JOURNAL FOR DERADICALIZATION

Spring 2026
No. 46
ISSN: 2363-9849

Research Method Note: Measuring and Evaluating P/CVE Initiatives with Respect to Human Rights and the Roles of Women

Michael J. Williams[a1]
[a]The Science of P/CVE

Abstract
The United Nations Global Counter Terrorism Strategy, adopted by the General Assembly, established respect for human rights and the rule of law as one of its four pillars: a fundamental basis for the fight against terrorism. Although large investments currently are made in a broad range of localized P/CVE programs, there has not been a commensurately large amount of empirically reliable evidence about them. This (mal)practice should change, given that this prospective body of knowledge is instrumental toward designing future prevention policies and programs: including whether/how to upscale localized projects and programs. Additionally, policies that affect women and girls, including policies ostensibly to benefit them, often lack substantial input, if not leadership, from the women and girls that said policies will affect. A gender-responsive approach to P/CVE is congruent with the UN PVE Plan of Action that promotes a wide range of actions, including: human rights and the rule of law, engaging communities, gender equality and empowering women. Social science methods, and associated evaluation research methods, have been designed to facilitate these objectives. The present work offers techniques and concrete recommendations for measurement and evaluation, to support those objectives: before, during, and after data collection.

Article History
Received Sep 17, 2025
Accepted Mar 14, 2026
Published Mar 27, 2026

Keywords: Measurement and Evaluation, M&E, Preventing Violent Extremism, PVE, Countering Violent Extremism, CVE, P/CVE, Human Rights, Women, Gender

Introduction

The United Nations Global Counter Terrorism Strategy, adopted by the General Assembly in 2006, established respect for human rights and the rule of law as one of its four pillars (see resolution 60/288, Shaheed, 2018). Additionally, scholars argue that P/CVE program design should not only uphold human rights and the rule of law, but also incorporate gender equality by ensuring that the roles of women and men are effectively integrated and coordinated within P/CVE programming (Ndung'u & Shadung, 2017). This is because research suggests that

[1] Corresponding Author Contact: Michael J. Williams, Email: m.williams@thescienceofpcve.org, Twitter: @MickWilliamsPhD, ORCID iD https://orcid.org/0000-0001-5630-9814

gender equality and women's empowerment may be key to the design of successful P/CVE programming (Ndung'u & Shadung, 2017; White, 2020).

Although there has been substantial investment currently in a wide array of locally-focused, internationally-funded P/CVE initiatives, there has been substantially less evidence of the development of a "culture of evaluation," among international donors, including the production and communication of empirically-reliable results (Frank, 2017). In short, "these practices need [to] shift quickly, as this body of knowledge will be central to directing the design of future preventive policies and programmes. [As such], donors have a key role to play in investing in these evidence-building activities" (Frank, 2017, p. 10). To support such priorities, the present work offers scientific techniques, and concrete recommendations, for measurement and evaluation (M&E) in support of human rights and the roles of women: before, during, and after data collection.

Scope and Contribution

The present article is intended as a research method brief: a practical synthesis of M&E considerations for P/CVE initiatives, focused on protecting human rights and strengthening women's meaningful participation. It does not introduce new empirical findings, per se; rather, it consolidates methodological and ethical implications and translates them with respect to common P/CVE implementation concerns (e.g., multi-actor delivery, sensitive data environments, and donor-driven reporting cycles). Recent reviews suggest that P/CVE evaluations remain variable in quality and sometimes generate negative downstream effects, which underscores the need for methodologically careful, rights-aware M&E guidance (Charkawi et al., 2024).

Conceptual Anchor

A rights-respecting, gender-responsive evaluation approach is not only a worthy normative requirement; it can also be understood as a mechanism that supports program legitimacy and trust: constructs repeatedly linked to participants' cooperation, disclosures, and sustained engagement with authorities and institutions (Jackson et al., 2012; Madon et al., 2017). It also aligns with a human security lens that shifts attention from state-centric threat

management to the protection of individuals' vital freedoms, safety, and dignity (Moayerian et al., 2025). Simply put, when programs (and their evaluations) are experienced as fair, transparent, and non-stigmatizing, participants and communities have greater reason to engage candidly; conversely, when they are experienced as coercive or extractive, disengagement and underreporting become more likely (Madon et al., 2017). This framework clarifies why human rights and gender safeguards belong inside M&E design rather than as an afterthought.

Methodological Rationale

Several tools referenced in this article (e.g., coalition assessment and scalability frameworks) originate in fields that routinely evaluate complex, multi-actor interventions. The value of such tools for P/CVE lies, not only in assessing programmatic outcomes, but in helping evaluators and other evaluation stakeholders to document, for example, a) the quality of collaboration and accountability structures, and (b) plan for transferability without assuming that impact automatically survives expansion (Milat et al., 2015; Milat et al., 2016). Because P/CVE often occurs in politicized contexts, these—and other such tools—should be treated as starting points and adapted through local stakeholder review and piloting, with explicit attention to unintended consequences such as stigma and distrust (Charkawi et al., 2024).

Advancing Human Rights Through M&E Before Data Collection Data Security[2]

As noted by a recent publication of UN Women, it is vitally important to safeguard the privacy and protection needs of program participants (UN Women, 2021). Therefore, with respect to M&E, such protections should be considered during a program evaluation's design phase. Included in such protections is data security, which should begin with proper training of those responsible for data collection, including attention to data storage: which can be expected to reduce the likelihood of data loss or breaches in data security (Centers for Disease

[2] Because data protection and confidentiality issues cut across multiple stages of evaluation, the recommendations below are stated here in full (e.g., testing for unintended consequences), then referenced in later sections, to avoid unnecessary repetition.

Control and Prevention, 2011). With respect to data collection, this should include preserving participants' anonymity to the greatest extent possible, including de-identifying their data. This can be easily accomplished through methods such as anonymous surveys but is not as easily accomplished through data collections that employ focus groups. For example, even if focus group discussants use only first names during focus group discussions, all (sighted) focus group participants will be aware of the likenesses (if not the identities) of their fellow discussants. A prospective workaround for this problem is to conduct focus groups virtually, and with participants' video capabilities disabled, or by teleconference. Naturally, such technological solutions are not always available; so, discussants should be made aware of the need to preserve one another's confidentiality beyond the focus group itself, and this potential limitation to privacy should be included among any related participation consent document(s).

Data security also includes the need to protect data from search and seizure by law enforcement or other security agencies: a plausible possibility in several regions of the world where P/CVE projects take place. Consequently, it is wise to assume—beginning during an evaluation's design phase—that unauthorized persons will attempt to confiscate data. Therefore, it is important to plan how, when, and how frequently data will be backed up (e.g., on a cloud-based server). Thus, even if local copies of data are confiscated, the project team will have preserved the backed-up data.

Similarly, it is prudent to assume that data collection staff will be questioned by authorities about the nature of the project. Consequently, data collection staff should be trained in how to respond to such questioning. Specifically, consider how the project might be described in honest, though general, terms that would not be objectionable to authorities. For example, a P/CVE project that intends to work with youth who are considered at risk of recruitment to violent extremist organizations might honestly be said to be engaging in "positive youth development," instead of P/CVE per se.

Duty to Report

Despite the aforementioned protections regarding data security, still, it is possible that ethical dilemmas could arise during data collection, if proper disclosures are not made to participants in advance. For example, during an evaluation project, if data collection staff

gain information about an impending attack, they would be caught between their duty to protect participants' confidentiality, and the project's overarching mission to prevent such violence (Williams & Kleinman, 2013). To circumvent such dilemmas, the project's informed consent document(s) should include disclosures that make clear to participants that project staff would be required to disclose any information, to the proper authorities, that they reasonably believe is suggestive of pending illegal activities (ibid.).

Plan to Test for Unintended Consequences

The notion of "do no harm" should not be lip service but can and should be put to empirical scrutiny. Specifically, unintended consequences should be anticipated to the best of the abilities of the project team, and M&E components should be built into the evaluation design to measure and subsequently analyse the extent to which any of the measured unintended consequences were manifested as a result of a given P/CVE project. Ideally, any such tests for unintended consequences should be built into pilot tests of the P/CVE project, to identify them in a stage early enough to revise the project if/as necessary.

For example, strategies to prevent violent extremism might inadvertently alienate various communities of faith, undermine the enjoyment of various fundamental freedoms, or run contrary to building tolerance or mutual understanding within a given intended population, and public trust might be eroded (Shaheed, 2018). Any/all of such outcomes could be measured, alongside other key outcomes of interest (for example as part of a pre-post or comparison-group survey design) and subsequently analysed for significant effects of such unintended effects. If a P/CVE project involves both male and female participants, testing for unintended consequences should include testing for significant differences in attrition by sex (see upcoming § "Test for Participant Attrition by Sex/Gender"). Naturally, one hopes for null effects along such dimensions, but—to reiterate—testing for unintended consequences is an empirical question, and one that deserves more than mere lip service or wishful thinking.

Advancing Human Rights Through M&E During Data Collection Implementation Constraints and Pragmatic Adaptations

P/CVE evaluations are frequently conducted in fragile, conflict-affected, or politically sensitive environments where access to participants may be brokered by security actors or other gatekeepers, and where evaluators and enumerators may face intimidation, surveillance, or physical risk (Bush & Duggan, 2013; Jansen et al., 2025). These conditions can compromise both data quality (e.g., nonresponse, underreporting, socially desirable responding) and research ethics (e.g., heightened risk of inadvertent harm). Accordingly, feasibility and safety constraints should be treated as design parameters rather than as implementation inconveniences, and they should be documented transparently in evaluation outputs (Bush & Duggan, 2013).

Practitioners also operate amid real tensions between donor reporting demands and rights-based safeguards, including pressures to deliver rapid indicators on short funding cycles with limited local analytic capacity (Sawadogo-Lewis et al., 2022). Likewise, small implementing organizations may have minimal staff time and expertise to support rigorous M&E, even when they value evaluation (Thurston & Potvin, 2003). In such settings, the practical aim is often not maximal measurement, but defensible measurement: collecting the least data necessary to answer core questions while minimizing risk, burden, and politicization (Bush & Duggan, 2013; Jansen et al., 2025).

Finally, evaluators should recognize that the term “human rights” can be politically loaded in some contexts, including being interpreted as external critique of state security practice. One pragmatic response is to maintain the substance of rights protections while using locally resonant language (e.g., “dignity,” “participant safety,” “fairness,” and “lawful practice”) in field-facing materials and briefings, and to ensure that “rights proofing” is not reduced to a purely technical exercise that inadvertently legitimizes coercive practice through oversecuritization (Hamilton & Lippert, 2020). Consequently, it is prudent to anticipate such constraints during an evaluation’s design phase, rather than treating them as after-the-fact caveats. Accordingly, the following pragmatic adaptations are offered as non-exhaustive starting points for balancing participant safety, data quality, and donor reporting demands:

- Conduct a brief, context-specific risk assessment that addresses confidentiality, duty-to-report boundaries, and enumerator safety, and revisit these as conditions change (Bush & Duggan, 2013; Jansen et al., 2025).
- Use sampling and recruitment procedures that reduce gatekeeper influence where possible (e.g., multi-channel recruitment, respondent-driven approaches when appropriate), and document unavoidable constraints (Bush & Duggan, 2013).
- Prefer low-intrusion, privacy-preserving data collection (e.g., short instruments, remote modes, de-identified storage), and plan in advance for secure backup and controlled access to data (Jansen et al., 2025).
- Where donor indicators conflict with participant safety or non-stigmatization, negotiate for proxy indicators or slower reporting rhythms and acknowledge the trade-offs in evaluation documentation (Sawadogo-Lewis et al., 2022).

To make the foregoing constraints and adaptations more concrete, the following two brief vignettes illustrate how an evaluation team might adjust data collection procedures under common access and security pressures while still preserving the substance of rights protections. These examples are intended as illustrations—not prescriptions—and should be adapted based on local consultation and ongoing risk monitoring. As a first example, in a setting where participation lists are routinely requested by local authorities, an implementing NGO may choose to collect anonymous, session-level feedback (no names, no phone numbers) and to store only aggregated counts by age and sex, etc., while documenting such group-level (i.e., disaggregable) data constraints in the evaluation report.

As a second example, in contexts where enumerators might be questioned and/or searched by authorities (e.g., at checkpoints), survey data should be collected via mobile devices instead of hardcopy. Barring mobile service, survey data could be aggregated (averaged over respondents) or otherwise summarized/updated on the fly (whether in hardcopy form or via an offline computing device): rendering data that are not only anonymous but unidentifiable at the individual level. If such challenges cannot be adequately surmounted, it would be methodologically and ethically preferable to suspend data collection during periods of elevated physical insecurity or unmanageable official interference.

To promote human rights, during data collection, a recommended practice is to accord meaningful roles for civil society actors to participate in those activities, including subsequent policymaking (Frank, 2017). Civil society has been called a "non-negotiable stakeholder" in P/CVE-related solutions, given that civil society has "a central role in the achievement of goals relating to human rights and the rule of law," by playing a "key role in balancing systems and politics to maintain these achievements" (Frank, 2017, p. 11). Consequently, such meaningful CSO involvement should be integrated within P/CVE responses at national, regional, and global levels, through transparent and participatory processes, and—likewise—should advance women's roles in P/CVE policy and programming processes (Frank, 2017; Ndung'u & Shadung, 2017). This is congruent with the UN's PVE Plan of Action which, in part, calls for engaging communities, empowering youth, gender equality, and empowering women (Frank, 2017, p. 36). Fortunately, there is a subfield of evaluation that can be brought to bear on assessing such participatory processes and can be integrated within a given data collection: namely coalition assessment.

Coalition Assessment

A coalition is a group of individuals who represent multiple organizations, who have agreed to work together to achieve a common goal (Smathers & Lobb, 2014). The underlying philosophy of a coalition is that members from multiple sectors or disciplines can, together, achieve better results than any one of the members could achieve alone (ibid.). As its name implies, coalition assessment is intended to measure and assess the quality of coalitions.

From an M&E perspective, coalition assessments are commonly of two types (Greenwald & Zukoski, 2018). The first type measures "relationships, climate, and expectations" (RCE): measurements that are typically based on the perceptions of coalition members regarding their satisfaction with the collaboration, the representativeness of the coalition with respect to its stakeholder ecosystem, and coalition members' confidence in the coalition's leadership, and related areas (ibid.). The second type of coalition assessment is intended to measure the "extent of collaboration" (EC): measures that are typically based on objective features of coalitions, such as the regularity of communications, and commitments of resources (ibid.).

Coalitions tend to form around a given issue, or cause (Smathers & Lobb, 2014), and the cause in the case of P/CVE programming is either P/CVE explicitly, or a P/CVE-related topic of common interest to the coalition members (e.g., enhancing public safety, safeguarding youth, etc.). In either case, coalition building can help to build support for the cause of P/CVE. In so doing, another potential advantage of coalitions is that they can help to spread the costs of working toward the common objective—costs of both time and funding—among coalition members.

Fortunately, there already exists an extensive array of available measures and other tools to facilitate coalition building and coalition assessment. Many such tools are available through the Society for Public Health Education (SOPHE, formerly presided over by Dr. Frances Butterfoss, an authority on the subjects of coalition building and coalition assessment). For further information, see the link in this endnote to investigate the offerings of that society.[i] A small sample of those freely available resources includes brief guides on the following topics:

- Coalition start-up tools
- Extant coalition planning tools
- Coalition building tools
- Coalition assessment and evaluation tools
- Coalition sustainability tools

Low-Resource Adaptations: A "Minimum Viable" M&E Package

Many locally delivered P/CVE initiatives are implemented by small organizations with limited evaluation budgets, no dedicated M&E staff, and constrained statistical analysis capacity. In such contexts, expecting "full-scale" designs can be counterproductive; what is often feasible is a staged approach that builds evaluative practice over time while maintaining basic standards of transparency, safety, and analytic defensibility (Sawadogo-Lewis et al., 2022).

A pragmatic starting point is to pre-specify a small set of core outcomes and safeguards (including sex-disaggregated reporting where relevant), and to prioritize data that are actionable for program improvement. This approach aligns with calls for more coherent

and comparable metrics in P/CVE evaluation, while acknowledging resource constraints (Charkawi et al., 2024). In short, the aim in such circumstances is not maximal measurement, but defensible measurement: a small set of core outcomes and safeguards that can realistically be implemented, repeated, and interpreted without imposing undue burden on staff or participants.

Accordingly, the following "minimum viable" M&E package is offered as a pragmatic starting point:

- Define 3–5 priority outcomes and 3–5 "do no harm" indicators that can be measured repeatedly with minimal burden (Charkawi et al., 2024).
- Use short pre-post instruments or brief structured interviews paired with routine implementation monitoring (attendance, reach, fidelity) to support plausible interpretation of change (Lu et al., 2019).
- Record and report feasibility constraints (access, gatekeeping, security incidents) as part of the evidence base rather than treating them as anecdotal context (Bush & Duggan, 2013).
- Build in a lightweight learning loop: a recurring, documented "pause and reflect" meeting with implementers and community representatives to interpret findings and revise practice accordingly (Sawadogo-Lewis et al., 2022).

Advancing Human Rights Through M&E After Data Collection Upscaling Programs

Funders have a key role to play with respect to building the P/CVE evidence base (Frank, 2017). This includes upscaling relatively small local programs to test the transferability of those programs to other contexts (Frank, 2017). Upscaling (or "Scaling up") entails efforts intended to increase the impact of programs or products—that have been successfully pilot tested—to benefit broader populations of people, or to benefit a given population in new/additional ways (Cooley & Linn, 2014). Additionally, scaling might be performed to increase the sustainability of an enterprise (e.g., by virtue of an economy of scale; World Health Organization, 2015).

Should a P/CVE Program Upscale?

From a humanitarian perspective, the answer might seem obvious; of course, we would want to serve more people. However, before forging ahead with ambitions to upscale a given P/CVE program, closely consider participant incentives, and keep in mind that such incentives might be partially (or wholly) intangible. For example, a P/CVE program might also promote human rights by helping marginalized people access jobs or education, or by ensuring their voices are heard by policymakers. Further consider whether such a program does so through social means: for example, support groups, job fairs, mentorship programs, or "town hall" style meetings. Likely, any of those events could be upscaled by migrating them online. However, intangible social benefits (e.g., sense of camaraderie, sense of accountability, etc.) might be diminished, or lost, by transferring in-person events to cyberspace: so much so, that the "online" versions of such programs might achieve their primary objectives to a lesser degree than the original small-scale versions. In short, if a P/CVE program cannot replicate its incentives (or sufficiently compensate for them) at a larger scale, then upscaling might produce diminishing returns—or worse—be counterproductive.

Scaling in P/CVE is rarely a purely technical exercise of replication. It commonly entails negotiating new partnerships, permissions, and community relationships, and it can shift how communities perceive a program (e.g., from one focused on "community support" to "security surveillance"), which can heighten the risk of stigma and suppress candid participation (Charkawi et al., 2024; Saraiva & Erfe, 2023). Consequently, scaleup decisions should be treated as hypotheses about transferability that require their own monitoring and evaluation plan, including safeguards to protect legitimacy and trust during expansion (Milat et al., 2015).

To preserve rights-based standards while scaling, evaluators can specify a small set of "non-negotiable" safeguards (e.g., voluntariness, confidentiality, non-discrimination, and meaningful gender inclusion) and track them as implementation outcomes alongside program outputs.

Planning Early to Upscale

If a given P/CVE program is not fully implemented, or if it remains empirically "unproven," why might it be advantageous to consider scalability early in a program's development? By beginning with an eye toward scalability—to have at least a basic strategy for achieving it—one might better structure a program, and measurement of its progress, to facilitate upscaling should the time come to do so (Cooley & Linn, 2014). For example, early planning to upscale might entail modifying a program's approach to reduce the unit cost of service delivery: reducing the time or materials required to serve a given number of program participants (ibid.). Also, as mentioned, any planning to upscale should consider the incentives that individuals have to participate in the program: to plan for those (or substantively similar) incentives to be available at a larger scale, in ways that do not induce disparities for any subpopulation(s) of prospective beneficiaries, or that are otherwise counterproductive (ibid.).

Scalability Assessment Tool

The topic of scaling a program is a subfield of business unto itself; so, it would be prudent to consult with specialists of business scaling prior to any bona fide attempt to upscale a P/CVE program. Nevertheless, to gain a broad idea of the evidence, processes, and other supports that are prudent to consider with respect to upscaling, consider the "Scalability Assessment Tool," available in appendix two (p. 17) of "Taking Innovations to Scale: Methods, Applications and Lessons," co-authored by one of the presidents of the Society for International Development, Larry Cooley, available via the link in this endnote (Cooley & Linn, 2014).[ii] It consists of a questionnaire that asks program managers (and perhaps other key stakeholders who are closely familiar with a given program) to rate a given program on such factors as the following.

- Is the program based on sound evidence?
- Does it address an important, persistent problem?
- Does it offer superior effectiveness, relative to competing programs?
- Is it cost-effective, relative to competing programs? (ibid.)

Though, to wit, there is no published data with respect to this tool's content validity or predictive validity, it can facilitate a project team's self-reflection: to gauge the extent to which a program is poised for scaling.

Additional Resources

The World Health Organization has developed a robust "Guide to Fostering Change to Scale Up Effective Health Services," available through the link in this endnote (2013).[iii] It was designed from a general, principles-based perspective about scaling, and—as such—can easily be read with P/CVE programs in mind. Additionally, the World Health Organization has developed an "Assessment and Planning for Scale Toolkit," available through the link in this endnote.[iv] Although it was written with the intent of scaling their "mHealth" program, again, it can be read with P/CVE programs in mind. At very least, the scaling of the mHealth program, as detailed in the aforementioned guide, can serve as a case study for how to scale a public health initiative.

Advancing the Roles of Women Through M&E

Gender mainstreaming, with respect to P/CVE initiatives, is often cited as a goal among stakeholders (White, 2020). However, P/CVE-relevant policies that affect women and girls, including policies ostensibly to benefit women and girls, often lack substantial input or leadership from the women and girls whom those policies will affect (Global Counterterrorism Forum, 2016; White, 2020). Not surprisingly, this has led to failures of meaningful inclusion of women and girls in program design and evaluation of P/CVE initiatives (White, 2020). Developing P/CVE programming through a gender-responsive approach is important, not only on moral grounds, but because research suggests that doing so will improve the effectiveness of P/CVE programs (Ndung'u & Shadung, 2017; White, 2020) and—hence—contribute to a more secure world (White, 2020).

Advancing the Roles of Women Through M&E Before Data Collection

First, gain clarity on whether the intent is to measure gender and/or sex as part of M&E data collection and subsequent analyses. What is commonly referred to as gender is simply a misnomer in place of the word sex. Sex can be recorded as male, female, or "other" (or other preferred terms); it is a categorical variable. In contrast, "gender" is a continuum of masculinity to femininity; it is a continuous/scalar variable.

This distinction between sex vs. gender is not at all vague and is well-known to virtually all sociologists and psychologists. If, for example, one is interested in accounting for different outcomes between men and women, one is interested in measuring sex, not gender. In the authors' experience, participants' sex has almost always been the variable of interest to governmental and NGO stakeholders with respect to P/CVE program participants and a given program's outcomes. This clarification is not simply semantic; whether a variable is measured categorically (e.g., female/male) vs. continuously (e.g., on a continuum of feminine to masculine) has profound implications: both for the type of statistical tests that can be performed with the data, and the sample sizes of participants required to detect statistically significant effects (see upcoming § "Statistical power").

If one wishes to measure participants' gender identity (vs. constraining responses to categorical male/female or male/female/other, etc.) one way to accomplish that simply is through an open-ended survey item. In the publication "Rethinking and updating demographic questions: Guidance to improve descriptions of evaluation samples," the following elegantly simple wording has been proposed: "How do you currently describe your gender identity? Please specify:" (Hughes et al., 2016).

Statistical Power

It has been asserted that "it should not be assumed that gendered impacts need not be evaluated" in the context of developing P/CVE programming (Ndung'u & Shadung, 2017, p. 13). Accordingly, the variables of sex or gender should always be anticipated for analysis

within a given evaluation of a P/CVE program. This translates into the need to recruit sufficient sample sizes to afford meaningful analysis of sex or gender.

Simply put, for inferential statistics to reveal significant differences (either between comparison groups and/or between time points) and to do so reliably, M&E designs need to achieve sufficient statistical power. In part, the number of subgroups (e.g., female vs. male, age groups, etc.) that will be analysed as part of a given quantitative (i.e., statistical) data analysis influences how many evaluation participants will be required to afford sufficient reliability in the findings.[3] Therefore, as alluded to earlier, the way that sex or gender will be measured as part of the evaluation of a given P/CVE program will fundamentally affect how many evaluation participants will be required for quantitative analyses to reveal statistically significant (in other words, sufficiently reliable) findings.

This is no trivial matter. In the authors' experience, reviewing myriad research and evaluation projects in the fields of P/CVE, a lack of statistical power is an all-too-common cause of M&E projects failing to demonstrate statistically significant findings. The shame of such null findings is that required minimum sample sizes can and should be estimated in advance: in a project's planning phase. By recruiting, at least, the minimum required number of research/evaluation participants, null findings—due to insufficient statistical power—are entirely preventable.

To estimate the number of participants needed for a given evaluation, specialized software is available for such calculations. The following endnote provides a link both to a widely-used, freely available statistical power calculator (G*Power), and its user's manuals.[v] However, even with the use of statistical power calculators, calculating statistical power is an arcane task; therefore, if an M&E team does not have at least one member who possesses a strong background in inferential statistics, it will be critical to consult with a statistician to estimate how many participants should be included to perform a given evaluation.

[3] Statistical power also is affected by other factors such as the size of a measured effect, the confidence level sought for an analysis, and the accuracy of the instrument used to measure a given effect.

JOURNAL FOR DERADICALIZATION

Spring 2026
No. 46
ISSN: 2363-9849

Advancing the Roles of Women Through M&E During Data Collection Facilitate Participation of Women: Minimize Barriers to Mitigate Sampling Bias

As discussed in the previous section, it is necessary to recruit sufficient numbers of participants to be able to perform meaningful data analysis. With respect to facilitating the participation of female participants, this entails ensuring that there are sufficient incentives and affordances in place during data collection.

Incentives

Participants are commonly offered financial compensation for participating in a given M&E initiative; however, as mentioned previously with respect to scalability, it is important to consider that participant incentives might be partially (or wholly) intangible. For example, participants might participate if they believe that their participation will help a cause that they support. For women, such a cause might be simply the opportunity to make women's voices heard through a given M&E initiative. Nevertheless, regardless of whether participant incentives are monetary or non-monetary, if participant incentives are insufficient, at least two problems may arise.

The first problem that might arise, as suggested above, is that insufficient participant incentives might lead to a failure to recruit sufficient sample sizes to perform meaningful data analysis. Bear in mind that women who are employed and/or serve as caregivers might be unable to set aside such responsibilities, to participate in an M&E initiative, without compensation. Therefore, sufficient incentives must be offered to allow women to justify their participation.

The aforementioned problem harkens to the second problem that might arise if participant incentives are insufficient. If little or no incentive is offered to participants, it risks biasing the evaluation sample—not only by sex or gender—but also by the sociodemographic characteristics of the women who choose to participate: an example of so-called selection bias. For example, offering little or no incentive might bias the sample toward those who have either the luxury of free time to participate (or who can afford to participate for low

compensation), or toward those who are desperate enough to participate for low compensation. This underscores a critical point with respect to offering sufficient participant incentives; participant sampling/recruitment is not only about procuring sufficient *numbers* of participants, but a sufficiently representative *cross-section* of participants. In other words, one needs to ensure—not only that a given M&E effort recruits a sufficient number of females (women and/or girls)—but that those who participate are a representatively diverse cross-section of females from a given population of interest.

Affordances

The aforementioned monetary compensation for evaluation participants is a kind of affordance: a tool that allows participants to participate who might otherwise be unable to participate. However, monetary compensation is not the only kind of affordance that might be necessary to facilitate the participation of a sufficiently large, sufficiently diverse sample of evaluation participants, including females. Other affordances include those that minimize logistical barriers to participation. For example, females (vs. males) in a given population might be disproportionately limited in their means of transportation (e.g., access, or license, to drive a car or motorbike). Consequently, to avoid unnecessarily biasing an evaluation sample with respect to socioeconomic status, and/or geography, it might be necessary to arrange free and convenient transportation for participants. Alternatively, to reduce logistical barriers, participants could be reachable by electronic means: for example, telephonic interviews, online surveys, or surveys sent via mobile phone text/SMS.

Code for the Sex (or Gender) of Respondents & Data Collectors

Another potential source of bias can be inadvertently introduced into data that are collected by in-person data collectors, by virtue of the sex/gender of those collectors. Specifically, participants (whether identifying as female or male) might respond to questions differently, depending upon the sex/gender of the data collector(s) with whom they interact (i.e., interviewer, focus group facilitator, or survey enumerator). Therefore, to discover the extent to which the sex/gender of data collectors might have influenced participants' responses, it is necessary to code for the sex/gender of participants (if, indeed, both females

and males are part of the overall evaluation sample), and for the sex/gender of the data collectors (if data collectors comprise both females and males). Subsequently, if such coding has been performed, it will be possible to analyse the extent to which participants' responses tended to vary based on the sex/gender of participants vs. those of the data collectors.

Such considerations might be especially important for the collection of personally sensitive (e.g., potentially embarrassing) information. If a given data collection will include the collection of such personally sensitive information, it might be best to match respondents with same-sex data collectors. Such matching does not eliminate the potential that respondents will (perhaps unknowingly) alter their responses based on the sex of the data collectors, but it might be the only plausible option to increase the likelihood that participants will be forthcoming and honest about sensitive subject matters.

Advancing the Roles of Women Through M&E After Data Collection Test for Theoretically Plausible Differences by Sex/Gender

Following data collection, as mentioned, among the data analytic tasks is to analyse the extent to which participants' responses differed based on their self-identified sex or gender. This should be considered standard operating procedure (and must be performed by persons well-trained in inferential statistics), but an important caveat is that such comparisons still should be justified according to theory. In other words, simply "fishing" (as it is known in the field of data analysis) for statistically significant differences between males and females, without a compelling theory to expect such differences, is tantamount to unethical statistical practice, because—as has been staunchly argued, and empirically demonstrated—it elevates the risk of false-positive findings (Christensen et al., 2021; Forstmeier et al., 2017).

Test for Gendered "Participant by Data Collector" Interactions

Another important data analytic task is to test for any statistically significant differences based on the sex/gender of participants vs. those of any data collectors who interface with participants. Generally, the hope is that such a statistical interaction will be null (statistically insignificant). Nevertheless, if such an interaction is found to be significant,

any associated findings need to be interpreted cautiously, because it means that something about the sex/gender of the data collectors inadvertently affected participants' responses. In other words, participants' average responses on a given variable of interest (or combination of such variables) do not represent their "true scores" but their scores made more or less extreme, as inadvertently influenced by the sex/gender of the data collectors. Naturally, following discovery of such unintended interactions, it behoves the evaluation team to consider why that might have occurred, and a) to adjust any future data collection procedures accordingly, to eliminate that source of bias, or b) to caveat any findings that were found to be influenced by the sex/gender of the data collectors.

Test for Participant Attrition by Sex/Gender

As mentioned, female vs. male evaluation participants might face disproportionate barriers to participation in a given M&E initiative. Furthermore, such barriers tend to be multiplied for M&E projects that collect data from participants at two or more timepoints (i.e., longitudinal, pre-post evaluation designs; West et al., 2004). Indeed, such barriers might cause participants to dropout disproportionately from an M&E data collection, correlated with their sex/gender. The best way to deal with such dropout (commonly referred to as attrition) is to prevent it, and a substantial body of knowledge, and associated techniques, have been developed to enhance participant retention (see Davis et al., 2002; see Williams, 2021).

Nevertheless, attrition can happen, but the question becomes: to what extent does it matter. Of course, if participants discontinue their participation prematurely due to a systematic reason—perhaps an uncomfortable feature of the data collection, or a factor external to the data collection (e.g., logistical barriers)—such factors bias the sample and restrict the generalizability of findings merely to the characteristics of those remaining in the study (Kazdin, 2003).

However, attrition might not be systematic; it might occur merely due to random factors that do not threaten the generalizability of findings. Therefore, attrition should be analysed at the earliest opportunity—if possible, while an evaluation is in a pilot phase—to diagnose whether the attrition represents a systemic problem that warrants correction. The only way to test for this is to analyse whether the outcomes of interest differed between those

who remained in the study vs. attriters. The procedure for statistically testing, and interpreting, whether participant attrition was problematic—systematic (vs. at random)—is detailed in a publication by Williams (2021). Nevertheless, even if it is assessed that attrition is completely at random, it is considered ethical practice to describe the sample characteristics of attriters in evaluation reports.

Conclusion

Concrete Next Steps and Research Priorities

Taken together, the foregoing recommendations imply several concrete, near-term steps that can improve the empirical defensibility and comparability of P/CVE evaluations, while advancing rights-based and gender-responsive standards. Accordingly, the following priorities are proposed for donors, implementers, or researchers alike. Their aim is to strengthen the evidence base in support of those standards:

- Develop and publish a concise, human rights and gender-responsive M&E protocol for P/CVE (including minimum ethical safeguards, suggested or exemplar indicators, and reporting templates; Charkawi et al., 2024).
- Invest in evaluator safety and "field-realistic" ethics practices, including participatory risk assessment with local teams and documentation of access constraints that might bias findings (Bush & Duggan, 2013; Jansen et al., 2025).
- Support lightweight learning systems for small implementers (e.g., periodic reflection cycles and simple dashboards) that balance accountability with feasible learning, rather than incentivizing superficial metrics (Lu et al., 2019; Thurston & Potvin, 2003; Sawadogo-Lewis et al., 2022).
- Encourage scaleup only when fidelity can be monitored and when expansion plans explicitly address perceived program legitimacy, stigma risk, and community consent (Hamilton & Lippert, 2020; Milat et al., 2015).

The present work offered techniques, and concrete recommendations, with respect to measurement and evaluation, to support the advancement of human rights and the roles of

women in P/CVE program design and evaluation. As mentioned, donors have a key role to play by investing in evidence-building P/CVE activities (Frank, 2017). In other words, donors are pivotal with respect to advancing M&E in the field of P/CVE. Given both the importance of preventing violent extremism, and donors' positions of power, donors have a responsibility—a duty—to empirically assess the effectiveness of policies and programs related to violent extremism and terrorism (ibid.). In the words of Cheryl Frank:

> This implies adopting a broad evidence-building approach including support towards the evidence-based design, implementation and evaluation of programs and policies. Donors are in a unique position to take up lessons from past initiatives into their own practices, discourage outdated and inappropriate methodologies and encourage the documentation and public dissemination of the program results. (2017, p. 12)

References

Bush, K., & Duggan, C. (2013). Evaluation in conflict zones: Methodological and ethical challenges. *Journal of Peacebuilding & Development, 8*(2), 5–25. https://doi.org/10.1080/15423166.2013.812891

Charkawi, W., Dunn, K., & Bliuc, A.-M. (2024). Evaluations of countering violent extremism programs: Linking success to content, approach, setting, and participants. *International Journal of Law, Crime and Justice, 77,* 100674. https://doi.org/10.1016/j.ijlcj.2024.100674

Christensen, J. D., Orquin, J. L., Perkovic, S., & Lagerkvist, C. J. (2021). Preregistration is important, but not enough: Many statistical analyses can inflate the risk of false-positives. *PsyArXiv*. https://doi.org/10.31234/osf.io/cj3xq

Cooley, L., & Linn, J. F. (2014). *Taking innovations to scale: Methods, applications and lessons*. https://www.msiworldwide.com/wp-content/uploads/2023/10/Taking-Innovations-to-Scale.pdf

Davis, L. L., Broome, M. E., & Cox, R. P. (2002). Maximizing retention in community-based clinical trials. *Journal of Nursing Scholarship*, *34*(1), 47–53.

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10.1111/brv.12315

Frank, C. (2017). How human rights and the rule of law can address terrorism. *Africa in the World Report*, *3*. https://community-democracies.org/app/uploads/2018/06/aitwr-3-1.pdf

Global Counterterrorism Forum. (2016). *Good practices on women and countering violent extremism*. https://www.thegctf.org/Portals/1/Documents/Framework Documents/GCTF Good Practices on Women and CVE.pdf?ver=2016-03-29-134644-853

Greenwald, H. P., & Zukoski, A. P. (2018). Assessing collaboration: Alternative measures and issues for evaluation. *American Journal of Evaluation*, *39*(3), 322–335. https://doi.org/10.1177/1098214017743813

Hamilton, C., & Lippert, R. K. (2020). Governing through human rights in counter-terrorism: Proofing, problematization and securitization. *Critical Criminology, 28*(1), 127–145. https://mural.maynoothuniversity.ie/id/eprint/16389/

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of*

Psychological Research, *21*(3), 138–151. https://cdn.ymaws.com/www.psichi.org/resource/resmgr/journal_2016/21_3Fall16JN.pdf

Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., & Tyler, T. R. (2012). Why do people comply with the law? Legitimacy and the influence of legal institutions. *The British Journal of Criminology, 52*(6), 1051–1071. https://doi.org/10.1093/bjc/azs032

Jansen, S., Niyonsenga, J., Nsabimana, E., Kagaba, M., Rutembesa, E., Slegh, H., Mihigo, B., & Mutabaruka, J. (2025). Real ethics has dirty feet – data collector perspectives on risk exposure during data collection in conflict-affected Eastern DRC. *Conflict and Health, 19*(21). https://doi.org/10.1186/s13031-025-00658-0

Kazdin, A. E. (2003). Drawing valid inferences I: Internal and external validity. In *Research design in clinical psychology*. Allyn & Bacon/Pearson Education.

Lu, S. K., Elliott, S. J., & Perlman, C. M. (2019). Perceived facilitators and barriers to evaluative thinking in a small development NGO. *Canadian Journal of Program Evaluation, 34*(1), 68–83. https://utppublishing.com/doi/pdf/10.3138/cjpe.43118

Madon, N. S., Murphy, K., & Cherney, A. (2017). Promoting community collaboration in counterterrorism: Do social identities and perceptions of legitimacy mediate reactions to procedural justice policing? *The British Journal of Criminology, 57*(5), 1144–1164. https://doi.org/10.1093/bjc/azw053

Milat, A. J., Bauman, A., & Redman, S. (2015). Narrative review of models and success factors for scaling up public health interventions. *Implementation Science, 10*(113). https://link.springer.com/content/pdf/10.1186/s13012-015-0301-6.pdf

Milat, A., Newson, R., King, L., Rissel, C., Wolfenden, L., Bauman, A., Redman, S., & Giffin, M. (2016). A guide to scaling up population health interventions. *Public Health Research & Practice, 26*(1), e2611604. https://connectsci.au/pu/article-pdf/doi/10.17061/phrp2611604/1821271/pu19056.pdf

Moayerian, N., Stephenson, M. O., Jr., & Stivachtis, Y. A. (2025). The human rights of displaced populations (pp. 15-30). In A. Otruba, M. O. Stephenson, Jr., Y. A. Stivachtis, & N. Dzotsenidze (Eds.), *Violent infrastructure: Protracted displacement and housing injustice in Tskaltubo, Georgia*. Virginia Tech Publishing. https://publishing.vt.edu/books/73/files/15eaf868-c1a0-4859-b9b1-10e446a56d7b.pdf

Ndung'u, I., & Shadung, M. (2017). Can a gendered approach improve responses to violent extremism? Institute for Security Studies. https://issafrica.s3.amazonaws.com/site/uploads/aitwr-5.pdf

Saraiva, R., & Erfe, A. (2023). Preventing violent extremism with resilience, adaptive peacebuilding, and community-embedded approaches. *Current Opinion in Environmental Sustainability, 61*, 101271. https://doi.org/10.1016/j.cosust.2023.101271

Sawadogo-Lewis, T., Bryant, R., & Roberton, T. (2022). NGO perspectives on the challenges and opportunities for real-world evaluation: A qualitative study. *Global Health Action, 15*(1). https://doi.org/10.1080/16549716.2022.2088083

Shaheed, A. (2018). Elimination of all forms of religious intolerance. *Interim report of the Special Rapporteur on freedom of religion or belief, submitted in accordance with General Assembly resolution*, *A/73/362*. https://docs.un.org/en/A/73/150

Smathers, C., & Lobb, J. (2014). Coalitions: Introduction (Community Development Fact Sheet 1 [CDFS-1], Building Coalitions Series). Ohio State University Extension. https://ohioline.osu.edu/factsheet/CDFS-1

Thurston, W. E., & Potvin, L. (2003). Evaluability assessment: A tool for incorporating evaluation in social change programmes. *Evaluation, 9*(4), 453-469.

UN Women. (2021). *Global digital consultation: Civil society voices on the gendered dimensions of violent extremism and counter-terrorism responses*. https://www.unwomen.org/en/digital-library/publications/2020/11/gendered-dimensions-of-violent-extremism-and-counterterrorism-responses

West, S. G., Biesanz, J. C., & Kwok, O. M. (2004). Within-subject and longitudinal experiments: Design and analysis issues. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The SAGE Handbook of Methods in Social Psychology* (pp. 287–312). https://doi.org/10.4135/9781412976190.n13

White, J. (2020). Gender in countering violent extremism program design, implementation and evaluation: Beyond instrumentalism. *Studies in Conflict and Terrorism*, *46*(7), 1192–1215. https://doi.org/10.1080/1057610X.2020.1818435

Williams, M. J. (2021). Attrition happens (and what to do about it). *Journal for Deradicalization*, *26*, 217–226. https://journals.sfu.ca/jd/index.php/jd/article/view/441/269

Williams, M. J., & Kleinman, S. M. (2013). A utilization-focused guide for conducting terrorism risk reduction program evaluations. *Behavioral Sciences of Terrorism and Political Aggression*, *6*(2), 102–146. https://doi.org/10.1080/19434472.2013.860183

World Health Organization. (2015). *The MAPS toolkit: mHealth assessment and planning for scale.* https://www.who.int/publications/i/item/9789241509510

About the JD Journal for Deradicalization

The JD Journal for Deradicalization is the world's only peer reviewed periodical for the theory and practice of deradicalization with a wide international audience. Named an "essential journal of our times" (Cheryl LaGuardia, Harvard University) the JD's editorial board of expert advisors includes some of the most renowned scholars in the field of deradicalization studies, such as Prof. Dr. John G. Horgan (Georgia State University); Prof. Dr. Tore Bjørgo (Norwegian Police University College); Prof. Dr. Mark Dechesne (Leiden University); Prof. Dr. Cynthia Miller-Idriss (American University Washington D.C.); Prof. Dr. Julie Chernov Hwang (Goucher College); Prof. Dr. Marco Lombardi, (Università Cattolica del Sacro Cuore Milano); Dr. Paul Jackson (University of Northampton); Professor Michael Freeden, (University of Nottingham); Professor Hamed El-Sa'id (Manchester Metropolitan University); Prof. Sadeq Rahimi (University of Saskatchewan, Harvard Medical School), Dr. Omar Ashour (University of Exeter), Prof. Neil Ferguson (Liverpool Hope University), Prof. Sarah Marsden (Lancaster University), Prof. Maura Conway (Dublin City University), Dr. Kurt Braddock (American University Washington D.C.), Dr. Michael J. Williams (The Science of P/CVE), Dr. Mary Beth Altier (New York University) and Dr. Aaron Y. Zelin (Washington Institute for Near East Policy), Prof. Dr. Adrian Cherney (University of Queensland), Dr. Wesley S. McCann (RTI International), and Dr. Daren Fisher (Hampton University).

For more information please see: www.journal-derad.com

Twitter: @JD_JournalDerad
Facebook: www.facebook.com/deradicalisation

The JD Journal for Deradicalization is a proud member of the Directory of Open Access Journals (DOAJ).

ISSN: 2363-9849

Editor in Chief: Daniel Koehler

Endnotes

[i] Society for Public Health Education, coalition building resources: https://elearn.sophe.org/coalition-building-resources

[ii] "Taking-Innovations-to-Scale" https://www.msiworldwide.com/wp-content/uploads/2023/10/Taking-Innovations-to-Scale.pdf

[iii] WHO "Guide to Fostering Change to Scale Up Effective Health Services:" https://tinyurl.com/WHO-guide-to-Upscaling

[iv] WHO "Assessment and Planning for Scale Toolkit:" https://tinyurl.com/WHO-Planning-to-Scale

[v] G*Power: https://tinyurl.com/G-Power-Calculator