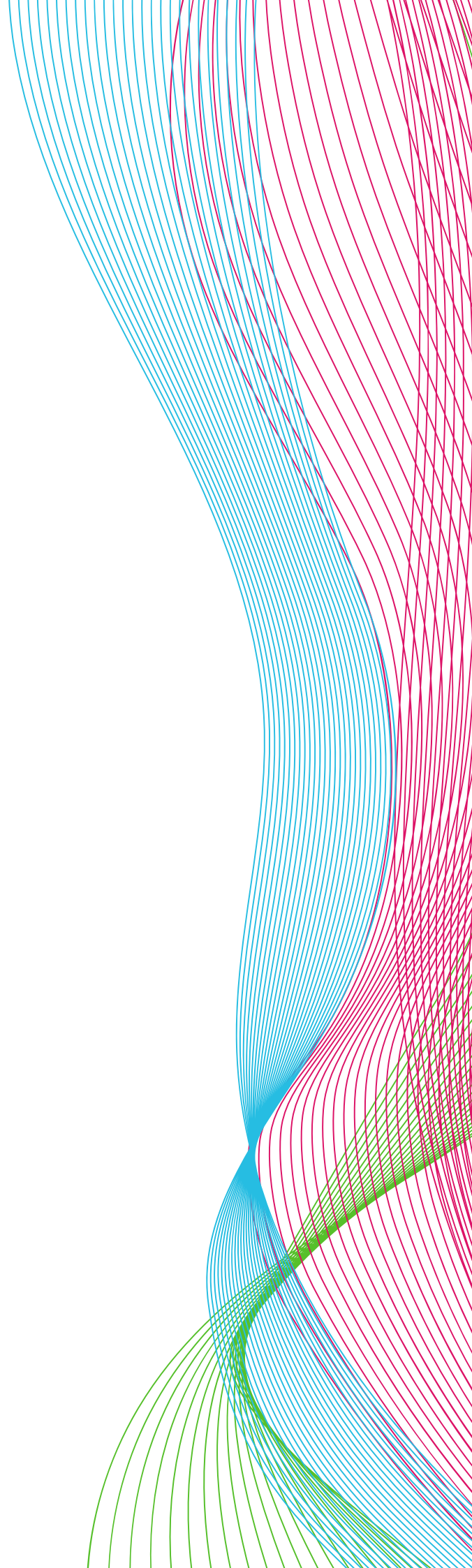


**The Global
SDG Synthesis
Coalition**



AIR[®]

Advancing Evidence.
Improving Lives.



16 PEACE, JUSTICE
AND STRONG
INSTITUTIONS



sdgsynthesiscoalition.org



@SDGSynthesis



The Global SDG Synthesis Coalition