

Evaluating programmes in conflict-affected areas¹

Carlos Bozzoli², Tilman Brück³ and Nina Wald⁴

HiCN Working Paper 142

April 2013

Abstract: We provide an overview over the concepts, pitfalls and challenges involved in conducting scientifically sound impact evaluations and randomized controlled trials (RCTs) in conflict-affected and fragile states or areas. We consider both interventions aimed at preventing and overcoming war, conflict and insecurity as well as interventions targeted at common socio-economic outcomes such as livelihoods, employment, microfinance, health or poverty which happen to be in insecure or post-conflict reconstruction environments. Surprisingly, this is a significant methodological research gap. Generating sound lessons from such interventions requires careful research design and a profound understanding of the underlying institutional environment. The good news is that rigorous impact evaluations and RCTs are possible even under adverse circumstances. We offer some specific suggestions and guidelines in this paper on how to achieve this. There are many opportunities for better interventions to help overcome the development challenges of violence and conflict.

This paper is forthcoming in P. Justino, T. Brück and P. Verwimp: XXX, Oxford University Press, Oxford.

¹ We are grateful for helpful comments from Andrew Tedesco.

² German Institute for Economic Research (DIW) and Households in Conflict Network

³ Stockholm International Peace Research Institute – SIPRI and Households in Conflict Network (director@sipri.org)

⁴ German Institute for Economic Research (DIW), Humboldt-University Berlin and Households in Conflict Network

Introduction

Recently, development economists have advocated an empirical approach to testing the effectiveness of public policy. Its advocates claim that effectiveness and efficiency of policies should be ‘tested’ in the field. Such testing – they suggest – should follow evaluation procedures resembling experiments in the medical sciences called randomized controlled trials (RCTs). That is, before approving a certain policy, the potential beneficiaries of a policy should be randomly assigned to two groups – ‘treated’ and ‘untreated’ individuals. After implementation, the results of the policy are based on a comparison between these two groups. Effective policies should improve the standing of the ‘treated’ group vis-à-vis the ‘untreated’ group.

The theoretical appeal of RCTs lies in eliminating confounding factors that potentially plague uncontrolled studies in the social sciences. Section 1 of our paper explains in detail why RCTs may give a better understanding of policy questions. Our approach here is general and does not deal with the specific challenges of research in conflict-affected areas. We provide two different explanations, one based on intuition (math-free) and another one based on statistical and mathematical tools. Our intention is to provide the reader with an introduction to the basic elements of this methodology, citing further references for those that are interested in a thorough study. However, RCTs are not free from problems when implemented in ‘real life’ situations. Therefore, in this section we also pay attention to potential pitfalls in the implementation of this methodology and its general limitations such the question of external validity.

In Section 2, we go beyond the uni-dimensional differentiation between ‘treated’ and ‘untreated’ groups. In particular, we add the conflict dimension – the focus of this volume – and analyse

different scenarios: studies based in conflict regions only, studies based in post-conflict regions, and a scenario where programmes are evaluated in a country with conflict and non-conflict regions. We posit that not accounting for conflict in impact evaluations may yield biased results. For instance, having experienced a violent conflict, farmers may have different time preferences or may prefer assets which can be concealed in the case of renewed fighting. Such changes may yield different behavioural responses to policy interventions than would occur in peace times. As the baseline case of Section 1, we consider studies where conflict is not important. Furthermore, we examine RCTs in conflict-affected areas which neglect to account for conflict (which, unfortunately, is quite common). For each of these scenarios, we discuss recent research addressing topics that range from reconstruction and its effect on violence to reintegration of special groups affected by conflict (such as combatants or women).

Finally, we consider in Section 3 practical issues to be considered in conducting impact evaluations under extreme circumstances such as those in conflict and post-conflict countries.

Overall, this paper is only a brief introduction to the tools used in randomized evaluation. The interested reader can read Duflo, Glennerster and Kremer (2006) or the book of Gertler et al. (2011). We also offer only a brief discussion of how to measure conflict in household surveys, which is discussed more fully in Brück et al. (2010).

Our main message in this paper is that, in comparison to many research environments where peace prevails, fairly little robust evidence from randomized controlled trials has been generated in conflict-affected areas. This is a significant shortcoming of both the research and policy

communities. Conducting serious and policy relevant research is possible even in fragile and conflict-affected countries and much can be learnt about policy interventions even in these settings. In fact, much should be learnt about these interventions given the extreme human misery concentrated in conflict-affected countries and the large sums of money spent on interventions in these countries, despite the absence of clear evidence of how best to help people achieve development under such circumstances. Used wisely, impact evaluation in conflict-affected countries is a feasible and useful tool for generating knowledge and improving policy concerned with how to escape the conflict cycle.

1 Methodology

Most studies in social sciences are based on observation; yet establishing causality between observed variables is extremely challenging. For example, one has access to survey data (information collected from a given population using a questionnaire), including individual records for years of completed education and wages. One may observe that individuals earning higher wages tend to have attained more years of schooling. This observation does not imply a causal relation from schooling to wages: one cannot conclude from this evidence that an additional year in school results in higher wage earnings. Third factors may influence both schooling and wages, so that the relation between education and earnings may not represent a causal relation, but the impact of other factors on these outcomes. This phenomenon is known as confounding: the effect of third factors blurring any inference that may be drawn from the relationship of two variables (in our case education and wages).

Randomized evaluations are becoming a popular research strategy in social sciences because they minimize the chance that a confounding variable is responsible for an outcome. They are used to understand the impact that a particular programme or policy proposal has on one or several outcomes of interest. The aim of randomized evaluations is to provide an accurate estimate of such an impact. In this sense, it attempts to overcome the problem of confounding. The key element of this strategy of evaluation is randomization, and in the next paragraphs we will see why.

To be concrete, let's concentrate on the following example. Suppose that the programme or policy that we want to study is a job-training programme intended to increase the skills of workers, so that they can earn higher wages. In this simple case, we have two elements of importance:

- a programme or policy: a job-training programme to increase the skills of workers
- an outcome of interest: wage earned per hour

In the next two subsections we explain why randomized evaluations can help social scientists. First, we give an intuitive answer without using basic statistical and mathematical tools. Second, we use tools from statistics to explain in more formal terms why randomized evaluations deal with the problem of confounding.

1.1 An intuitive approach

Suppose that to analyse the impact of the job-training programme on wages (without changing anything else), a researcher enrolls all workers of a firm in a study (with their consent). Say half of

all these workers then attend a training programme. After the training programme finishes, the workers of that firm will belong to two categories: trained and not trained.

In the jargon of randomized evaluations, the job-training programme is called a ‘treatment’ (T). Those having received the treatment are called ‘treated’ (T=1) and those who did not are considered ‘untreated’ (T=0). Some authors call the ‘untreated’ group a control group, for reasons that will be later obvious. Notice that the decision to be treated or not is an individual decision, which may depend on the individual’s characteristics (for example their motivation).

The researchers now intend to discover the impact of their job-training programme on wages and to do so they collect information about workers using a survey questionnaire administered after the training programme was finished. Suppose that they collect information on treatment status (treated, T=1; and untreated, T=0) and wages, w .

Their survey indicates that wages for the treated amount, on average, to \$5.50 per hour and \$5.20 for those not treated. They conclude that the ‘effect’ of the job-training programme on wages was an increase in \$0.30 per hour on wages ($=5.50-5.20$). Is this a reasonable estimate of the job-training programme effect? Possibly not, because the treated group may be systematically different to those in the untreated group. For example, those in the treated group could have more years of job experience than those in the untreated group, and the \$0.30 ‘plus’ may be due to years of experience rather than job training. In this example, this is not a big problem because interviewers could be aware of this and include ‘age experience’ as one of the variables of interest in the study. Then, using statistical methods, they could try to estimate the impact of job

training, taking into account (controlling for) the systematic difference between treated and untreated in terms of job experience.

But treated and untreated individuals may differ in other ways. Suppose that those enrolling in the job programme are more motivated than those that did not. This difference (in motivation) is difficult to measure. Moreover, other traits linked (correlated with) treatment status and wage earned could also exist, reducing the reliability of the results.

These pitfalls can be overcome by allocating individuals into treatment in a random fashion. In this sense the treated and untreated groups will not have systematic differences since the mechanism selecting them out or into treatment does not depend on an individual decision. By randomizing, all individuals have the same (ex-ante) probability of being treated, and this helps to make treated and untreated groups comparable – that is, they do not differ from each other in any systematic way.

Put in another way, individuals in the treatment group will have similar characteristics to those in the non-treatment group, except that one group received ‘job training’ and the other did not. Now, the difference in average wages between treated and untreated groups is a reliable estimator for the effect of job training, since both groups are comparable except for receiving or not the training programme.

1.2 *Mathematical foundations*

To be able to analyse the impact of the training programme on wages, we must ‘keep other factors constant’. In doing so, we implicitly resort to answering the two related counterfactual questions: How would individuals participating in the job-training programme have fared in the absence of the programme? And how would individuals not participating in the programme have fared in the presence of the programme? Answering these questions in real life is not feasible since individuals will either belong to the treated or the untreated group.

To answer these questions, we hence obtain the impact of a programme by comparing the average outcome variable in a group affected by the programme to the average outcome in a group not affected (untreated) that is ex-ante comparable to the treated group. The latter is called the comparison group; a group that would have had similar outcomes to those in the treatment group had the comparison group received treatment. Comparison groups, or control groups, are difficult to construct. Non-treated individuals may systematically differ from treated individuals in many dimensions, many of them not easily observable such as ability, motivation, personality, preferences, or social networks.

Let us call the outcome of individual i as Y_i^T if treated and Y_i^C if not treated. We are interested in the effect of the treatment on the outcome, that is $Y_i^T - Y_i^C$. This is not possible to obtain in reality, since an individual assigned to treatment cannot be also non-treated. In our case, a worker can receive training or not, but in theory it can have two *potential outcomes*, one under treatment, Y_i^T , and another one without, Y_i^C .

However, we could try to see if we can come up with a good population estimator for $E(Y_i^T - Y_i^C)$, where E is the mathematical expectation operator. From observing the data, we have the mean of wages in job trained individuals, which in large samples converges to $E(Y_i^T | \text{received job training})$ and the mean of wages in those that were not trained, that in large samples converges to $E(Y_i^C | \text{did not receive job training})$. Therefore, we have a consistent estimator for $E(Y_i^T | T=1) - E(Y_i^C | T=0)$. Subtracting and adding $E(Y_i^C | T=1)$ we have $E(Y_i^T | T=1) - E(Y_i^C | T=0) + E(Y_i^C | T=1) - E(Y_i^C | T=1) = E(Y_i^T | T=1) - E(Y_i^C | T=1) - E(Y_i^C | T=0) + E(Y_i^C | T=1) = E(Y_i^T - Y_i^C | T=1) + E(Y_i^C | T=1) - E(Y_i^C | T=0)$

Then

$$E(Y_i^T | T = 1) - E(Y_i^C | T = 0) = E(Y_i^T - Y_i^C | T=1) + E(Y_i^C | T=1) - E(Y_i^C | T=0)$$

where

$E(Y_i^T | T = 1) - E(Y_i^C | T = 1)$ can be estimated with the data from the simple

$E(Y_i^T - Y_i^C | T=1)$ is the treatment effect (on the treated), the effect we want to estimate

$E(Y_i^C | T=1) - E(Y_i^C | T=0)$ is a bias between $E(Y_i^T | T = 1) - E(Y_i^C | T = 0)$ and the treatment effect, due to omitted variables. Thus, it is called omitted variable bias (OBV).

We would like to have $E(Y_i^C | T=1) - E(Y_i^C | T=0) = 0$, since then

$$E(Y_i^T | T = 1) - E(Y_i^C | T = 0) = E(Y_i^T - Y_i^C | T=1)$$

But in general, $E(Y_i^T | T = 1) - E(Y_i^C | T = 0)$, that is, the omitted variable bias (OBV) will differ from 0. What is more, one cannot estimate the counterfactual $E(Y_i^C | T=1)$: the outcome in treated individuals had they not received job training.

However, there is one special case where the $OBV=0$: the case where treatment status is randomized. Under randomization, T is a random variable independent from the potential outcomes. Then $E(Y_i^C | T=1) = E(Y_i^C | T=0) \Rightarrow E(Y_i^C | T=1) - E(Y_i^C | T=0) = 0$

As a result,

$$E(Y_i^T | T = 1) - E(Y_i^C | T = 0) = E(Y_i^T - Y_i^C | T=1)$$

The implementation of this result is simple. In a sample of individuals, one randomizes access to treatment (assignment to the programme) across individuals. As a result, some individuals will be treated and some will not. Since the assignment was random, systematic differences between treated and not treated individuals can be neglected.

The theoretical term $E(Y_i^T | T = 1)$ has an empirical counterpart in our sample: it is the average of outcome Y in the treated group. Likewise, $E(Y_i^C | T = 0)$ can be estimated from the average of outcome Y in the non-treated group. Thus, $E(Y_i^T - Y_i^C | T=1)$ is estimated as the difference in means (average Y in treated vs. non-treated groups).

1.3 Threats and potential biases

As we have explained, randomization makes ‘treated’ and ‘untreated’ groups comparable, that is, it eliminates (from an ex-ante perspective) any other systematic differences between these groups, except for two elements. First, one group has been ‘treated’ while the other has not. Second, as a consequence of treatment, the value of outcome Y could differ between groups (in a systematic way). In fact, if this occurs, it shows that the treatment has a non-negligible impact.

Usually, researchers provide sample evidence that treatment and no-treatment groups are ‘similar’. To do so, they provide the means of key characteristics (for example age and sex) and perform equality tests. As expected in a sample, the average of a given variable may be different between groups (because of chance), but this difference should be ‘small’ in statistical terms. If this occurs, it can be taken as evidence that the groups do not differ in a systematic way, at least for the characteristics being tested. But individuals may differ in other characteristics that are difficult to measure, usually referred to as ‘unobservables’: factors that cannot be measured but that may influence outcomes. If randomization is done properly, one should not worry about this, because their effect will be negligible.

Randomization is a powerful tool to obtain unbiased estimates of the impact of a programme. But even if individuals are randomized across groups, problems can occur that threaten the validity of inference made in the study. We will briefly mention some of these problems to make the reader aware of potential pitfalls. However, it is not an exhaustive list; we refer the reader to the specialized literature for a thorough exposition of these potential threats.

Attrition: A programme or treatment can take some time to yield observable results. For example, the training programme may last six months, and the researcher may want to record multiple measurements of the outcome of interest. For example, a baseline (pre-programme) measurement may be taken, followed by readings at month 6, 12, 18, and so on. Individuals enrolled in the programme may drop out of it (incomplete treatment), or they may not be reachable during or after the treatment, and therefore no measurements of outcome Y will be available. Even if multiple measurements of Y are taken over time on some individuals, some of these may not be taken on other individuals. This can also happen with individuals assigned to the ‘untreated’ group. The problem here is whether the individuals dropping out of the study differ from those staying on it. If there is no systematic difference between drop-outs and stayers (consider the case of both ‘treatment’ and ‘control’ groups), the attrition can be considered ‘random’, and this will have negligible impact on the results of the study (estimates will be reliable, or ‘consistent’). If attrition is not ‘random’, different techniques could be applied to ameliorate potential bias. However, these techniques have limitations. Researchers should plan ahead and implement measures to reduce the magnitude of attrition by employing ‘retention strategies’.

Diffusion or contamination: This problem occurs when individuals in either group ‘influence’ those in the remaining group. For example, non-participants in our ‘job-training programme’ may ‘learn’ skills from participants in the programme. This could occur if, for example, there are ‘treated’ and ‘untreated’ individuals living in the same household. If the training programme ‘works’ (learning job skills increase wages), and there is a ‘diffusion problem’, untreated individuals will also benefit from the programme (their skills will improve as result of the

‘contamination’ and their wages will rise). This will result in downward biased estimates of the programme, since wages for untreated individuals will be higher than if no ‘diffusion’ had occurred. To avoid this problem, researchers usually design randomization so that ‘treated’ and ‘untreated’ units are distant from each other. For example, treatment may be allocated to villages and all individuals (or relevant demographic group) in the village be offered treatment. ‘Treated’ and ‘untreated’ villages may not be close to each other: for example, it should be difficult for individuals from both villages to meet or interact in food markets. However, each programme being evaluated is particular and provisions to avoid diffusion or contamination should be made on a case by case basis.

Resentment: this problem occurs if ‘untreated’ individuals feel they are treated unfairly and this affects their behaviour in such a way that it alters outcomes. In our job-training case, individuals not being offered training may become less productive, or exert less effort (either knowingly or unknowingly), and their wages may be affected. An opposite effect may occur in treated individuals. This problem may not occur in every setting, and its magnitude/plausibility may be limited or negligible in certain cases. However, it is important to gauge the potential of resentment before embarking on fieldwork.

Finally, another problem of randomization is that of *external validity*: If a policy is evaluated in a given social context, it does not follow that the results can be extrapolated to other situations. Countries, and even regions within a given country differ in economic, social and cultural realms, and there is no reason why such differences might not affect the effectiveness of a programme (or the magnitude of it).

2 Theory

There are many issues which have to be taken into account when planning, implementing and evaluating programmes in conflict-affected areas, such as the nature of the conflict (as argued elsewhere in this volume). For instance, if a key impact of a conflict is displacement, this may be quite obvious to the researcher, yet providing evidence may pose challenges of attrition. If the key impact of conflict is fear, researchers may be less aware of this, yet such fear may drive key behavioural responses of study participants. In this section, we will first outline five possible scenarios of evaluating programmes in conflict areas. After this, we will explain and analyse explicitly the special requirements for project evaluation in conflict and post-conflict countries in order to be beneficial for the people participating in these programmes.

2.1 Usual scenario

Table 1

	Treatment	Control
Conflict		
No Conflict		

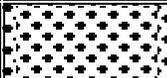
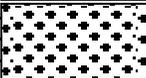
Source: Authors' own.

The great majority of development projects and programmes are undertaken in non-conflict areas. One example of this type of evaluation is Conditional Cash Transfer (CCT) programmes

which have been implemented and evaluated in many developing countries, especially in Latin America. In general, CCT programmes just measure the overall programme effects and do not allow heterogeneity for variables different from individual, household or programme level. Evaluation of these programmes is mainly done as described in the former section on methodology.

2.2 *Neglecting-conflict scenario*

Table 2

	Treatment	Control
Conflict		
No Conflict		

Source: Authors' own.

Some programmes are implemented in (former) conflict-affected areas but do not take into account possible effects of conflict on the programme explicitly. One example for this category is Colombia's CCT programme 'Familias en Acción'. To the best of our knowledge there is no differentiation between conflict- and non-conflict-affected communities when evaluating programme effects on child health and education.

However, there might be some problematic issues associated with this approach. On the one hand, by neglecting the conflict dimension, the impact of the programme might actually be underestimated because some of the positive effects of the programme are taken away by the

negative impact of conflict when just using the usual econometric tools. On the other hand, a programme could even have negative impacts on the people if it reinforces some of the negative consequences of the conflict.

Just a simple example: imagine that there is a CCT programme intended to increase school enrolment for secondary school. In non-conflict areas, school enrolment increases due to the programme delivering the desired positive impact. In conflict regions, however, it might be more difficult for teenagers to attend school for various reasons (for example it may be dangerous to go to school or be in school, more profitable to participate in armed groups, there may be a higher necessity to help in the household compared to peaceful regions, and so on). Thus, programme effects in conflict settings are smaller or even insignificant. If large parts of the programme actually take place in conflict areas, the overall effect of the programme might be relatively small even though the programme is very successful in peaceful regions. If worse comes to worst it could also be the case that the programme has a negative impact in conflict areas, for example if the money given to the participants is taken away by armed groups or if the money is used for conflict-related issues like buying drugs or arms – even if students do attend classes. By not taking into account conflict explicitly, these negative impacts might have been missed since they are not reflected in lower enrolment rates. Yet, when they do result in lower enrolment rates for secondary school the impact of the programme as a whole may turn out to be negative even though for peaceful regions (the intended target of the programme) it was planned appropriately. In a nutshell, leaving aside relevant conflict-related aspects is sub-optimal for designing, implementing and evaluating a programme.

2.3 Conflict scenario

Table 3

	Treatment	Control
Conflict		
No Conflict		

Source: Authors' own.

In the third scenario, the programme is carried out only in conflict-affected areas. The majority of projects in this category try to reduce conflict and prevent counter-insurgencies. The hope is that a programme has a conflict-reducing effect, or at least that a project does not fuel conflict. In these cases, conflict is the dependent variable and the interesting issue is to find the impact of a development programme on conflict.

In this strand of research, many authors make their assumptions based on two different theories: the first one is the so-called 'hearts and minds theory' which suggests that governments can win the hearts and minds of a population by the successful implementation of development projects. Two papers support this hypothesis at least partially. First, Berman, Shapiro and Felter (2011) find that in Iraq the provision of services leads to a decrease in violence. However, reconstruction spending does not have a conflict-reducing effect. According to Berman, Shapiro and Felter (2011) it would be very important for policy makers to know the cause of this failure, which may include poor local knowledge or poor oversight. Second, Beath, Fontini and Enikolopov (2012) find that in Afghanistan the National Solidarity Program – a large

development programme targeting the delivery of services and infrastructure – leads to a reduction in violence if the level of violence was not too high initially.

The second theory is called the ‘bargaining model approach’ which suggests that the introduction of development projects in conflict-affected countries can alter the balance of power among the government and the rebel groups. As a consequence, the implementation of development projects might lead to an increase in violence in the short run if the project weakens the power of one party. Crost, Felter and Johnston (2012a) provide some evidence for this hypothesis. They analyse the impact of a big nationwide anti-poverty programme (KALADHI-CIDSS) on conflict in the Philippines. This programme targets the country’s poorest provinces with the aim to improve local infrastructure, governance, participation and social cohesion. The authors of the study find that provinces receiving the programme experience an increase in conflict casualties in the short run. Their explanation for this result is that rebels anticipate the weakening of their position by a successful implementation of a development project. As a consequence they try to prevent the project implementation which leads to an increase of violence in these areas. In the long run, however, once the project has been implemented, conflict levels in treatment and control municipalities are no longer significantly different.

Another paper by Crost, Felter and Johnston (2012b) investigates the impact of a Conditional Cash Transfer programme on conflict in the Philippines. Contrary to their results for the KALADHI-CIDSS programme, they find that the CCT programme leads to a reduction of conflict in treatment communities and that there is even a spillover effect to nearby villages which did not receive the programme. They assume that this programme was more successful in

decreasing conflict because it is less visible to insurgents than infrastructure projects. As a consequence it is more difficult for rebels to find a target and destroy the project. Additionally, cash transfer recipients might be less willing to join rebel groups so that the support and the size of armed groups shrinks, leading to a reduction of conflict in the long run.

2.4 *Post-conflict scenario*

Table 4

	Treatment	Control
Post-Conflict		
No Conflict		

Source: Authors' own.

These projects dealing with post-conflict scenarios comprise topics directly related to conflict like demobilization, displacement, peace building, food aid, infrastructure and social trust as well as the usual programmes on education, health, job training and gender. Naturally, in the first group of variables, causes and consequences of conflict take an important role in the evaluation process, whereas in the second group of variables, conflict does not necessarily play a key role in measuring programme impacts but could just be added as a control variable. Realistic scenarios for this type of study are countries which have been affected by conflict as a whole and want to improve their citizens' living conditions by implementing suitable projects. In this case, the inclusion of conflict as a control variable is the adequate choice in order to avoid an omitted variable bias and to take into account the past intensity of conflict in different regions. Here, the

impact of conflict on the project is not in the centre of the study, but the country is ‘by coincidence’ affected by conflict and thus it has to be taken into account. Yet, the majority of studies in this field of research address topics directly related to particularities of post-conflict zones.

In the following paragraphs, we will present three examples of programmes putting emphasis on different subgroups of the population and targeting various aspects of living in a post-conflict environment. The first one is a project carried out in post-conflict Liberia and evaluated by Fearon, Humphreys and Weinstein (2009). They use a randomized field experiment in a former conflict prone area in northern Liberia to measure the impact of a community-driven reconstruction programme. These programmes focus on supporting participatory processes to improve local provision of public goods and eradicate poverty and are often implemented in post-conflict countries. The idea behind these programmes is that community members start cooperating and helping each other for a common benefit. The authors of the study come to the conclusion that these projects indeed foster social cohesion among community members which persists after the programme ended.

Another project in Liberia aims to reintegrate young ex-combatants and is evaluated using a randomized controlled trial by Blattman and Annan (2011). The project consists of agricultural training, literacy and numeracy courses, psychosocial counselling and material support during the training. After having completed the training, each person receives a set of agricultural tools needed for sustainable agriculture. Results indicate a higher engagement in agriculture for programme beneficiaries, slightly higher social engagement, and less interest in participating in

the election violence in neighbouring Cote d'Ivoire. However, due to limited access to markets, there was no significant increase in income.

The third programme – Women's Income Generating Support Project (WINGS) – was implemented in Northern Uganda to help poor displaced women to build small businesses and is still on-going (AVSI Uganda 2010). Beneficiaries receive business skill training and have to develop a business plan afterwards in order to get grants for their enterprises. Annan, Blattman, Green and Jamison will conduct a randomized impact evaluation in a post-conflict setting to find out what is working for whom and why (see <http://chrisblattman.com/projects/wings/>, accessed 24 November 2012). Preliminary results show that participants managed to increase their earnings, savings and consumption relative to the control group. However, there is little impact on empowerment, psychological well-being and social integration. One reason for these ambiguous results might be that participants experience higher levels of community hostility and support both at the same time.

Programme effects in these examples are just measured in former conflict areas. The comparison group lives in post-conflict areas as well so one cannot compare the size of these effects with those in non-conflict areas.

2.5 *'Perfect' scenario*

Table 5

	Treatment	Control
--	-----------	---------

Conflict		
No Conflict		

Source: Authors' own.

There might be some situations where one only has the means to implement one programme and has to choose the most efficient way to support the most people with the least amount of money. This is a possible scenario in a country which is partly affected by conflict and partly peaceful (for example, Colombia, Uganda, India) and policy makers have to decide whether to help the few heavily deprived people in conflict-affected areas or many moderately deprived people in peaceful regions. For these reasons it might be very useful to have control and treatment groups in peaceful and conflict-affected regions as displayed in the 'perfect' scenario.

This scenario could be the best one for countries which are partly affected by conflict since, as mentioned already, one is able to compare programme outcomes in conflict and non-conflict areas. By analysing the outcomes one can see what works in conflict areas and what has to be adjusted to these special circumstances people are living in. Additionally, one is not confusing the results of conflict with peaceful regions as in the neglecting-conflict scenario, avoiding the risk of under- or overestimation of programme effects due to conflict.

The question is why are there almost no evaluations taking into consideration conflict and non-conflict-affected areas. One explanation may be that there are few countries which have conflict in one area and peace in another area at the same time. Second, it might be difficult to implement a programme in conflict areas for various reasons which we will analyse in more detail later on.

Still, it would be a great advantage to split evaluation in post-conflict and non-conflict regions. Third, it might be more costly to implement programmes in conflict and non-conflict regions at the same time so that some programmes are just restricted to peaceful regions and those programmes dealing with the consequences and causes of a conflict might just be implemented in (post)-conflict parts of the country. Different programmes make it difficult to compare their efficiency.

Other reasons why evaluations are not done this way might be political or ethical reasons. For policy makers it might not be the best strategy to deprive parts of the population of a programme just because it is not as efficient as in other parts of the country.

All these examples show that many factors have to be taken into account when implementing development programmes in conflict-affected countries. In the remaining paragraphs of this paper, we will elaborate some of the major issues in conflict-affected settings. These issues concern the country context specific problems of the project, choosing the type of programme, problems with monitoring and evaluation, and difficulties with local partners and the government.

3 Practice

By contextual issues, we mean the peculiarities of conflict-affected areas. The first and probably most obvious one is the lack of security in conflict zones. Our guess is that many projects are not implemented in conflict-affected regions due to a high level of insecurity for staff members and an associated high risk of project failure. Moreover, it might also pose a danger for programme

recipients if one of the conflicting parties is against the project and thus threatening its participants. As a consequence, people decide not to engage in project activities and the project does not have the expected benefit for the population.

Another often observed development in conflict-affected countries is the economic and social instability which makes it difficult to plan and implement long term projects. These areas may lack adequate infrastructure and functioning markets, suffer from high levels of corruption and have weak state capacity, all of which are important for conditional cash transfer programmes, for example. In the course of a project, the possession of the programme area could change, causing one to deal with different parties. Voluntary and forced displacement is often a consequence of violent attacks leading to a dramatic change in project participants. All these circumstances make it complicated to plan, implement, monitor and evaluate projects in conflict-affected areas.

To sum up, it is very important to include the characteristics of the respective conflict, the balance of power between the conflicting parties, the economic and social situation in conflict areas, and the timing of the project, as well as perceptions of the population regarding the project.

3.1 Choosing the type of programme

As already mentioned by Crost, Felter and Johnston (2012b), it could make a difference for the success of a programme whether you implement CCT programmes or infrastructure projects. Infrastructure projects are more visible to people and could thus be targeted and destroyed by

rebel groups more easily. On the other hand, the visibility of infrastructure projects and the provision of services could also be a signal to the population that circumstances are changing for the better and that the government and local partners are willing to support the conflict-affected population. It could also be a signal to the rebels from the government that it is not willing to give up on these areas and its citizens. For example, Böhnke, Koehler and Zürcher (2010) find that in Afghanistan aid helps the Afghan district and provincial government to be seen in a positive way, which is one reason to implement a more visible project.

Willibald (2006) and Haider (2011) collect some pros and cons of cash transfers in conflict-affected regions: (i) weak banking systems making it difficult to deliver cash to recipients, (ii) states lacking the ability to implement the programmes, (iii) the likelihood of creating inflation in instable markets, (iv) the possibility of corruption and diversion of money by irregular armed groups and civil servants, and (v) high migration rates making it difficult to target the right population group. Haider (2011) mention some measures/preconditions which could turn CCT programmes into a successful instrument in conflict-affected countries. First, when implementing CCT programmes it is very useful to not only look at the demand side (the CCT programme) but also to change the supply side accordingly in order to satisfy needs resulting from the increased availability of cash. Second, it is crucial to ensure a safe delivery of money to recipients. Third, CCT programme planners should address the issues of corruption and weak state capacities. Last, but not least, it is essential in conflict-affected areas to find an equilibrium between poverty reduction and social cohesion. According to Crost, Felter and Johnston (2012b), it was an advantage of the CCT programme in the Philippines that it was less visible than other projects

and thus less vulnerable to attacks because, in general, programme participants were not known to the public when they went to the bank to get the programme money.

Willibald (2006) summarizes the potential advantages and disadvantages of using cash in Disarmament, Demobilization and Reintegration (DDR) programmes. Implementing CCT programmes is expected to be less troublesome in post-conflict countries with weak institutions and missing infrastructure than commodity based programmes. In the long run, cash might even create a demand for the establishment of institutions like banks, health centres, schools, and so on. Cash given to ex-combatants can also facilitate the process of reintegration in their families, since they do not have the burden of supporting a ‘new’ family member without any additional means. During the disarmament and demobilization process, cash might help the willingness of ex-combatants to hand in arms and move back to their former communities. But there are also some potential risks associated with giving out cash to ex-combatants. The cash might allow them to buy newer and better arms, leading to the creation of illegal weapons markets and trade. Ex-combatants receiving cash might lead to social unrest in their communities if the ‘peaceful’ members do not receive support and feel disadvantaged. Moreover, ex-combatants receiving subsidies may have little incentive to find a job and/or little experience to manage funds successfully. That is why many programmes do not just distribute cash but combine transfers with some kind of training. According to Willibald (2006), it is not a question of whether to implement cash transfers but how to implement them.

As we can see from the examples in this paper, different environments require different programmes. Most likely the best strategy – but also the most expensive one – is to implement different types of programmes in the same areas which complement each other.

3.2 *Evaluation*

As can be imagined, obtaining reliable data in conflict-affected regions is very challenging, for various reasons. First, establishing a baseline is harder than in stable environments. In insecure and volatile contexts, people migrate more frequently, leading to higher attrition rates and to selection bias if those who move are different from the group of stayers. Over time, the economic and social situation of participants may change dramatically in conflict-affected areas so that programme goals have to be adjusted accordingly to satisfy the needs of people. Of course, this makes it more difficult to analyse the impact of the project. In general, measuring the impact of a programme in conflict-affected regions is a complex enterprise because many different factors are at play influencing the outcomes. Therefore, programme evaluators should be extremely careful not to neglect the conflict-sensitive programme impact which may either fuel or reduce conflict. In order to adjust to changing programme objectives and number of participants, data collection should be on a regular basis to get reliable results.

Second, one might encounter problems in finding reliable secondary data, or in some cases getting access to secondary data at all. Additionally, data from the government might be manipulated in a certain way. In this situation, it is best to collect as many data as possible from different sources and compare them in order to get an idea of the magnitude of the values and the economic and social situation of communities.

Third, it is of great importance to choose the adequate conflict indicators for measuring conflict and conflict-affected participants. As with other secondary data, conflict data could be subject to manipulation and measurement error. One solution is to take data from different sources and construct an index. Another approach is to mix quantitative and qualitative (for example threat perception by the population) measures in order to get a more reliable estimate of conflict (DFID 2009).

3.3 Monitoring and project partners

Finally, when implementing a project, it is essential to have a reliable and regular monitoring system to guarantee the success of the project. Therefore you need trustworthy local partners from governments, NGO's or international organizations. It is not only important that project owners trust local partners but also that project participants do as well. Therefore, it would be desirable that local partners are impartial and not involved in the conflict, which is sometimes hard to find.

4 Lessons learnt

In this paper, we introduced the basic tools of randomized evaluation both in peaceful and conflict-affected settings, providing examples from the literature on conflict and development. Randomized evaluation serves a useful purpose, in peaceful and conflict settings alike, for identifying causality from a programme to an outcome, controlling for various outside influences. There are a few lessons learnt from the exposition above.

First, it is important to stress the need for careful preparation of randomized controlled evaluations in conflict-affected areas. Researchers and donors must be aware of security threats that may compromise fieldworkers and even the subjects of the study. Contextual issues (discussed in Section 2) should be carefully taken into account and contingency plans should be drawn up accordingly.

Second, and like in any other evaluation study, outcomes of interest (including unintended side-effects) should be clearly outlined at the design stage of the study. Moreover, it is important for the researcher to be able to link policy and outcomes, understanding the process behind the results. Often, the appeal of RCTs lies not only in the outcomes (whether the policy is effective or not) but also in an explanation of the mechanism linking policy and results, based on survey data. The researcher should be able to track ‘mediating factors’ that could explain the success or failure of the programme. For example, a schooling programme may fail because perceptions of security prevent children from going to school, or because the programme does not cover the cost of school related items. Explaining failures helps design better policies in the future.

Third, the period elapsing between the implementation of a given policy and the materialization of results is variable and depends on the policy and context in which it is applied. A baseline survey is necessary to establish that control and treatment groups are similar to each other in terms of observables. At this stage, individuals participating in the survey should be made aware that they will be followed in the future to monitor policy outcomes. One of the key aspects behind success in the implementation of a study is a good ‘sample retention’ strategy, that is, plans to avoid drop out of interviewees between survey waves. It may also be necessary to do

several waves of follow-up interviews to check when the impact may materialize (if at all). Conduction of only a single follow-up survey risks missing important effects occurring at a different point in time.

Fourth, data matter. A randomized evaluation is based on survey information. The survey should be designed by an interdisciplinary group and a pilot survey should be run to evaluate whether the interview collects all relevant information. In conflict areas census data may be outdated, and so the researcher may find useful the guidance of NGOs and agencies (for example WFO or UNOCHA) to select sampling clusters.

Finally, the single most important issue in the design of any study in a conflict-affected area is to account for the conflict dynamics. Disregarding conflict will yield biased research findings if conflict shapes institutions and people's behaviour and well-being. Measuring conflict requires a strong awareness of conflict dynamics and a careful design of the household survey. In that sense, we see our paper as a compendium to the related paper on measuring conflict in household surveys (Brück et al. 2010). Measuring conflict in surveys and designing impact evaluation studies are two important and related tools in a researcher's tool kit which can help lead to a better understanding of how conflicts shape development.

References

- AVSI Uganda (2010), *Women's Income Generating Support (WINGS) Project Report*, <http://www.avsi-usa.org/docs/pdf/WINGSReport.pdf> (accessed 24 November 2012)
- Beath, A.; Fontini, C. and Enikolopov, R. (2012), *Winning Hearts and Minds through Development? Evidence from a Field Experiment in Afghanistan*, Research Policy Working Paper 6129, The World Bank
- Berman, E.; Shapiro, J. and Felter, J. (2011), 'Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq', *Journal of Political Economy* 119(4): 766–819
- Blattman, C. and Annan, J. (2011), *Reintegrating and Employing High Risk Youth in Liberia: Lessons from a randomized evaluation of a Landmine Action agricultural training program for ex-combatants*, Evidence from Randomized Evaluations of Peacebuilding in Liberia, Policy Report 2011.1, Innovations for Poverty Action, Yale University
- Böhnke, J.; Koehler, J. and Zürcher, C. (2010), *Assessing the Impact of Development Cooperation in North East Afghanistan 2005 – 2009: Final Report*, Evaluation Reports 049, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn
- Brück, T., P. Justino, P. Verwimp and A. Avdeenko (2010), 'Identifying Conflict and Violence in Micro-Level Surveys'. HiCN Working Paper 79

- Crost, B.; Felter, J. and Johnston, P. (2012a), *Aid under Fire: Development Projects and Civil Conflict. Revise and Resubmit AER*
- (2012b), *Government Transfers and Civil Conflict: Experimental Evidence from the Philippines*, Unpublished Manuscript,
<http://patrickjohnston.info/materials/4Ps.pdf> (accessed 25 November 2012)
- DFID (2009), *Working Effectively in Conflict-affected and Fragile Situations*, A DFID Practice Paper, Department for International Development, UK
- Duflo, E; Glennerster, R. and Kremer, M. (2006), *Using Randomization in Development Economics Research: A Toolkit*, MIT Department of Economics Working Paper 06-36, Massachusetts Institute of Technology, Cambridge MA
- Fearon, J.; Humphreys, M. and Weinstein, J. (2009), ‘Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-conflict Liberia’, *American Economic Review* 99(2): 287–91
- Gertler, P.; Martinez, S.; Premand, P.; Rawlings, L. and Verweers, C. (2011), *Impact Evaluation in Practice*, The International Bank for Reconstruction and Development / The World Bank, Washington DC
- Haider, H. (2011): *Cash transfers in fragile/conflict-affected environments. Helpdesk Research Report*. Governance and Social Development Resource Centre
- Willibald, S. (2006), ‘Does Money Work? Cash Transfers to Ex-combatants in Disarmament, Demobilisation and Reintegration Processes’, *Disasters* 30: 316–39