The Life You Can Save

Quality of evidence form

Date: 27 June 2025

Intervention: Community-based program to reduce violence against women (SASA! Together)

1. Studies

Please list all studies considered for this appraisal. Provide their reference in APA style. Create a folder 'Sources' in the prospective RNP evaluation folder and save a copy of each study. Include a link for each study file next to the reference in the field below. Include a justification for why a subset of studies are examined if applicable.

We review three studies:

- Ullman et al. (2025), the most recent and comprehensive review on the effectiveness of community-based interventions aimed at improving attitudes toward women and girls and reducing violence against them. It has a particular focus on low- and middle-income countries (LMICs).
- Randomized Controlled Trial study on SASA! Together where CEDOVIP was an implementing partner. There are two documents on the same RCT that we review. The journal article by Abramsky et al., 2014 and the 3ie Impact Evaluation report by Watts et al., 2015 since they may provide complementary information.
- A quasi-experimental evaluation conducted in 2015–2016, assessed the impact of the Government of Uganda–Irish Aid Joint Programme to address gender-based violence (GBV) across eight districts in the Busoga region (UWONET and Ministry of Gender, Labour and Social Development, 2016).

Please note that this is not intended as a systematic review; rather, it represents a targeted scoping of the evidence to highlight the most pertinent and methodologically robust studies for our purposes.

We do not review the quality of the following two relevant studies

- Abramsky et al. (2016) because it is a secondary data analysis based on the Abramsky et al., 2014 RCT. Our quality of evidence criteria do not fit this type of study. However, from reviewing the paper we consider a robust analysis method was used and since it relies on experimental data results are likely highly reliable.

- Starmann et al. (2018) because it is a qualitative study and we prioritize assessing studies with experimental designs.

References of reviewed documentation

Ullman, C., Amin, A., Bourassa, A., Chandarana, S., Dutra, F., & Ellsberg, M. (2025). Interventions to prevent violence against women and girls globally: a global systematic review of reviews to update the RESPECT women framework. *BMJ Public Health*, *3*(1). https://bmjpublichealth.bmj.com/content/3/1/e001126

Abramsky, T., Devries, K., Kiss, L., Nakuti, J., Kyegombe, N., Starmann, E., ... & Watts, C. (2014). Findings from the SASA! Study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. *BMC Medicine*, 12, 1–17. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-014-0122-5

Watts, C., Devries, K., Kiss, L., Abramsky, T., Kyegombe, N., & Michau, L. (2015). The SASA! study: A cluster randomised trial to assess the impact of a violence and HIV prevention programme in Kampala, Uganda. *International Initiative for Impact Evaluation*, *24*, 1-31. https://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/sasa-study-cluster-randomised-trial-assess-impact

Uganda Women's Network (UWONET) & Ministry of Gender, Labour and Social Development. (2016, August 3). *Final evaluation of the GoU–Irish Aid joint programme to address gender-based violence in Busoga Region* [Final report]. UWONET. https://www.uwonet.or.ug/download/final-evaluation-of-the-gou-irish-aid-joint-programme-to-address-gender-based-violence-in-busoga-region-2016

2. Appraisal

2.1) Study design

Please provide details on the design of the studies considered in this appraisal. At the end indicate what is the highest study design level across the studies considered for this intervention.

Ullman et al. (2025) - systematic review of reviews

Abramsky et al. (2014) & Watts et al. (2015) - cluster Randomized Controlled Trial

UWONET and Ministry of Gender, Labour and Social Development (2016) - A quasi-experimental, mixed-methods evaluation with cross-sectional household surveys and qualitative methods, designed to assess community-level impacts of the SASA! intervention on gender-based violence prevention.

Appraisal for study design

Level	Description
☑ High	Evidence comes from one or more systematic reviews, meta-analyses or experimental studies such as Randomised Controlled Trials (RCTs).
☐ Medium	Evidence comes from quasi-experimental designs. These are well designed but non-randomised studies.
☐ Low	Evidence comes from non-experimental studies (e.g., observational studies).
☐ Not considered	Evidence comes from case studies, anecdotal reports, expert opinion or narrative reviews.

2.2) Risk of bias

Please provide details around potential risk of bias across the studies considered in this appraisal. If needed refer to *risk of bias checklists* prepared by the Joanna Briggs Institute[1] to aid the assessment of limitations of systematic reviews,[2] RCTs[3] and quasi-experimental[4] studies. Indicate for which study this is the case with a brief explanation as to why. At the end provide an overall appraisal for risk of bias across the studies considered.

We consider that the risk of bias is overall low. We give greater weight to the experimental study and the systematic review of reviews.

Ullman et al. (2025) - we consider the study has a low risk of bias.

1. Is the review question clearly and explicitly stated? Yes. The review questions are explicitly listed under the "Research questions" section (p. 4).
2. Were the inclusion criteria appropriate for the review question? Yes. Clear criteria are outlined, including populations, intervention types, outcome measures, and study designs (p. 4–5).
3. Was the search strategy appropriate? Yes. It was comprehensive and updated from

[1] https://jbi.global/critical-appraisal-tools
[2] Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003;3:25 doi:10.1186/1471-2288-3-25.
[3] Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, Aromataris E, Munn Z. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evidence Synthesis. 2023;21(3):494-506
[4] Barker TH, Habibi N, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias quasi-experimental studies. JBI Evid Synth. 2024;22(3):378-88.

a prior review, with search terms developed in collaboration with librarians and applied across multiple databases (p. 5–6).

4. Were the sources and resources used to search for studies adequate? Yes. Both academic and grey literature sources were searched, including regional databases and organisational websites (p. 5–6).
5. Were the criteria for appraising studies appropriate? Yes. AMSTAR-2 was used to assess systematic reviews, this is an established tool although the authors note a potential limitation was that it is not as suited to assess systematic reviews (better suited to assess individual studies) - but this study was a review of systematic reviews. (p. 6).
6. Was critical appraisal conducted by two or more reviewers independently? Yes. Reviews were coded independently and in duplicate using AMSTAR-2 (p. 6).
7. Were there methods to minimize errors in data extraction? Yes. Data were extracted by two independent reviewers using a shared form (p. 6).
8. Were the methods used to combine studies appropriate? Yes. The review synthesised findings qualitatively given the heterogeneity across studies and types of reviews (p. 6–7).
9. Was the likelihood of publication bias assessed? No direct assessment of publication bias was reported, though the inclusion of grey literature and expert solicitation partially mitigates this issue (p. 5–6).
10. Were recommendations for policy and/or practice supported by the reported data? Yes. Recommendations reflect patterns found in the reviews and are tied to the strength and consistency of evidence (p. 10–11).
11. Were the specific directives for new research appropriate? Yes. The review identifies key evidence gaps and provides clear directions for future research (p. 11).

Abramsky et al. (2014) & Watts et al. (2015) (page numbers based on Ambrasky)

Bias related to selection and allocation

1. Was true randomization used for assignment of participants to treatment groups? Yes. The study employed cluster randomisation with random selection of communities within matched pairs. Names of communities were drawn blindly from a bag (p. 4).
2. Was allocation to treatment groups concealed? Yes. Randomisation was conducted by the research team with allocation masked until selection (p. 4).
3. Were treatment groups similar at the baseline? Yes. Tables 2 and 3 demonstrate high comparability between intervention and control communities in demographics and key variables at baseline (p. 10–11).

Bias related to administration of intervention/exposure

4. Were participants blind to treatment assignment? No. Participants were not blinded; the intervention was public and community-based, making blinding unfeasible (p. 5).
5. Were those delivering the treatment blind to treatment assignment? Not explicitly stated but unlikely. Intervention delivery staff and community activists were aware of treatment allocation (p. 5).
6. Were treatment groups treated identically other than the intervention of interest? Yes. Both groups received standard services, but intervention sites received additional SASA! community mobilisation. Police and healthcare provider engagement occurred in both, but only intervention sites received intensive community-level programming (p. 4).

Bias related to assessment, detection and measurement of the outcome

7. Were outcome assessors blind to treatment assignment? At baseline, yes. At follow-up, no—it was not possible to blind follow-up interviewers (p. 5).
8. Were outcomes measured in the same way for treatment groups? Yes. The same standardised questionnaires were used in both groups (p. 5–6).
9. Were outcomes measured in a reliable way? Yes. Interviewers were thoroughly trained (3+ weeks), and tools were adapted from validated sources (e.g. WHO surveys) to ensure reliability (p. 5–6).

Bias related to participant retention

10. Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed? Yes. Follow-up was cross-sectional and re-sampled rather than longitudinal, so attrition is not relevant. Response rates were high in both groups (97–99%) (p. 8).

Statistical Conclusion Validity

11. Were participants analysed in the groups to which they were randomized? Yes. An intention-to-treat (ITT) analysis was conducted (p. 7–8).
12. Was appropriate statistical analysis used? Yes. A two-stage adjusted cluster-level analysis, weighted and accounting for clustering, was used, with sensitivity analysis via unpaired t-tests (p. 7–8).
13. Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial? Yes. A cluster-randomised trial was used appropriately for a community-level intervention, with proper statistical handling (p. 7–8).

UWONET and Ministry of Gender, Labour and Social Development (2016)

Bias related to temporal precedence

1. Is it clear in the study what is the "cause" and what is the "effect" (i.e. there is no confusion about which variable comes first)? Yes, the study clearly identifies the intervention (GBV programme activities such as community dialogues, training, and sensitisation) as the cause and the intended outcomes (e.g., changes in attitudes, increased reporting, reduced incidence of GBV) as the effects. The evaluation design is based on assessing these effects following the implementation of the intervention (p. 5–6, 19).

Bias related to selection and allocation

2. Was there a control group? Yes, there was a comparison between intervention and non-intervention districts (e.g., Kamuli vs. Buyende and Bugiri), which functioned as a de facto control group to assess changes in awareness, service usage, and attitudes (p. 18, 26, 31).

Bias related to confounding factors

3. Were participants included in any comparisons similar? Yes, the evaluation acknowledges similarities in the context of the intervention and comparison districts. However, it also notes contextual challenges and population mobility that may have influenced results. Nonetheless, participant profiles and baseline conditions were similar enough to allow comparison (p. 7, 18).

Bias related to administration of intervention/exposure

4. Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest? Generally yes. The main difference between groups was the exposure to the GBV programme. Outside of that, no other significant differences in services or support structures were identified between the intervention and non-intervention sites. However, there was overlap where national policies or unrelated NGO programmes also influenced communities (p. 18, 31).

Bias related to assessment, detection and measurement of the outcome

5. Were there multiple measurements of the outcome, both pre and post the intervention/exposure? No. The evaluation appears to rely on a single round of post-implementation data collection with retrospective recall. There is no baseline quantitative measurement presented in the evaluation (p. 18–19).
6. Were the outcomes of participants included in any comparisons measured in the same way? Yes. Data collection tools (e.g., household survey, qualitative interviews)

were administered consistently across intervention and non-intervention districts. (p. 18)

7. Were outcomes measured in a reliable way? Partially. While consistent tools were used, the evaluation notes reliance on recall and self-reporting, which may introduce bias. There is no mention of validation of responses or external triangulation beyond qualitative accounts (p. 18–19, 26).

Bias related to participant retention

8. Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed? Not applicable. This was not a longitudinal design with participant follow-up. It used cross-sectional data collection from different individuals at the end of the programme, so issues of attrition are not relevant (p. 18–19).

Statistical Conclusion Validity

9. Was appropriate statistical analysis used? Mostly yes. The evaluation presents descriptive statistics and some comparative analysis between districts. However, the analysis remains mostly descriptive and does not use inferential statistical methods such as confidence intervals or tests of significance. (p. 18, 24–26)

Appraisal for risk of bias

Level	Description
☑ Low	The *majority* of RCTs, systematic review(s) or meta-analysis considered show a *low* risk of bias.
☐ Medium	Across *some* of the considered studies the risk of bias may be *moderate*, but not high.
☐ High	A *high* risk of bias is identified across studies.

2.3) Imprecision

Please provide an assessment on the potential of imprecision in the reviewed studies. Imprecision refers to the uncertainty in how close the estimated effect size (e.g., the impact of an intervention) is to the true effect. It is influenced by the sample size, which directly impacts the statistical power of a study. In your answer specify for which studies and to what extent imprecision is of concern. At the end provide an overall appraisal for imprecision across the studies considered.

Moderate concern for imprecision. Abramsky et al. (2014) reported meaningful effects using appropriate methods, but some key outcomes—particularly on IPV—had wide confidence intervals due to a small number of clusters. Ullman et al. (2025) highlighted that many reviews, especially in LMICs, included studies with limited power or only intermediate outcomes. The UWONET and Ministry of Gender (2016) evaluation relied on small samples and lacked baseline data, reducing certainty in the observed changes.

Ullman et al. (2025)
Imprecision is a concern, particularly in LMICs and for newer or less-studied intervention types. While some interventions (e.g. empowerment + microfinance, school-based curricula in LMICs) had more precise findings, many others did not (p. 9–11). Many reviews included evaluations with small sample sizes or lacked statistical power to detect meaningful differences, especially in LMICs. The review notes that many interventions reported only intermediate outcomes (like norm change), not direct measures of violence, which increases uncertainty (p. 7–9).

Abramsky et al. (2014) & Watts et al. (2015)
While effect sizes suggest meaningful impacts, the precision of some estimates is limited. Imprecision is of moderate concern, particularly for IPV-related outcomes. The study authors acknowledge limited statistical power due to modest number of clusters and potential inter-cluster variation, especially for IPV outcomes (p. 13–14). Examples of concern:

- Past year physical IPV: a large effect size (adjusted RR 0.48) but wide CI (0.16 to 1.39) due to high inter-cluster variation (p. 12).
- Community response to IPV: very wide CI (2.11, 95% CI 0.52 to 8.59) due to small sample sizes among women who experienced IPV (p. 12).
- Sexual IPV: adjusted RR 0.76, CI (0.33 to 1.72), indicating imprecision (p. 12).

UWONET and Ministry of Gender, Labour and Social Development (2016)
Imprecision is a moderate concern. The evaluation draws on a small sample (e.g., household surveys in only three districts) and relies on self-reported recall-based data without baseline comparators. This limits the ability to quantify effect sizes or assess confidence in reported changes. For example, changes in attitudes or behaviour are reported but not quantified with margins of error. The absence of baseline and small sample sizes mean the study lacks statistical power to detect small-to-moderate changes with confidence. Overall, imprecision limits how definitively one can interpret the impact. (p. 18–19, 24)

Appraisal for imprecision

Level	Description

☐ Low	The studies have *high* statistical power to detect significant differences. Confidence intervals are sufficiently narrow meaning that statistical imprecision has negligible impacts on decision making. This typically means that the average study has large sample sizes.
☑ Medium	The studies only have *moderate* statistical power to detect significant differences. Confidence intervals may introduce some uncertainty into decision making. Some of the considered studies have small sample sizes.
☐ High	The studies have *low* statistical power to detect significant differences with confidence intervals that introduce uncertainty into decision making. This typically means that the studies have small sample sizes.

2.4) Inconsistency

Please provide an assessment on the potential of inconsistency among the reviewed studies. Inconsistency refers to variation in the results across studies evaluating the same intervention. At then end provide an overall appraisal for inconsistency across the studies considered.

Medium concern for inconsistency. Abramsky et al. (2014) showed consistent, directionally aligned results across outcomes. However, Ullman et al. (2025) reported variation in effects across intervention types and contexts—particularly for bystander, perpetrator, and economic empowerment interventions. The UWONET and Ministry of Gender (2016) evaluation found mixed outcomes across districts, with uneven shifts in attitudes and service awareness, suggesting contextual differences influenced results.

Ullman et al. (2025)
Inconsistency is evident in several domains, often due to contextual variation, study design differences, and varying outcome measures. However, for some interventions (like school-based curricula in LMICs), the results are more consistent (p. 7–10).

- Perpetrator interventions, where findings were mixed even within similar settings (p. 8).
- Bystander interventions and infrastructure improvements, which were primarily tested in HICs and showed variable results (p. 9).
- Empowerment and microfinance-based interventions, where impacts varied by population and context (p. 7–8).

Abramsky et al. (2014) & Watts et al. (2015)
There is no concern regarding inconsistency. Results were directionally consistent across outcomes and aligned with the intervention theory of change (p. 13–14). No contradictory effects were observed.

UWONET and Ministry of Gender, Labour and Social Development (2016)
There is some inconsistency across findings. While positive impacts are reported (e.g., increased awareness and service uptake), challenges persist such as continued underreporting, harmful norms, and lack of knowledge about some services. Some outcomes show positive change in one district and not another. For instance, awareness of referral pathways varied across districts. The evaluation also notes mixed effects on men's attitudes and lingering acceptance of GBV in some areas. Overall, inconsistency is a moderate concern but does not invalidate the observed positive trends (p. 24–28, 30–32).

Appraisal for inconsistency

Level	Description
☐ Low	The estimated effects of the intervention are broadly consistent across the studies (e.g., RCTs). There may be some minor variation.
☑ Medium	The estimated effects are broadly consistent across the studies, although there may be some moderate variation.
☐ High	There is inconsistency in the estimated effects across studies.

2.5) Indirectness

Please provide an assessment on the potential of indirectness among the reviewed studies. Indirectness refers to situations where the evidence does not directly apply to the intervention being evaluated or the specific context of interest. At the end provide an overall appraisal for indirectness across the studies considered.

There is low directness since there is robust evidence from LMICs as evidenced by the systematic review of reviews by Ullman et al. (2025), and direct evidence from two studies where the program was implemented by CEDOVIP in Uganda.

Ullman et al. (2025) - Indirectness is low. The reviewed evidence aligns well with the intervention focus (community and school-based programmes to shift norms and reduce VAWG) and includes relevant geographic focus areas such as India (p. 2, 9–10).

- Most of the evidence is directly relevant to the intervention of interest: community-based programmes to change attitudes and reduce violence.
- The review includes studies from India, where Breakthrough operates, and covers relevant interventions like group-based workshops, school-based programmes, and community mobilisation (p. 2–4, 7–10).
- Interventions in schools (e.g. life skills, dating violence prevention) are explicitly included and assessed, especially under the child and adolescent abuse prevention strategy (p. 9–10).

The cluster RCT (Abramsky et al., 2014; Watts et al., 2015) and quasi-experimental evaluation (UWONET and Ministry of Gender, Labour and Social Development, 2016) have low indirectness since these directly evaluated the SASA!Together program and CEDOVIP was an implementation partner. Implementation was in Uganda where CEDOVIP operates.

Appraisal for indirectness

Level	Description
☑ Low	The RCTs or quasi-experimental studies study the intervention directly as it is implemented by the prospective RNP or in highly relevant contexts.
☐ Medium	The studies are only moderately relevant to the context in which the intervention is implemented.
☐ High	The studies have very low relevance to the context in which the intervention is implemented.

3. Summary

Criteria	Appraisal		
Study design	☑ High	☐ Medium	☐ Low
Risk of bias	☐ High	☐ Medium	☑ Low
Imprecision	☐ High	☑ Medium	☐ Low
Inconsistency	☐ High	☑ Medium	☐ Low
Indirectness	☐ High	☐ Medium	☑ Low

4. Quality of evidence final score

Please award the prospective RNP a final *quality of evidence score*:

Level	Description

☑ High	There is a wide range of well conducted experimental and quasi-experimental evidence showing the effectiveness of the intervention (i.e., the studies show a low risk of bias and the estimates are precisely estimated). This includes evidence in the geography where the organization operates. There is at least one study assessing the impact of the intervention as implemented by the RNP (direct evidence). There is broad consistency on the effect of the intervention across considered studies.
☐ Medium	There is a medium range of experimental, quasi-experimental and non-experimental evidence showing the effectiveness of the intervention. Some studies have a moderate risk of bias, but none show a high risk. Some studies show moderate imprecision. There is primarily indirect evidence for the intervention. There is moderate variation of the results across the studies.
☐ Low	There is a low range of studies showing the effectiveness of the intervention. There is a high risk of bias and imprecision across some studies. There is only indirect evidence on the effectiveness of the intervention (none coming from studies where intervention is implemented by prospective RNP). There is high variation of the results across studies.

Please provide any comments to back up the decision if considered necessary. Only one quality score should be awarded to the RNP even if it implements more than one intervention. Hence, to inform this final score, first award an intervention specific score. Take the average across the interventions' final scores.