RCTS FOR PEACEBUILDING Lessons from Nigeria's Middle Belt

MARCH 2019

Background

Many peacebuilding interventions are rooted in the theory that if we bring conflicting groups together to work toward mutual goals, then trust between the two groups will improve, and violence will decrease.¹ Mercy Corps implemented a study that examines the effectiveness of such an intervention in Nigeria's Middle Belt, where resource-based conflict between farmers and pastoralists has led to thousands of deaths and costs the Nigerian economy nearly \$14 billion a year.²

The research focused on the first two years of Mercy Corps' USAID-funded Engaging Communities for Peace in Nigeria (ECPN), which aimed to prevent violence between farmer and pastoralist communities in Benue and Nasarawa states and was implemented in partnership with Pastoral Resolve (PARE). The program was designed to 1) strengthen the capacity of farmer and pastoralist leaders to resolve disputes in an inclusive, sustainable manner; 2) build trust among farmers and pastoralists by creating opportunities for contact and collaborations to achieve common goals; and 3) foster engagement among farmerpastoralist communities, local authorities and neighboring communities to prevent conflict. The project used a randomized controlled trial (RCT) to examine how the overall intervention affects communities. We triangulated these data by looking at effects of the program on individual participants who participated fully in the program compared to both individuals in treatment sites who did not participate in the committees and individuals in control communities.

WHAT IS AN RCT?

The use of randomized controlled trials (RCTs), often referred to as the "gold standard" of evidence, comes from the medical research field. In RCTs, researchers randomly assign some people or groups to receive the 'treatment' (in development programming, the intervention) and others to receive no treatment, while keeping other variables constant. RCTs are designed to minimize bias in research and measure the effectiveness of the treatment independently from other factors. By comparison, a quasiexperimental evaluation uses a comparison group that is not randomly selected but matched in a different way, and a traditional performance evaluation includes no comparison group and relies on pre- and post-program data.

² https://www.crisisgroup.org/africa/west-africa/nigeria/262-stopping-nigerias-spiralling-farmer-herder-violence. https://www.mercycorps.org/research-resources/economic-costs-conflict-nigeria.





¹ This is otherwise known as the Contact Hypothesis, which states that intergroup attitudes will improve and prejudice will decrease if people get to know one another under defined conditions.

Why conduct an RCT of a peacebuilding program?

While RCTs are proliferating across non- and post-conflict environments, they are rare in active conflicts, and even rarer for peacebuilding programs.³ A number of challenges stemming from insecurity make RCTs in the midst of conflict difficult: 1) safety of staff; 2) consistency of access to research populations; 3) a fluid environment, affecting implementation; and 4) high costs. In addition to ethical questions (discussed below), valid concerns have led to a resistance among many peacebuilders to quantitatively assess the results of their programs. Among these concerns are fears that more simplified measures will not capture the complexity of the peacebuilding intervention and that attribution will be difficult because peace is affected by many elements outside of the control of the program. However, peacebuilders need rigorous evidence from interventions in active conflicts in order to know what is most effective so we can replicate and adapt it, as well as justify continued investments in peacebuilding.

Specifically, benefits of conducting this RCT included:

- **Creating a counterfactua**l: Conflicts are characterized by changing dynamics: conflict spreads to other areas, it escalates and de-escalates, the political climate changes, and new actors become involved. All of these complexities can make it difficult to know if an intervention or the environment led to a change. Toward the end of the ECPN program, for example, an anti-grazing law was passed by one of the state governments, leading to an escalation of tensions and violence in the area. For some of the program's outcomes, such as attitudes and contact, our treatment communities stayed the same or improved only slightly. However, we found that on some outcomes, the control communities actually did worse over time. Without the control, we would have thought the program had little effect, but with it, we see the program bolstered communities in the face of negative environmental changes.
- Examining the widely applied theory that collaboration across conflicting groups will improve peace outcomes: Contact theory (Allport, 1954) is a central theory underpinning many peacebuilding programs, where bringing people in contact across lines of division will reduce prejudice. This theory forms the basis of what is the People-to-People approach, a USAID

THEORY OF CHANGE TESTED:

If farmers and pastoralists collaborate on quick impact projects that demonstrate the benefits of cooperation and economic projects that address underlying tensions, then there will be an increase in positive interactions, thereby building trust and economic interdependence, while reducing violence.

congressionally-mandated fund, which provides \$25M a year for these types of programs. However, little empirical evidence exists to know how valid contact theory (Paluck, Green and Green, 2018), or other leading peacebuilding theories, are when applied outside of the laboratory.⁴ Consequently, there is a significant opportunity to advance the peacebuilding field by testing these central theories.

³ 3IE (2015). Peacebuilding evidence gap map. Retrieved from: http://gapmaps.3ieimpact.org/evidence-maps/evidence-peacebuildingevidence-gap-map.

⁴ Sherif, M. (1958). Superordinate goals in the reduction of intergroup conflict. American Journal of Sociology, 63, 349-356. Tajfel, H., & Turner, J. C. (1979). "An integrative theory of intergroup conflict". In W. G. Austin & S. Worchel. The social psychology of intergroup relations. Monterey, CA: Brooks/Cole. pp. 33–47.

• Overcoming self-selection challenges in peacebuilding programs: As the aim of most peacebuilding programs is to bring people together across lines of division, one criticism leveled at these programs is that they often work with those who are already amenable to crossing divides and easy to reach. Consequently, it is difficult to know if these programs are effective for those who need the program the most, i.e., those who are most resistant to peace. Comparisons between those who participated and those who did not are not particularly informative, because those who decided to participate are likely different than those who did not, particularly on a central outcome peacebuilding programs care about: openness to the other group. Randomization can help address these shortcomings by selecting not only people who are most amenable to the program, but also those who are resistant, or in between. To address this limitation, in this study, we tried to randomly select people from target communities to be on the project committees and therefore have frequent contact with members of the other group (farmers or pastoralists), in order to increase the generalizability of our results. A downside of randomizing in this way is attrition; as people were not signing up for the program, it may be difficult to have them committed to the full length of the program (see below).

Overview of the Research

The research team conducted the baseline survey between September and December 2015 and the endline between January and April 2018. For the community-level RCT, we first established a list of sites eligible for the ECPN intervention, where each site contained one farmer and one pastoralist community. Eligible sites had a "demonstrated need," defined as the communities engaging in violent clashes in the previous year. The scoping exercise initially led to the identification of 30 eligible sites and through further visits and interviews, we narrowed the list to 15 sites (30 communities). We then randomly selected ten of these 15 sites to receive the program and monitored five of the sites as a control group. To both measure spillover and to increase power due to the limited number of communities, we also planned to randomize individual participation or non-participation in specific activities, but had to abandon the randomization effort due to various challenges (see below). Instead, for the individual-level analysis, we allowed people to self-select into participation, and collected both pre- and post data from them. Our final sample for the individual-level analysis included 74 individuals from treatment communities who participated in ECPN project committees, 121 individuals from treatment communities who garticipate in any ECPN activities, and 92 individuals from control communities.

This research measures peacebuilding outcomes in multiple ways. First, we used a survey, drawing equal numbers of respondents from farmer and pastoralist communities in each site, to measures several outcomes, including trust, social contact, attitudes toward violence, and outgroup attitudes. Second, we trained a team to observe certain behaviors of the participants, including interactions in markets, attendance at weddings, and behaviors at project meetings. Third, we engaged communities in a public goods game (PGG), a behavioral "game" used in economics and psychological research. In the PGG, participants were given a cash gift and asked to contribute part of that gift to a community fund that would finance a development project that benefits their community and the other community. Participants were told that any contribution they make to the community fund would be tripled, so that their giving 100 Naira to the community fund becomes 300 Naira for the community fund. In essence, the exercise asked participants to make a difficult trade-off between their own interests and the interests of the broader community.

Key findings from the study include the following:⁵

- Inter-group contact and trust between farmer and pastoralist communities changed more positively
 or deteriorated significantly less in ECPN sites relative to control, even as regional tensions
 increased. Contact between farmers and pastoralists in the control sites decreased by approximately
 15%, while contact in ECPN sites stayed the same. Further, though trust between farmers and
 pastoralists in control sites decreased, trust within ECPN sites increased, leading to a 13%
 difference in the overall level of trust between intervention and control sites.
- Perceptions of security increased significantly more in ECPN communities than in control communities. By the end of the first phase of the program, perceptions of security in ECPN communities improved 15% more than control sites.
- Among individuals, as a result of the program, direct participants' attitudes and behaviors improved more than indirect participants in ECPN communities, who improved more than individuals in control groups. We found that the benefits to those who participated did spread to the wider community. . These trends were strongest for intergroup trust and perceived security.
- ECPN communities did not improve their perceptions of the effectiveness of dispute resolution mechanisms, relative to control communities.

More detail, including methods, interpretation of findings, and a discussion are included in the full evaluation report.

Overcoming Obstacles to Conducting an RCT for Peacebuilding

The ECPN research team had to confront difficult questions and challenges, some of which are typical across RCTs in any sector, some of which are common for research in conflict zones, and others of which are unique to the intersection of RCTs on peacebuilding programs. What follow are some of the bigger questions and how the joint program and research team responded.:

Is it ethical to do an RCT in a conflict environment?

As a development or peacebuilding program in a conflict-affected area necessarily involves people who are suffering in some way, whether from poverty, food insecurity, trauma, or violence, to many it seems unethical to deny some people the program's intended benefits while asking for their time to participate in research. The team carefully weighed this concern against potential benefits and concluded that because there were not enough resources to intervene in every community that needed assistance in any case--which is true of almost all development interventions--we wanted to implement the program in a way where we could learn the most.

A concern that often emerges is whether the people who need the program the most will receive it when using randomization. As noted above, more communities were experiencing violence than the program could reach. We created a list of fifteen possible sites based on recent experience of violence, as well as demographic variables, such as levels of poverty and ethnic composition. Among those that looked similar and met this threshold of violence--meaning they were considered to be actively in conflict--we randomly chose ten of those to be in the program, which was the most our resources could stretch. A potential method

⁵ For more detail and an interpretation of findings, please see the full impact evaluation report of the Engaging Communities for Peace in Nigeria program, forthcoming.

for managing the tension that some will not receive the program is using a waitlist model, which means that all communities or individuals eventually become fully engaged.⁶

A final ethical concern relates to the fairness of randomization, specifically why a lottery rather than criteria, usually related to vulnerability and need, should be used. However, when randomization is done through some type of lottery, it is often more transparent than how people are selected for many programs, as the potential participants rarely learn why they were or were not selected.⁷ The ECPN program team found that explaining the system of randomization to communities made sense on an immediate level, and that local leaders were willing and able to communicate the process easily to their communities.

Conflict environments are always changing. How can you balance adaptive management with the need to keep the conditions in place for an RCT?

RCTs in any intervention can make aspects of adaptive management difficult or impossible because they require holding the tested intervention at least somewhat constant, most notably who participates in which activities. And because peacebuilding programs work in contexts where the key conflict dynamics change, adaptive management is arguably particularly important for peacebuilding. For this reason, not all contexts--particularly when the conflict is in the early stage and major dynamics are changing on a daily or weekly basis--are ripe for an RCT. But in many conflicts, the actors, underlying causes, and general locations stay at least somewhat constant for a few years, which is enough time to implement and evaluate a peacebuilding intervention. In these places, such as Nigeria's Middle Belt, we found that the conflict had reached enough of a stasis that it was possible to conduct an RCT, though we still had to make compromises in the RCT design because of context and programmatic changes.

Changing dynamics, such as policy changes like the anti-grazing law implementation noted above or conflict spillover, affected the ECPN program, and by extension the RCT. In one example, changing rain conditions along with insecurity in a neighboring area delayed the return of a semi-nomadic pastoralist group to its dry-season community, a treatment location where this group had been in conflict with farmers. While some pastoralists from this group--who had been part of the baseline survey--eventually returned, the program team decided to move forward on some activities with the pastoralists who had remained in the community, or new pastoralists who had arrived. This adaptation was the best decision for the program's success, given the uncertainty around when the original pastoralist group would return, but it meant that some of the originally surveyed pastoralists were not available for all of the program activities or the final survey. This compromise kept the integrity of the community-level randomization, but risked the individual-level randomization. For this reason and others mentioned below, the research team learned that surveying even more individuals at baseline would have provided space in the research for more individuals to drop out and may have enabled the design to maintain the individual-level RCT. Alternatively, if we had been able to include more communities at the outset, we could have dropped this community from the research with minimal impact to our ability to detect an effect of the program (ie., statistical power).

⁶ While ECPN did not set out to use a waitlist approach, after a year of implementation, the donor agreed to extend the program, and therefore, in essence, we were able to use a waitlist approach. All control communities were included in ECPN program activities in the second phase of the program.

⁷ An open lottery is how Blattman, Jamison and Sheridan (2017) conducted the randomization in their study on the effects of cash and cognitive behavioral therapy on crime in Monrovia, Liberia. All the potential participants were present when the lottery took place, and the potential participants literally saw numbers being pulled out of a bowl. As a result, people saw that there was no nepotism or corruption in how people were selected, and they knew that they were or were not selected due to chance, not something about who they are or what they did.

Conflict programs' success often ride on their ability to identify who the key players in the conflict are, and ensure those key players participate in program activities. However, in almost all cases, the program team's analysis and understanding of key stakeholder evolves throughout the program, as they get to know communities and conflict dynamics better, and as events change. This steady evolution underscores the need for adaptive management, which can create barriers for RCTs. For example, the RCT first set out to measure the impacts of the program on 'partial participants' - those community members who participated in one inter-community dialogue or other event. However, program team members found it impossible to keep some community members out of these events, especially as their understanding of key actors in the conflicts increased, and feared that doing so would threaten the success of those activities. So in the end, we dropped the tier of partial participants and ended up comparing only the full participants--those who participated in regular joint planning meetings--to everyone else in the community, as well as those in the control communities. In a different example, the team conducted an inter-community peace forum in one treatment site, but community leaders requested that leaders from a neighboring site--which happened to be a control site--attend the forum because of a recent conflict event that spread across the area. The program team decided to include the control site in that one forum and risk contamination of the research for the sake of the program's success.

Conflict dynamics are often very localized, even within a state or province. How can an RCT manage those unique community-based dynamics?

For field-based RCTs on peacebuilding programs, there is a tension between being localized and contextually relevant versus creating generalized knowledge. For the ECPN study, we agreed that certain process elements would be standardized (i.e., the how), as that was what the RCT was testing (e.g., how contact improves relationships), but community members were allowed flexibility to address conflict the way they saw as most relevant (i.e., the what). Communities led the joint initiatives in each program site following a similar process, even though the projects were different. The process for selecting projects included a series of community meetings—starting with separate farmer or pastoralist meetings and building up to joint decision-making meetings—as well as participatory needs assessments designed to gather the opinions of various demographic groups. Joint project committees included some people who had been randomly assigned to the committees as well as volunteers, drawing from an even number of farmers and pastoralists, as well as women and youth. Each committee conducted resource mapping exercises to determine needs and then managed the implementation of two grants: one for a quick impact project to build trust early in the project, the other for a joint development project.

This project selection and implementation process was all the same for each of the communities. However, what was contextually specific were the types of projects committees picked. Examples of quick impact projects included construction or rehabilitation of market stalls and schools, rehabilitation of health centers and construction of fences along grazing routes to protect both cattle and farmlands. The joint economic development projects were to address the underlying issues related to the conflict: sharing of resources that impacted their livelihoods. Pollution of water that affected both farming and livestock was the primary issue people raised, and so each site received a new borehole, with farmer and pastoralist youth helping to construct them. What was important was that each community was free to identify and implement projects that addressed conflict drivers specific to their location. But a common set of parameters around the process helped create a consistent definition of this activity that enabled researchers to measure social contact across sites.

Once randomization is complete, how do you maintain your sample size? Isn't attrition high in conflict settings?

Attrition of participants, a challenge in all impact evaluations, is an even greater risk in conflict-related programs, given the potential for displacement, higher levels of mistrust of interventions because of the conflict, or fear that prevents people from attending events. All of these challenges presented themselves for the ECPN RCT.

At the end of 2017, and immediately before the planned final evaluation, a new anti-grazing policy in Benue state, one of the two treatment states, upended many communities and increased tensions. Not only did this policy change increase mistrust and threaten program outcomes, it also made tracking down specific communities and individuals extremely difficult because several entire groups had been displaced. While this would be a challenge for any final evaluation, the amount of time investment that had gone into the RCT design made the stakes even higher for finding specific participants and respondents within a narrow timeframe. Mercy Corps extended the time of the baseline by several months in order to locate and survey as many people as possible who had also participated in the baseline. Thanks to the strength of the relationships between Mercy Corps and partners and the communities, the research team was able to locate and survey some of the targeted respondents. Even for those respondents who had been displaced because of the broader insecurity or, for pastoralists, because they were no longer allowed to graze in the state, the majority of respondents in treatment groups stated that their relationships with and trust in the other specific conflict group stayed the same or increased. However, one limitation of the data is that the community members whom we were able to locate may have been different from others, meaning that the final sample in those displaced communities may not have been fully representative of the community.

A related challenge was that in the baseline survey, because of heightened tensions and deep mistrust of outsiders given the conflict history, some entire groups of participants provided false names to the researchers. While the program team was able to gain their trust through implementation and eventually learn community members' real names, they had the added burden of having to piece together which individuals had participated in the baseline survey. This required extreme diligence on the part of partner staff to identify some of these baseline respondents, and also meant that some baseline respondents were never re-identified and so had to be replaced by others (non-randomly) for the activities and then in the final survey.

Another source of attrition was the time required to participate in treatment groups and the self-selection challenge mentioned above. Individuals who were randomized into committees needed to commit to two years' worth of periodic volunteer work. As discussed earlier, we set up three committees in treatment sites: (1) joint natural resource, (2) early warning system, and (3) peace. The research team initially randomly assigned baseline survey respondents to form part of these committees, with the other part reserved for community leaders, but random assignment proved difficult. Many people who were not selected wanted to be on the committees and some people who were selected were not able to participate or were not locatable when the committees launched. As a result, in addition to some randomized committee members, others self-selected into these committees. Therefore, during the program, the team had to drop the individual randomization and rely instead on a less rigorous comparative analysis across individuals combined with the community-level RCT.⁸

⁸ Despite relying on this self-selection for the individual analysis, however, committee members and non-committee members who were resurveyed were not statistically different on relevant attitudes and perceptions various attributes at baseline, before ECPN began. Their

Overall, the program team was able to confront some of these challenges and persevere in maintaining contact with enough participants, and the study design was able to accommodate this high attrition by dropping certain parts of the research. One lesson Mercy Corps learned is that it would have been easier to hire dedicated staff for tracking participants and maintaining communication with them so as to reduce attrition.

Program staff are already overburdened with implementation and program M&E. Is an RCT too much to ask?

Many worry that RCTs can overburden program and M&E staff, diverting precious attention from ongoing conflict analysis and program implementation. However, if the RCT is well-managed, it should not add an unmanageable burden on staff. RCTs do require staff who are committed, knowledgeable, and passionate about evaluation, and creative in solving problems. We found that program staff were excited about what they could contribute to the broader peacebuilding field, and that they were open to the possibility of failure. These attributes were essential to successfully completing the RCT and ensured that the team was motivated to persist in the face of major challenges. Additionally, team members had to be comfortable with surveying and observing people in communities where the program was not intervening, and with no promise of a future intervention. The program team and partners, as the face of the program, had the emotionally taxing work of researching sites that would not benefit from the program. In this context, the staff were able to agree to this difficult work because of their belief in what the research could tell us about the intervention.

Structural components can also reduce the time and energy required of staff: the RCT should be fully integrated and embedded within the M&E system. In our case, the RCT baseline simply replaced the typical baseline and included all of the outcomes of interest to the program team--not just those of interest to the research team. All M&E staff, not just the researchers, became experts in the RCT methodology and therefore were an essential part of its success. One concern shared by the program and research team was that after investing so much time, energy, and to some extent, funds (see below) into the RCT, if the results were inconclusive, the efforts would have felt 'wasted.' One way the research design addressed this was by including multiple ways of measuring outcomes, to ensure that we felt confident in our results. This also would help us deflect criticisms that our results--particularly self-reported attitudes--were due to a presentation bias. In addition to the survey, we added observations and the public goods game, as mentioned. This meant that even if one tool or part of a tool failed to provide learning, other tools taking advantage of the treatment and control groups might be able to contribute to our understanding of the intervention.

Peacebuilding programs are generally already under-resourced. Isn't an RCT too expensive?

RCTs do require additional funds, on top of a typical M&E system. However, the extra amount required, in the case of the ECPN RCT, was not orders of magnitude higher than existing M&E costs. Additional funds were needed to cover the costs for of a research consultant to lead the design, as well as cover the costs of surveying, collecting the observational data, and implementing the behavioral game in control sites, in addition to the treatment sites. But mostly, the program drew on its existing M&E budget of about 5% of the program's cost, though with intensive support from Mercy Corps' in-house research team. One lesson

baseline similarity increases our confidence that committee membership, not pre-existing dispositions, explains differences in attitudinal and behavioral change between committee members and non-committee from baseline to endline.

learned for Mercy Corps was that keeping track of randomized participants (see section above) would have benefited from hiring and funding one full-time staff person in-country for that sole purpose. A related challenge is that peacebuilding programs tend to be small, and this makes RCTs difficult because programs cannot reach enough communities to provide enough statistical power for rigorousto research. This was true in the case of ECPN--with only 30 communities participating in the research--but the researchers found creative methodologies to account for these small numbers. Ideally, a larger program with greater scale would provide an opportunity for more robust results.

Recommendations

The following summary of lessons is designed to help practitioners, donors, and researchers as they consider implementing an RCT of a peacebuilding program:⁹

- 1) When considering an RCT of a peacebuilding program, look for these favorable conditions:
 - a) The conflict situation, including the main conflict actors and groups, locations, and drivers, is stable enough to conduct an RCT. If the environment is less stable, the intervention evaluated would need to be discrete and short, around six months or less.
 - b) The program or stream of activities is based on a theory that has broad relevance to the peacebuilding community; and
 - c) The program has a wide enough reach across communities that it will yield a large enough sample size.
- 2) Hire program staff who are committed to learning and are unafraid to confront failure, and help build their research knowledge and skills so they can fully contribute to the research objectives.
- 3) Hire a dedicated M&E staff person (or people) who can track communities and participants to reduce the amount of attrition.
- 4) When creating the analysis plan and drafting the methodology, plan for a higher level of attrition than in non-conflict programs.
- 5) Use mixed methods and multiple approaches--in addition to surveys--to take full advantage of a randomized research design and ensure robust, triangulated results.

⁹ International Alert's "<u>Improving the impact of preventing violent extremism programming: A toolkit for design, monitoring, and evaluation</u>" includes guidance for decision-making when considering a counter-factual evaluation method (p. 114). This table is a further helpful reference that is also applicable to many conflict contexts absent violent extremism.

CONTACT

REBECCA WOLFE, PhD Director of Evidence and Influence | Peace and Conflict rwolfe@mercycorps.org

LISA INKS Senior Advisor | Peace and Conflict links@mercycorps.org

About Mercy Corps

Mercy Corps is a leading global organization powered by the belief that a better world is possible. In disaster, in hardship, in more than 40 countries around the world, we partner to put bold solutions into action — helping people triumph over adversity and build stronger communities from within. Now, and for the future.



45 SW Ankeny Street Portland, Oregon 97204 888.842.0842 mercycorps.org

This brief was made possible with the support of the Peacebuilding Evaluation Consortium.

About the Peacebuilding Evaluation Consortium (PEC)

The Peacebuilding Evaluation Consortium (PEC) is a project of Alliance for Peacebuilding (AfP) in partnership with CDA Collaborative Learning Projects, Mercy Corps and Search for Common Ground (SFCG). The project is funded by the Carnegie Corporation of New York (CCNY) and is field-wide effort to address the unique challenges to measuring and learning from peacebuilding programs. The PEC convenes donors, scholars, policymakers, local and international practitioners, and evaluation experts in an unprecedented open dialogue, exchange, and joint learning. It seeks to address the root causes of weak evaluation practices and disincentives for better learning by fostering field-wide change through three strategic and reinforcing initiatives: 1) Developing Methodological Rigor; 2) Improving the Culture of Evaluation and Shared Learning; and 3) Fostering the Use of Evidence to Inform Peacebuilding Policy.