Reference Paper for the 70th Anniversary of the 1951 Refugee Convention

Impact Evaluations in Forced Displacement Contexts:

A Guide for Practitioners

Paola Elice, Impact Evaluation Specialist, World Bank Fragility, Conflict, and Violence Unit

October 2021

Abstract:

This paper reviews evidence on program effectiveness generated by impact evaluations and how this evidence has informed program design and policy in low- and middle-income country contexts. It recognizes that there is a growing number of impact evaluation initiatives to study forced displacement programs. It reviews examples of this work and reflects on differences, challenges, and innovations that characterize these impact evaluations. The paper shows that impact evaluations are starting to generate much-needed evidence on program effectiveness in displacement contexts and helping to inform resource allocation for the benefit of displaced people and host communities.

This work is part of the program "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership''. The program is funded by UK aid from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), it is managed by the World Bank Group (WBG) and was established in partnership with the United Nations High Commissioner for Refugees (UNHCR). The scope of the program is to expand the global knowledge on forced displacement by funding quality research and disseminating results for the use of practitioners and policy makers. This work does not necessarily reflect the views of FCDO, the WBG or UNHCR.

Contents

Acronyms and abbreviations

3ie	International Initiative for Impact Evaluation
ACLED	The Armed Conflict Location & Event Data Project
CCT	Conditional Cash Transfer
CDD	Community-Driven Development
CEGA	Center for Effective Global Action
CFS	Child Friendly Spaces
CGD	Center for Global Development
CRRF	Comprehensive Refugee Response Framework
DIME	The World Bank's Development Impact Evaluation Unit
DRC	Democratic Republic of the Congo
DRDIP	Development Response to Displacement Impacts Project
EGAP	Evidence in Governance and Politics
FCDO	The United Kingdom's Foreign, Commonwealth and Development Office
FCV	Fragility, Conflict and Violence
GCR	Global Compact on Refugees
ICRC	International Committee of the Red Cross
IDA	International Development Association
IDMC	Internal Displacement Monitoring Center
IDP(s)	Internally Displaced Person (People)
IE(s)	Impact Evaluation(s)
IPA	Innovations for Poverty Action
IPV	Intimate Partner Violence
IRC	International Rescue Committee
JDC	UNHCR-WB Joint Data Center on Forced Displacement
J-PAL	Abdul Latif Jameel Poverty Action Lab
LMCI(s)	Low- and Middle-Income Country(ies)
LSHTM	London School of Hygiene and Tropical Medicine
M&E	Monitoring and Evaluation
MISFA	Microfinance Investment Support Facility for Afghanistan
NGO	Non-Governmental Organization
PDS	Public Distribution System
POWER-J	Providing Opportunities with Education for Refugees and Jordanians
RCT	Randomized Controlled Trial
RDD	Regression Discontinuity Design
SWIFT	Survey of Wellbeing via Instant, Frequent Tracking
TaRL	Teaching at the Right Level
TUP	Targeting the Ultra-Poor
UK	United Kingdom
UN	United Nations
UNHCR	United Nations High Commissioner for Refugees
US	United States
USD	United States Dollar
VAM	Vulnerability Analysis Mapping

WB	World Bank
WBG	World Bank Group
WFP	World Food Program
WfWI	Women for Women International
WHO	World Health Organization

Introduction

By the end of 2020, 82.4 million people, of which 34.6 million children (42%), had fled their homes due to war, conflict, persecution, violence, and human rights violations (UNHCR, 2021). This number is more than double the amount of forcibly displaced reported in 2010 (41 million) by the United Nations High Commissioner for Refugees (UNHCR). This scale of displacement hasn't been experienced since post-World War II. Among the displaced, an estimated 48 million (58%) were still living within the borders of their own country (or internally displaced), 26.4 million (32%) were refugees in a foreign country, 4.1 million were asylum seekers and 3.9 million were Venezuelans displaced abroad (UNHCR, 2021).

Although most of the media coverage centers on the spill over of refugees into Europe and the United States—for instance the arrival of Syrians to Europe following the Syrian war—the majority of those displaced (8.6 in 10) remain in fragile, conflict and violence-affected (FCV) countries. They also live as refugees in neighboring low- and middle-income countries (LMICs). Most displacement crises are protracted in time, with more than 3 in 4 refugees (76%) residing in host countries for more than 5 consecutive years[1].

Until recently, forced displacement has been seen as a humanitarian issue. This is now changing due to the protracted nature of forced displacement situations and how they threaten the economic and social development of LMICs. In 2018, after two years of extensive consultations led by the UNHCR, the United Nations General Assembly affirmed the Global Compact on Refugees (GCR) which forms the institutional basis to design more sustainable policy responses in contexts characterized by forced displacement.

The GCR states that going forward, the governments of countries affected by forced displacement, humanitarian organizations and organizations with a mandate to promote economic development, like the World Bank Group, will coordinate the delivery of their programs to ensure a more sustainable and predictable policy response that relies on strengthening existing national systems to meet the needs of both displaced and host populations. This enhanced policy response will be anchored on the principle of 'shared responsibility' and on an effort to employ resources as effectively and efficiently as possible (Global Compact on Refugees, 2018).

But how should the new approach be implemented in practice? How can governments make sure funds are channeled to programs that are effective and efficient? How can governments know what programs are good value for money? To guide policy and program implementation and to ensure resources are deployed to bring the largest possible benefit to the biggest number of people in need, it is important to rely on evidence on program impact to guide policy making.

In 'stable' LMICs, starting in the mid-90s, impact evaluations (IEs) – especially randomized evaluations or randomized controlled trials (RCTs) (impact evaluations based on random assignment of a program to a target population[2]) – have been applied to rigorously measure program impact and to inform program implementation, including decisions to scale, course-correct or cancel a program based on the evaluation findings.

The effectiveness of impact evaluations on programs in 'stable' LMICs is widely documented: for example, the results of impact evaluations of microfinance interventions debunked these programs by showing that the actual impact of these interventions was not as positive as what previously widely thought; on the other hand, the reassuring findings have contributed to the scale-up of cash and graduation programs and important health and educational programs such as deworming interventions, remedial education interventions, etc.

This paper was prepared as a background paper for the UNHCR *People Forced to Flee: History, Change and Challenge* publication and aims to demonstrate the extent to which impact evaluations are applicable to study programs in contexts of

[1] UNCHR defines a protracted refugee situation as one in which 25,000 or more refugees of the same nationality have been in exile for at least five consecutive years in a specified host country.

[2] For an overview of randomized evaluations, we recommend the following J-PAL resource: https://www.povertyactionlab.org/resource/introduction-randomized-evaluations.

forced displacement and how their application in these contexts differ. The paper is structured as follows: section 1 introduces impact evaluations; section 2 provides examples of the evidence generated through impact evaluations in `stable' LMICs and how it has been used in policy making; section 3 discusses the specific issues related to the application of impact evaluations to evaluate humanitarian and forced displacement programs; section 4 illustrates examples of emerging impact evaluation work in contexts of forced displacement; and section 5 provides recommendations for the future expansion of impact evaluations to inform forced displacement policies.

The content of this paper is derived from various sources: first, the author's first-hand experience working with Innovations for Poverty Action and the World Bank on the implementation of impact evaluations; second, consultation of publicly accessible impact evaluation resources curated by the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA), the International Initiative for Impact Evaluation (3ie), the World Bank Development Impact Evaluation unit (DIME) and the Center for Global Development (CGD); third, the collection of insights generated by impact evaluations of forced displacement programs now in the field under the FCDO-UNHCR-WB research program *Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership*[3]; fourth, an evaluation of the impact evaluation literature; and fifth, consultations with impact evaluation researchers and practitioners in the UNHCR, the IRC, the WFP, the London School of Hygiene and Tropical Medicine (LSHTM), and Women for Women International (WfWI)[4].

1. A brief introduction to impact evaluation

Evaluation and impact evaluation

Evaluations can include entire country programs, sectoral programs, organizational strategies, etc.[5] It is important to note that impact evaluations have a narrower and specific focus on interventions and/or projects, which may be implemented as components of a broader program, or in the context of a specific policy. For example, when developing a new policy that grants refugees the right to work, a government can engage a service provider to provide job seeking support services to refugees and an impact evaluation can be designed to determine whether the service provider's intervention is attaining the planned results (refugees are able to secure and retain jobs).

It is also important to understand the distinction between impact evaluations and process evaluations. Impact evaluations are considered 'complimentary' to process evaluations. While a process evaluation is a type of formative evaluation whose objective is to verify whether a program is being implemented as planned, an impact evaluation is a type of 'summative' evaluation whose objective is to assess the extent to which a program is achieving its intermediate (i.e.: changes in participants' knowledge, attitudes, behaviors) and final outcomes (i.e.: desired human capital, environmental, health outcomes).

Finally, within the realm of 'summative' evaluations, impact evaluations are quantitative exercises and employ rigorous methods that allow to measure the impact of an intervention on a sample population by comparing it to a comparison population which is similar but did not receive the intervention. Box 1 illustrates this concept with an example based on a project's theory of change[6].

[3] The program was established in 2016 as a partnership between UK Aid, funded by the Foreign, Commonwealth and Development Office (FCDO), the World Bank, and the UNHCR.

[4] I wish to thank Ninette Kelley (UNHCR), Lori Bell (UNHCR), Karoline Elisabeth Eberle (UNHCR), Caroline Sergeant (World Bank), Paolo Verme (World Bank), Aprille Knox (J-PAL), Nessa Kenny (IPA) and Erin Kelley (World Bank) for their helpful comments and suggestions. I wish to also thank Kathryn Falb (IRC), Eva Noble (WfWI), Jonas Heirman (WFP), Felipe Dunsch (WFP) and Camilla Fabbri (LSHTM) for their inputs to this paper during consultations.

[5] The reader can refer to the following note for a concise introduction to the discipline of evaluation: https://meera.snre.umich.edu/evaluation-what-it-and-why-do-it

[6] A theory of change is the illustration of how a project uses its inputs to produce specific outputs that are in turn expected to cause specific outcomes that will contribute to achieving an ultimate objective (or impact).

Box 1 – Theory of change, process evaluation and impact evaluation

Inputs	Outputs	Intermediate outcomes	Final outcomes
Time, money, and resources (personnel)	Products resulting from the combination of inputs. Outputs need to abide by some quality standards.	This is how the behavior, relationships, activities, actions of an individual, or group are expected to change.	This is the broader goal that we wish to achieve with the project.
Example: materials, lead engineer, construction workers	Example: borehole (functioning, well maintained).	Example: households fetch clean water from the borehole.	Example: improved child health.
Process evaluation	**Process evaluation**	**Impact evaluation**	**Impact evaluation**

Causal impact and comparison group

An impact evaluation can estimate the **causal impact** of a project/intervention by comparing outcomes between a population which participated in an intervention and a comparable/similar population that didn't (the comparison group).

Causal impact is the share of change (if any) for a specific outcome/s, i.e.: income, that is attributable solely to the participation in the project/intervention. Distilling the effect of a project on specific outcomes from other factors that could concomitantly be affecting the same outcomes, making it difficult to understand what causes what, is achieved by comparing outcomes for the group that participated in the project to outcomes for a comparison group which is similar to the group that participated, but that did not participate. Impact evaluations rely on the identification of a comparison group that is as similar as possible to the group that was part of an intervention.

Impact evaluation methods

There are different impact evaluation methods, each distinctly based on how the comparison group is identified and assumptions made regarding the comparison group. Because the choice of the impact evaluation method is important to the success of the evaluation, researchers and practitioners work hand in hand to carefully assess pros and cons and to identify the most suitable method[7]. One of the simplest evaluation methods relies on the comparison of data at the baseline (before the program is implemented) and at the endline (after the program is implemented). This evaluation method is called 'pre-post' and has a disadvantage in that it can't disentangle the effect of the program from other factors other than the program that could be changing between baseline and endline. Appendix 2 illustrates the drawbacks of this evaluation method with an example.

In the ***experimental method***[8] (or randomized evaluation or randomized controlled trial), the comparison group is identified by random assignment, a procedure equivalent to the one adopted to test a drug or a vaccine in the medical field. Random assignment is desirable because it allows overcoming the issue of **selection bias**, which is the concern that some unobservable characteristics may in fact be driving the observed change in outcomes – fully or to some extent – rather than the intervention being evaluated. Appendix 1 illustrates the problem of selection bias with an example. In the *experimental method*,

[7] For a succinct summary of impact evaluation methods and underlying assumptions for each method, we recommend the following note: https://www.povertyactionlab.org/sites/default/files/research-resources/impact-evaluation-methods-table.pdf

[8] The following are recommended non-technical handbooks for those interested in learning more about randomized evaluation: *Running Randomized Evaluations: A Practical Guide* (Glennerster and Kudzai, 2013) and *Using Randomization in Development Economics Research: A Toolkit* (Glennerster *et al.*, 2006).

practitioners and researchers work together to identify the two groups – the one that receives the intervention and the comparison one - by random assignment. For example, for a training intervention that has more applicants than training slots available, one can randomly pick participants that will receive the training first from the list of eligible applicants and collect data both on trained individuals and those that will be invited for the training at a later stage. In this case, the researcher would use the list of applicants eligible for the training and "flip a coin" using a statistical software to select those that will be invited for the training first. Random assignment is in practice not always a viable method. For example, to evaluate the impact of a company selling solar-powered products, randomizing who the company sells its products to among people who expressed interest in purchasing these products could likely cause the company reputational harm. Appendix 4 discusses in greater detail what is required from implementing organizations when engaging in an impact evaluation, and especially a randomized evaluation.

Non-experimental methods can be applied in those cases where random assignment isn't possible or when similar populations already exist 'naturally' that received or didn't receive an intervention. For example, Esper *et al.* (forthcoming) exploited a sudden change to the cash reintegration assistance program for Afghan returnees from Pakistan: those that returned before a certain date received a higher cash assistance amount than those that returned after (more details on this forthcoming study can be found in Appendix 3 here). Instead, Quattrochi *et al.* (2020) exploited one of the program household poverty targeting features to evaluate the impact of a voucher program by using households that didn't make it into the program because they were falling slightly below the cut-off score assigned by the implementing NGO in the targeting procedure as the comparison group. Finally, where other methods aren't applicable, the comparison group can be constructed using statistical techniques: for example, Sonne *et al*. (forthcoming) evaluate the impact of a community driven development project in refugee-hosting areas of Ethiopia and construct a comparison group by identifying similar households in refugee-hosting areas that were not selected for the program (more details on this ongoing study can be found in Appendix 3 here).

2. Seminal impact evaluations

J-PAL, the global research organization specialized in impact evaluations, estimates that more than 400 million people have been reached by programs that were tested with randomized evaluations. This estimate considers only studies authored by J-PAL researchers and doesn't consider impact evaluations that yielded poor results and led to preventing further investments on ineffective programs – which are as important for policy as studies that found impact. As an example, Banerjee *et al.* (2008) found that a program to reduce health workers' absenteeism had only marginal impact and these results were taken on board by the government that cancelled the planned scale-up. As a result of thousands of impact evaluations conducted in 'stable' LMICs to date, there is now a wealth of evidence on program effectiveness available in academic articles, policy papers, blogs, and literature reviews[9].

This section illustrates some examples of evidence generated by impact evaluations and how it has informed policy in selected sectors (education, jobs, social protection, and health).

Education

Impact evaluation methods were first applied to study programs in LMICs aimed at improving educational outcomes such as enrollment, attendance, and learning. Randomized experiments conducted in the mid-1990s in Western Kenya to identify the most cost-effective strategies to improve attendance and learning are well-known examples. Two of the experiments focused on testing supply-side interventions, namely the provision of flipcharts (Glewwe *et al*, 2004) and textbooks (Glewwe *et al*, 2009); other experiments tested a school-based deworming program (Miguel and Kremer, 2004) and the provision of school meals (Vermeersch and Kremer, 2004); another field experiment studied the effect of providing financial incentives to teachers based on children's test scores (Glewwe *et al*, 2010). These studies found that: 1) the provision of textbooks improved test scores only for the most able students; 2) the use of flipcharts had no effects on student learning; 3) school-

[9] Examples of reviews of the available evidence categorized by theme can be found on the website of the International Initiative for Impact Evaluation (3ie): https://www.3ieimpact.org/evidence-hub/publications/briefs/evidence-use.

based deworming[10] and the provision of meals were effective in reducing school absenteeism but did not improve test scores in the short run; and 4) financial incentives tied to students' test scores had the perverse effect of making teachers 'teach to the test'. Inexpensive deworming treatments (less than $0.50 per child per year) were scaled-up of to millions of children in Kenya, India, Ethiopia, Vietnam, Nigeria, and Pakistan[11].

Another notable randomized experiment assessed the effectiveness of "Progresa", a conditional cash transfer (CCT) program in rural Mexico that targeted poor mothers and provided them with a cash transfer, conditional on children's school enrollment and attendance (Schultz, 2004). The randomized evaluation found positive program effects on enrollment which were sustained up to 18 months after the cash was transferred. Given the positive findings, the program was scaled up to the national level and adopted by other Latin American countries.

Finally, other evaluations focused on interventions to improve the quality of teaching. In India, Banerjee *et al.* (2007) partnered with the NGO Pratham to evaluate remedial education interventions and found very positive results for a simple, easily scalable intervention (2-week long training) that trained young high-school graduates on basic core competencies and employed them to be remedial education teachers ("Balzakhi"). The program is now widespread in India and has been scaled through the Teaching at the Right Level (TaRL)[12] initiative to other countries, including Ghana[13].

Jobs and microenterprise

Technical and vocational education and training programs that are employed worldwide to improve employment outcomes – especially for the youth – have been largely evaluated. Results on their effectiveness in LMICs are positive for employment, formal employment, and earnings, although small (Chinen *et al*, 2018). Promising evidence is emerging from more recent evaluations focusing instead on training on soft skills such as negotiation, teamwork, and leadership[14].

Microcredit can help small businesses grow by giving them access to credit, one of the main barriers to growth of small businesses[15]. Although in the beginning these programs seemed revolutionary, based on the Grameen Bank model, their popularity was challenged by the results of randomized evaluations done in multiple countries, all yielding similar and non-reassuring results (Banerjee *et al*, 2015).

Social protection

Cash transfer programs –conditional and unconditional– are proliferating in LMICs and have also been largely evaluated. Social protection raises many questions at the operation level. For example, is giving cash alone (an easily scalable model) a more cost-effective model to providing cash paired with other types of livelihood support? While the debate continues on whether programs should focus on simply providing cash, there is a consensus that graduation programs (cash paired with other support services such as business training, mentorship, etc.) targeting the ultra-poor seem to work well. These programs are based on the BRAC ultra-poor graduation model, which is a comprehensive poverty-alleviation program which includes targeting the poorest households, providing them with cash (or a productive asset) and business training, and other types of support (savings schemes, etc.). RCTs conducted in six countries found significant program effects across the ten outcome measures sustained 24 months after the asset transfer and significant effects for 8 out of 10 outcomes 12 months 36 months after the asset transfer (Banerjee et al, 2015a)[16]. These programs are now being widely scaled.

[10] The following WHO note explains the importance of deworming: https://www.who.int/elena/titles/bbc/deworming/en/
[11] https://www.evidenceaction.org/dewormtheworld-2/
[12] https://www.teachingattherightlevel.org/
[13] https://www.poverty-action.org/impact/scaling-targeted-instruction-ghana
[14] 3ie Youth and Transferable Skills Evidence Gap Map https://gapmaps.3ieimpact.org/evidence-maps/youth-transferable-skills-evidence-gap-map
[15] For a comprehensive overview of barriers to growth of small businesses in LMICs, refer to the IGC-ANDE-Argidius working paper *Addressing Constraints to Small and Growing Businesses* available at this link: https://www.theigc.org/wp-content/uploads/2018/11/IGC_ANDE-review-paper_final-revised.pdf
[16] A summary of this research can be found here: https://www.poverty-action.org/impact/ultra-poor-graduation-model

Health

Despite considerable progress made globally since the 90s, in 2019, an estimated 5.2 million children under 5 years died because of preventable and treatable illnesses such as pneumonia, diarrhea, and malaria[17]. Even in low-income settings, many of these premature deaths could have been averted with simple and affordable preventative health care such as immunization, distribution and correct use of insecticide-treated bed nets, adequate nutrition, and access to safe drinking water. Impact evaluations have focused on understanding the best strategies to ensure beneficial health behaviors are adopted. For example, impact evaluations helped build the evidence base for the importance of free distribution of essential health care products —such as insecticide–treated bed nets and deworming treatment— through randomized evaluations that measured take-up and utilization of these products at different prices (Cohen and Dupas, 2010). Impact evaluations have also been used in India to test alternative strategies to improve child immunization and identify what strategy is most cost-effective (Banerjee *et al*., 2020). Finally, impact evaluations have also been helpful to evaluate the success of different strategies to prevent HIV infections. For example, an impact evaluation of a new informational campaign —targeting teenagers in Kenya and explaining the increased risk of HIV transmission when having relationships with older partners— found that this type of informational intervention led to a 28% decrease in teenage pregnancy, suggesting teenagers were using contraceptive methods more frequently (Dupas, 2011), as opposed to the traditional abstinence-only informational campaign which was shown to be ineffective. However, these findings did not replicate in an experiment in Botswana, and this led the implementing NGO to explore other approaches[18].

3. Considerations on the application of impact evaluations to forced displacement projects

Programs in humanitarian contexts are being evaluated although rigorous impact evaluation methods aren't common. A review of humanitarian program evaluations (Puri *et al.*, 2015) identified only 38 impact evaluations from more than 900 studies. Similarly, in a review of cash and "near-cash" transfer programs in emergency settings, Tappis and Doocy (2018) identified only 5 impact evaluations from 108 studies[19]. More recently, IPA documented fewer than 25 published randomized evaluations of forced displacement programs, the bulk of which were evaluations of psychosocial support programs[20]. However, this landscape is changing, and there is increased interest in the application of rigorous evaluation to measure the impact of humanitarian and forced displacement programs alike, especially given the protractedness, size, mounting financial needs of these crises and the realization that reaching solutions quickly for many of these crises does not seem attainable. Furthermore, the emergence of innovative financing instruments that tie disbursements to *demonstrated* achieved results[21] and which are already being piloted by humanitarian organizations (for example the ICRC[22]) will also contribute to increasing willingness for rigorous measurement.

[17] https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality

[18] This article summarizes how the NGO Young 1ove utilized the findings from the impact evaluation in Botswana: https://www.globalinnovation.fund/does-sugar-daddies-replicate-the-preliminary-results-are-in-for-botswana/.

[19] We reference available reviews of evaluations of 'humanitarian programs' in absence of a review focusing on forced displacement programs. Humanitarian crises not always imply forced displacement and vice versa.

[20] The list of evaluations is annexed to this blog: https://www.poverty-action.org/blog/filling-forced-displacement-evidence-gap-taking-stock-world-refugee-day.

[21] These innovative financial instruments are social bonds, development bonds and result-based financing instruments. A social bond is a financial instrument in which an investor funds a project, and a government serves as outcome payer (pays back the investor only when results are achieved). In a development bond, an investor anticipates the funds needed to implement the project, and a philanthropic organization pays back the investor based on achieved outcomes. In a result-based financing model instead, the funder pays an amount that varies based on the magnitude of the achieved outcomes. For all these new financing instruments, payments are tied to achieve results that can be rigorously measured.

[22] In 2017, the ICRC launched the first impact bond in the humanitarian sector. The outcome payer will pay for demonstrated achieved results on the efficiency of treatment at rehabilitation centers in conflict-affected areas of Mali, Nigeria, and the DRC. The project is summarized in this article: https://www.devex.com/news/icrc-launches-world-s-first-humanitarian-impact-bond-90981.

To respond to the increased demand for rigorous evaluation, organizations specializing in impact evaluation are already offering impact evaluation trainings tailored to humanitarian practitioners (see for example J-PAL's training *Generating Rigorous Evidence on the Effectiveness of Humanitarian Programming*[23] and DIME's training for WFP teams on evaluating cash-based interventions and specifically their impact on gender outcomes[24]). It is expected that these trainings will contribute to further nurturing the desire for rigorous evaluation in the humanitarian and forced displacement sectors. Furthermore, there are examples of pioneer randomized evaluations that can deliver important lessons learnt for future evaluations: for example, Quattrochi *et al.* (2020) share lessons learnt from the evaluation of a humanitarian program in Eastern Democratic Republic of the Congo.

The following paragraphs discuss what result to expect from the application of impact evaluations in humanitarian and forced displacement settings: what will be different, what additional challenges there may be, and what opportunities could arise.

Research questions

Impact evaluations in forced displacement settings will be developed to provide answers to **specific research questions** which may differ from research questions relevant to a 'stable' LMIC setting. For example, while in a LMIC setting one may wonder whether a given program has any effect at all –justifying a traditional randomized controlled trial design to include a 'pure' comparison group that doesn't receive an intervention for the duration of the study (which could last several years)– in a humanitarian and/or forced displacement situation the critical questions may not be whether a program works or not, but rather **which program is most effective**, or **what is the most effective way to deliver it**. For example, Aker *et al.* (2016) evaluate traditional vs. digital delivery of aid in cash to people displaced by drought in Niger and find that digital delivery (via mobile money) was more effective and allowed people to purchase more diverse types of foods and have better dietary diversity than people who received cash. Impact evaluations are also suitable to investigate **new research questions** specific to forced displacement such as program's effects on social cohesion. In a forthcoming study, Baseler *et al.* study the role that a refugee-led intervention to support local small business owners in Uganda can have on strengthening the support for Uganda's inclusive refugee policy (more details on this study here).

Methodological considerations

Shocks that may characterize humanitarian and forced displacement contexts, such as sudden program changes, can create suitable situations to conduct an impact evaluation employing a quasi-experimental method. For example, in the evaluation of the UNHCR cash reintegration assistance for Afghan returnees from Pakistan, researchers leverage the sudden change in the cash amount given to returnees upon return (more details of this evaluation are included here). Specific program targeting features can also be leveraged to design a rigorous evaluation. For example, Lehmann and Masterson (2014) evaluate the impact of the UNHCR winter cash assistance program for Syrian refugees in Lebanon leveraging a key program feature to identify treatment and control households: the program was offered to people living at or about a 500m altitude. This meant that households living at 499m would not receive the program and were used as comparison group, while households living at 500m would receive it. Randomized controlled trials can be more feasible than expected in such difficult contexts. In fact, organizations work with specific budgets and timelines: it is common to have more people in need than there are resources to implement a program for all and implementing immediately to all eligible participants may often be logistically unfeasible. Hence, randomization may in some instances even provide a 'fairer' path for program implementation.

Innovative and nimble research designs

One critical challenge to implementing impact evaluation of projects that can be delivered in emergency situations is that evaluation activities (such as data collection, or randomization where applicable) will need to also be deployed at the same speed as program activities. This is a challenge that organizations working in emergencies, such as the World Food Program, are already grappling to solve: the WFP for example can leverage a research budget that can be easily deployed in a country or another depending on where the emergency occurs. Similarly, the WFP is working on the development of 'adaptive'

[23] https://www.povertyactionlab.org/event/training-randomised-impact-evaluations-humanitarian-interventions
[24] https://www.worldbank.org/en/events/2019/05/27/cash-based-transfers-and-gender#2

research designs that can be suitable to deploy during emergencies. In contexts where new displacement waves can be expected, organizations that already have an established presence on the ground can make evaluation plans in anticipation of new displacement waves and roll-out the research as it occurs. This is the strategy used by Aker C. (2017): the study compares two different modalities to deliver emergency relief (cash and vouchers) and finds that there were no differences in food consumption and other measures of well-being across modalities, suggesting that cash was a better strategy given it is a cheaper program to implement.

Challenges and opportunities for innovation

Attrition. Attrition, which is the inability to survey individuals that are in the study sample at baseline in follow-up surveys, is problematic for an impact evaluation when the sample that cannot be surveyed has different characteristics from the sample that was tracked ('systematic attrition'). Given the high levels of population mobility that characterize forced displacement contexts, it is challenging when the impact evaluation design relies on following up the same individuals over time. For example, in an impact evaluation of a business training program implemented in a displacement context, participation in the program could have improved business revenues for a subgroup of successful business owners to the point that they were able to earn an income sufficient for them to move to the city or return to their countries of origin, which could make it impossible to reach them again. The inability to survey these individuals during follow-up surveys could lead to underestimating the impact of the program because the population who stayed behind is likely on average to be less successful than the population that left, whose information is no longer available. This is a challenge that can be pronounced depending on population mobility in a specific context. Carefully designed tracking protocols that allow keeping in touch with study participants even after they have moved can mitigate this challenge, as well as leveraging innovations in tracking methods[25]. Finally, relying on cross-sectional data (different samples at baseline and follow-up) could also be a viable alternative. For instance, in the case of an intervention delivered at a school level, where it is difficult to keep track of the same children over time, but still maintain the school-level randomization of treatment. A forthcoming IPA study is a good example of how researchers were able to innovate to adjust to the difficult context of working within a refugee camp in Tanzania[26].

Contamination. When a humanitarian emergency occurs, different humanitarian organizations can intervene with similar or different programs (shelter, WASH, food aid, cash aid, etc.). An environment in which many programs are implemented at the same time isn't conducive to implementing an impact evaluation because it can be hard to control 'who receives what'. The comparison group identified for the evaluation can be 'targeted' to receive programs offered by other organizations and/or both the intervention and comparison groups may be receiving so many other programs that the marginal effect of the program being evaluated can be smaller and harder to detect. Contamination can be minimized by coordinating interventions by different implementing organizations and by implementing follow-up surveys as soon as possible (to reduce the likelihood of other programs coming in). Contamination may also not interfere with the evaluation if it is equally spread across study populations (those that participate in the program and those that don't).

Shorter, more frequent, and more focused data collections. The need to collect data rapidly as humanitarian and forced displacement situations unfold and humanitarian support programs are rolled out will drive the need to develop short and focused surveys that identify key indicators of interest to be monitored over time. For example, the WFP conducts routine data collections for their Vulnerability Analysis Mapping (VAM) and these datasets can be leveraged for impact evaluation. Furthermore, it's probable that impact evaluations in emergency contexts will rely on shorter time spans between baseline and follow-up surveys. More frequent surveys are in fact useful to monitor changes in situations that are rapidly evolving. The frequent collection of data also allows for more frequent analysis and the ability to feed back information about program performance to field teams which can use the data to correct implementation as needed.

[25] The IPA Peace and Recovery Program, with funding from UK-Aid, is piloting for example tracking protocols for displaced people in Lebanon and Colombia that rely on WhatsApp https://immigrationlab.org/2020/10/01/webinar-whatsapp-surveys/.
[26] This is the strategy utilized in an ongoing study evaluating a violence prevention intervention in schools in a refugee camp in Tanzania https://www.poverty-action.org/study/preventing-violence-against-children-refugee-camp-schools-tanzania.

Quality data on forced displacement. Until recently, the scarcity of studies focusing on forced displacement has been in part dictated by a scarcity of data on these populations. However, this landscape is changing and there is now more and better-quality data available or being collected on forced displacement that can be useful for impact evaluation: for example, the establishment of the UNHCR-WB Joint Data Center will ensure that going forward displaced populations are included in national household surveys. Some cross-sectional surveys have already been collected that cover both displaced and host populations, such as the Uganda Refugee and Host Communities 2018 Household Survey, which covers both host and refugee households, and the Profile of Internally Displaced Persons in North-East Nigeria 2018 which is representative of IDPs and host households. Finally there is the collection of representative panel surveys[27], for example the Cox's Bazar Panel Survey in Bangladesh which is representative of Rohingya refugees and host households.

Emerging innovative data can also be leveraged for impact evaluation in forced displacement: for example, The Armed Conflict Location & Event Data Project (ACLED) is a platform that collects georeferenced data on conflict events by crowdsourcing information from the local news; satellite nighttime lights data, which is used to approximate economic activity; and rainfall data to identify areas affected by drought and/or flooding.

Cost-effectiveness analysis. A shortcoming of many impact evaluation papers is that they are silent on intervention cost and cost structure–information that is critical for policy. Obviously, the cheaper the intervention, the easier to scale it up if successful. However, the cost structure is also important, especially for evaluations of interventions implemented by entities other than the government, such as NGOs, but relying on the government structures to scale. Among impact evaluation researchers and practitioners alike the debate on how to strengthen this important aspect is open[28]. J-PAL and IPA have made steps towards addressing this shortcoming by making cost-effectiveness analysis a requirement for all evaluation projects funded[29]. It is expected that the application of impact evaluations to study humanitarian and forced displacement programs will contribute to pushing the cost-effectiveness analysis agenda forward.

Ethical considerations

Conducting research and impact evaluations in displacement situations can involve risks both to the research teams (enumerators) and to study participants. For this reason, ethical considerations are especially important in these contexts and of interest throughout all stages of research, from the choice of the most suitable impact evaluation method, to the evaluation design, implementation, and data analysis. For example, in the design phase, it is necessary to assess whether it is ethical to have a 'pure' comparison group–a group of people who participate in the research but do not receive the intervention being evaluated for the duration of the study (until all follow-up data has been collected). This design would be unfeasible for an emergency relief program providing food and health care because the program would need to be rolled out soonest to all people in need.

Ethical considerations are important also for questionnaire design: for example, it is important to decide whether it is critical for the study to ask sensitive questions, such as questions that may evoke painful memories and possibly reactivate trauma (asking somebody if they've been raped for instance). If it is determined that this is an important question to ask for the success of the study, the questionnaire and the questionnaire administration manual can be designed to minimize the risk of provoking distress to participants. Furthermore, enumerators need to be trained to recognize signs of distress and what to do if a respondent manifests these signs.

Conducting randomized controlled trial (RCT) evaluations – impact evaluations with a 'pure' comparison group that does not receive the intervention during the study – is feasible in humanitarian contexts. First, there may be interventions that could

[27] A panel survey is a survey in which the same individuals are tracked over time.
[28] To get more insights on the ongoing debate on cost-effectiveness and impact evaluation, we suggest watching the recoding of this virtual event organized by 3ie: https://www.youtube.com/watch?v=RbnPPIExn_o&feature=youtu.be
[29] https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea

do more harm than good, such as those that may involve reevoking traumatic experiences: it is important to evaluate and recognize anything unethical before proceeding with an intervention will not work as intended. Second, agencies face budget and operational constraints that make it unfeasible to reach everybody or everybody at once, necessitating a gradual phase-in of the intervention. In such circumstances, randomizing the intervention may in fact constitute a 'fairer' way to proceed with rollout. A study in Eastern DRC is an example of how researchers have dealt with ethical considerations, where the program evaluated was targeting the most vulnerable households. To evaluate the program, the research budget was used to reach more households than would have been reached by the program (those that weren't targeted) and the comparison group (Quattrochi *et al.*, 2019).

Given the increased interest in applying impact evaluations to evaluate programs in humanitarian and forced displacement contexts, there is a momentum to be seized to develop ethical principles tailored to the review of studies in humanitarian contexts (Falb *et al.*, 2019). The IRC conducted more than 75 research studies between 2007 and 2017, 40 of which were impact evaluations in 28 crisis-affected countries. Based on the experience providing ethical review to their studies via their internal Ethics Review Board, they make the following recommendations to advance in the development of standard ethics reviews for studies in humanitarian contexts: introducing a staged review of research protocols and flexible reviews of modification requests to account for the need to adjust research protocols quickly to adjust to changing circumstances on the ground or to deal with uncertain information on the sample that can be confirmed later on; introducing principles to address the traumatic experiences of participants; addressing the difficulties in attaining meaningful informed consent among populations that may be dependent on the interventions provided; and ensuring reviewers are knowledgeable of the population involved in the research.

4. Impact evaluations in contexts of forced displacement

Currently there is an increase in the number of impact evaluations in contexts of forced displacement. New studies are emerging from various initiatives[30]: the IPA and J-PAL Governance, Crime and Conflict Initiative, funded by UK Aid, which is expected to produce evidence on the effectiveness of policies to promote peace and good governance, reduce crime, and support individuals and communities recovering from conflict and humanitarian emergencies[31]; J-PAL's newly established Regional Research Center in Cairo, which is expected to boost the capacity to conduct rigorous research in several of the North-Africa and Middle Eastern countries affected by forced displacement[32]; IRC's Airbell Research Center which has been active for quite a few years; WFP's impact evaluation partnership with DIME[33]; the UK Aid – UNHCR - WB *Building the Evidence on Forced Displacement* research program; the *PlayMatters* program, funded by the LEGO Foundation, which is expected to produce research on play-based interventions for children; and IPA and J-PAL research hubs and data infrastructure to support research in locations critical for humanitarian response (for example Bangladesh and Tanzania).

As presented in section 2 with evaluations in 'stable' LMIC contexts, the following will discuss trends in impact evaluations in displacement contexts in the same four selected sectors: education, jobs and microenterprise, social protection, and health.

Education

The education sector in displacement contexts faces new challenges that require tailored solutions. First, education infrastructure both in contexts where displaced children stay in camps and contexts where they may be able to access schools in hosting areas may be strained: there may be a shortage of classrooms, teachers, learning materials and water and sanitation facilities that prevent refugee children from accessing education (UNHCR, 2019). Children who are able to access schools in hosting areas may lack the documentation and residency requirements necessary for school enrollment. Once in

[30] This list is not exhaustive.
[31] The launch of this initiative was covered in this article: https://www.devex.com/news/dfid-backed-rcts-take-on-new-frontiers-in-fragile-countries-92952
[32] https://www.povertyactionlab.org/updates/release-j-pal-launches-new-regional-research-center-cairo-auc-help-improve-lives-across
[33] https://www.wfp.org/publications/wfp-impact-evaluation-strategy-2019-2026

school, they can experience discrimination and social exclusion, and their ability to learn may be compromised by specific factors like the trauma they have experienced with displacement (UNICEF, 2019).

Jordan, Lebanon, and Turkey are countries facing huge challenges in the education sector following the massive influx of Syrian refugee children into schools. A completed J-PAL study in Turkey (Alan *et al.*, 2021) found positive results from an educational program that built social cohesion in ethnically mixed schools by developing perspective-taking abilities in children. Similarly, the Lebanese and Jordanian governments are launching new programs with the goal to preserve the quality of education in government schools, while ensuring inclusiveness. Some of these projects are being evaluated with financial support from the FCDO-UNHCR-WB *Building the Evidence on Forced Displacement* research program: in Lebanon, an impact evaluation will test the effectiveness of a new training curriculum for teachers' coaches, while in Jordan, both a pedagogical and psychosocial interventions will be tested.

Research is scant on interventions targeting children, yet children living in situations of displacement are estimated to be a huge share of the total number of displaced (42% in 2020 according to the UNHCR). For example, Child-Friendly Spaces (CFS) in camp settings are a popular program funded by the UNHCR, but evidence of their impact is lacking. Some initial evidence suggests effects may not be sustained over time: a randomized controlled trial evaluation (RCT) of child friendly spaces in a refugee settlement in Uganda found some impact on psychosocial wellbeing and developmental assets especially for girls in the short term (while CFS were still active), but no effects in the longer term on any outcome (Metzler *et al.*,2019). These findings call for further research on the impact of these types of interventions. The *PlayMatters* initiative[34] is expected to generate more evidence in this space: the program (with implementation done by a consortium led by IRC) and related research (research institutions involved include IPA and the Behavioral Insights Team) will promote play-based, early learning solutions for pre-primary and primary school-aged children impacted by the humanitarian crises in East Africa.

Jobs

As with the education sector, displacement brings unique issues to the labor market for both the displaced and host communities. Among the challenges that the displaced face in obtaining and retaining a job on the host labor market are whether they have the right to work in the first place, physical and mental health issues they may be experiencing, lack of social networks, and separation from family members (Caron and Schuettler, 2020). Moreover, displaced job seekers are highly likely to find themselves in saturated job markets, face discrimination, lack the skills required to work in the host labor market (such as language) or face downgrading (as documented for Venezuelan refugees and migrants living in Brazil in Shamsuddin *et al.*, 2021). As per the impact of the arrival of displaced people on labor markets of host LMICs, the literature suggests that refugees compete with hosts mostly in the short run for low-skill informal jobs, especially jobs done by women (Verme and Schuettler, 2021). Specific policies and programs are being designed that consider the challenges faced by the forcibly displaced and some evaluation work is ongoing. For example, Ethiopia adopted the Comprehensive Refugee Response Framework (CRRF) and pledged to grant rights to work to up to 30,000 refugees. In this context, the World Bank, in partnership with external researchers, the UNHCR and the government of Ethiopia are designing an impact evaluation to test a 'positive imagery' mental health intervention that could be easily scalable and help refugees overcome mental health barriers that prevent them from tapping into the full potential of being able to access formal jobs (more details on this forthcoming evaluation can be found here). In Bangladesh, researchers at the World Bank and other institutions are engaging to design a rigorous analytical program to inform the humanitarian response to the Rohingya refugee crisis and help inform the policy dialogue with the government, which has made it illegal for refugees to work and very difficult to access informal jobs given movement restrictions that keep them confined to the camps. In a forthcoming randomized evaluation, Hussam *et al.* study the psychosocial effects of working as compared to merely receiving the equivalent in cash for Rohingya refugees. The findings do show that the cash-only arm doesn't improve psychosocial wellbeing, while employment brings significant improvements in psychosocial wellbeing. These findings are important to inform the policy dialogue on refugees in Bangladesh.

[34] To read more on *PlayMatters*: https://www.legofoundation.com/en/about-us/news/the-lego-foundation-awards-us-100-million-grant-to-the-international-rescue-committee-to-bring-learning-through-play-to-children-impacted-by-crises-in-ethiopia-and-uganda/

Social Protection

Social protection programs are increasingly important, especially following the COVID-19 pandemic, and are extremely important in helping displaced populations. However, many displaced people struggle to access these programs, as found in Phadera *et al.* 2020 who studied the impact of accessing the Public Distribution System program in Iraq for displaced households (more details on this study can be found here). As discussed in section 2 graduation programs were rigorously tested in many different country contexts and are now being evaluated in displacement contexts in Afghanistan, Uganda[35], and Mozambique. There are good reasons to believe that this type of program should work given their effectiveness with ultra-poor households in LMICs. But given the unique challenges faced by forcibly displaced people, impact evaluation research can reassure on their effectiveness and optimal design (*Should these programs include psychosocial support when they involve displaced people, and if so, how should this be delivered? Besides the likelihood of trauma, are there other hindering factors that may make these programs not as effective for displaced households?*). Evidence from the Afghanistan experiment (Bedoya *et al.*, 2019) is encouraging, with strong positive impacts of the program that measured 12 months after implementation ended and 24 months after the asset transfer.

Other ongoing impact evaluation efforts focus on cash transfer programs or programs accompanying cash transfers and are aimed at enhancing their impact. In Lebanon, the UNHCR and WFP evaluated the short- and long-term impact of a multi-purpose cash assistance program. The impact evaluation determined that impact on most dimensions materialize in the long term and that the effects of receiving assistance fade away quickly for households that received cash for a shorter time frame, suggesting that longer cash assistance cycles are more effective and that likely a cash 'plus' approach could lead to more sustainable outcomes and should be explored (Chaaban *et al.*, 2020). As per evaluations of programs accompanying cash transfers, two such evaluations are ongoing: one in Niger, where the treatment group receives cash and entrepreneurship training and one in Cameroon where the treatment group receives cash and participates in an intervention to prevent intimate partner violence (IPV).

Health

While the forced displacement impact evaluation literature available already focusses on mental health interventions, there still is a need to generate more evidence to allow to derive more general and applicable findings. Tol *et al.* (2011) note that "*overall, research and evidence focus on interventions that are infrequently implemented, whereas the most commonly used interventions have had little rigorous scrutiny*". The emphasis of future impact evaluations should be on interventions that can be delivered as efficiently as possible (by non-specialist facilitators and to groups rather than individually). An example of this is a randomized controlled trial that tested the WHO multimedia guided self-help intervention with South-Sudanese women refugees in Northern Uganda (Tol *et al.*, 2020). In Bangladesh, the World Bank Development Impact Evaluation unit is developing a randomized experiment to test the effectiveness of a photography training for adolescent girls that aims to improve their psycho-emotional wellbeing (more details on this evaluation can be found here). Other impact evaluation work is ongoing in Ethiopia and Kenya to understand the effectiveness of 'community-driven-development' programs (more details here). The available evidence on these programs in Eastern DRC suggests that they may not be effective: After IRC engaged in evaluations of their 'community-driven reconstruction (CDR)' program, they found negligible program effects that, together with the cost of implementing these programs, led the IRC to discontinue the program. (Laudati *et al.*, 2018a and Laudati *et al.*, 2018b)[36].

5. Conclusion and recommendations

[35] A summary of the ongoing evaluation in Uganda can be found here: https://www.poverty-action.org/study/impact-graduation-program-livelihoods-refugee-and-host-communities-uganda

[36] Results from the RE of the first phase of the program were null on all intermediate and final outcomes, while some positive results were found on some indicators from the phase 2 evaluation, though still insufficient to justify further investment in this type of program given its cost.

1. Millions are spent every year on programs to support the increasing number of people forced to flee from their homes due to conflict, natural disasters, and persecution (forcibly displaced people). For years humanitarian organizations have been struggling to reach all in need because of limited funding. Going forward, based on the principle of "shared responsibility", established with the 2018 Global Compact on Refugees, organizations working towards poverty eradication, such as the World Bank Group and governments of countries affected by displacement (either directly or indirectly by hosting refugees), will join forces with humanitarian organizations to ensure a coordinated and strengthened displacement response. To ensure funding is channeled to effective programs, rigorous evidence is needed. Impact evaluations provide this rigorous evidence base by measuring the causal impact between a program and the outcomes it intends to improve. Impact evaluations have been widely applied in the past three decades to study programs in 'stable' LMICs. There is an increased interest by donors and practitioners in the application of such evaluation methods in contexts of forced displacement, as evidenced by a growing number of dedicated research initiatives.

2. Although conducting impact evaluations in humanitarian settings is a difficult task given the volatile environments in FCV countries, some examples of their application exist and are delivering important findings and lessons learnt to inform future research in this area. For example, the IRC and DIME evaluated the impact on health and educational outcomes of a Community-Driven Reconstruction (CDR) program in the Eastern DRC and found negligible program effects. As a result of these findings, the IRC discontinued the program. Future research can build and expand on the lessons from the implementation of these pioneer studies.

3. Conducting impact evaluations in humanitarian and forced displacement contexts will not only contribute the much-needed evidence on program effectiveness, but also push researchers to innovate and deliver findings that are timely and operationally relevant. For example, cost-effectiveness analysis, which is often overlooked in impact evaluation will be a key component of such research, together with tailored and more comprehensive ethical standards. Impact evaluation results will also need to be completed and acted on more quickly. It is also anticipated that researchers will be 'pushed' to develop innovative methods that respond to the evidence needs, while operating within the constraints of humanitarian and forced displacement environments.

4. Currently, there is a scarcity of experienced researchers and economists working on forced displacement and humanitarian issues. Consequently, partnerships are key in launching successful impact evaluations, and key partnerships are already taking shape. For example, FCDO, the WB and the UNHCR have launched in 2016 the *Building the Evidence on Forced Displacement* research program and in 2019 the WB-UNHCR Joint Data Center was established; the WFP's Office of Evaluation and DIME have launched an impact evaluation partnership; the LEGO Foundation, IRC, IPA, the Behavioral Insights Team, and others will work under the *Play Matters* initiative on impact evaluations to specifically evaluate children play-based interventions. Partnerships between implementing organizations, academic institutions and research organizations specializing in impact evaluations will be crucial in producing evidence and gains that go beyond the impact evaluation results: practitioners engaging in impact evaluation will develop an evidence-based culture to implementation, while further strengthening organizations' internal M&E processes for the benefit of both the impact evaluation and the organization itself; researchers will be pushed to innovate to develop impact evaluation designs and protocols that apply to humanitarian and forced displacement contexts.

5. Data is core to an impact evaluation. Impact evaluations will benefit from new data collection initiatives in forced displacement contexts, such as the collection of cross-sectional and panel surveys representing both host and displaced populations[37]. Going forward, more and higher quality data on displaced populations and hosts will be available and will

[37] The World Bank and the UNHCR have launched several initiatives to collect data representative of both displaced and host communities. For example, in Uganda, the Uganda Refugee and Host Communities 2018 Household Survey is a survey representative of both host and refugee households. In Nigeria, the Profile of Internally Displaced Persons in North-East Nigeria 2018 is a representative survey of IDPs living in camps, host communities and host households. In Bangladesh, the WB and IPA have collected a survey representative of Rohingya refuges and host communities (see Cox's Bazar Panel Survey).

be useful to rigorous evaluation. Investing in such primary data collections will be important, as well exploring ways to collect and use other types of available data (satellite imagery, rainfall data, social media data, etc.).

References

Aker, J. C. (2017). Comparing Cash and Voucher Transfers in a Humanitarian Context: Evidence from the Democratic Republic of Congo. The World Bank Economic Review, 31, 44–70, available at: https://doi-org.libproxy-wb.imf.org/10.1093/wber/lhv055.

Aker, J. C., Boumnijel, R., McClelland, A. and Tierney, N. (2016). Payment Mechanisms and Anti-Poverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger. Economic Development and Cultural Change, 65 (1), 1–37.

Alan, S., Ceren Baysan, C., Gumren, M. and Kubilay, E. (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. The Quarterly Journal of Economics, 136, 2147–2194, available at: https: //doi-org.libproxy-wb.imf.org/10.1093/qje/qjab009.

Banerjee, A., Chandrasekhar, A., Duflo, E., Dalpath, S., Floretta, J., Jackson, M., Kannan, H., Schrimpf, A. and Shrestha, M. (2020). Evaluating the impact of interventions to improve full immunisation rates in Haryana, India. 3ie Impact Evaluation Report 126, available at: https://www.3ieimpact.org/evidence-hub/ publications/impact-evaluations/evaluating-impact-interventions-improve-full.

Banerjee, A., Duflo, E. and Glennerster, R. (2008). Putting a Band Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System. Journal of the European Economic Association, pp. 487–500.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B. and Udry, C. (2015a). Development economics. A multifaceted program causes lasting progress for the very poor: evidence from six countries. Science (American Association for the Advancement of Science), 348, 1260799–1260799.

Banerjee, A., Karlan, D. and Zinman, J. (2015b). Six Randomized Evaluations of Microcredit: Introduction and Further Steps. American Economic Journal: Applied Economics, 7 (1), 1–21.

Banerjee, A. V., Cole, S., Duflo, E. and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. The Quarterly Journal of Economics, 122 (3), 1235–1264.

Baseler, T., Ginn, T., Hakiza, R., Ogude, E. and Woldemikael, O. (forthcoming). Can Aid Change Attitudes toward Refugees? Experimental Evidence from Microentrepreneurs in Urban Uganda. Journal of Development Economics.

Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M. and Shapiro, J. P. (2019). No Household Left Behind: Afghanistan Targeting the Ultra Poor Impact Evaluation. Policy Research working paper, no. WPS 8877 Washington, D.C. : World Bank Group., available at: http://documents.worldbank.org/curated/en/855831560172245349/ No-Household-Left-Behind-Afghanistan-Targeting-the-Ultra-Poor-Impact-Evaluation.

Bousquet, J., Kiemtoré, I., Mvukiyehe, E., Verpoorten, M. and of the Democratic Republic of Congo (FSRDC), S. F. (2021). Strengthening Community Resilience in Fragile and Conflict-Affected Societies. Evidence from a Randomized Controlled Trial of a CDD Program with a Conflict Resolution Dimension in Eastern Congo. Endline Impact Evaluation Report. Available at: https://documents1.worldbank.org/ curated/en/619651625059705063/pdf/Endline-Impact-Evaluation-Report.pdf.

Caron, L. and Schuettler, K. (2020). Jobs Interventions for Refugees and Internally Displaced Persons. Journal of Development Economics, available at: https: //openknowledge.worldbank.org/handle/10986/33953.

Chaaban, J., Ghattas, H., Salti, N., Moussa, W., Irani, A., Jamaluddine, Z. and Al Mokdad, R. (2020). Multi-purpose cash assistance in Lebanon. Impact evaluation on the well-being of Syrian refugees.

Chinen, M., de Hoop, T., Balarin, M., Alcazar, L., Sennett, J. and Mezarina, J. (2018). Vocational and Business Training to Improve Women's Labor Market Outcomes in Low and Middle Income Countries — A systematic review.

Cohen, J. and Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. The Quarterly Journal of Economics, 125 (1).

Dupas, P. (2011). Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya. American Economic Journal: Applied Economics, 3 (1), 1–34.

Esper, H., Krishnan, N. and Wieser, C. (forthcoming). More is better: Evaluating the impact of a variation in cash assistance on the reintegration outcomes of returning Afghan refugees.

Falb, K., Laird, B., Ratnayake, R., Rodrigues, K. and Annan, J. (2019). The ethical contours of research in crisis settings: five practical considerations for academic institutional review boards and researchers. Disasters, 43, 711–726.

Glennerster, R., Duflo, E. and Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit. National Bureau of Economic Research Technical Working Paper Series, 333, available at: http://www.nber.org/papers/t0333.

Glennerster, R. and Kudzai, T. (2013). Running Randomized Evaluations: A Practical Guide. Princeton University Press, available at: http://runningres.com/.

Glewwe, P., Ilias, N. and Kremer, M. (2010). Teacher Incentives. American Economic Journal: Applied Economics 2, pp. 205–227.

Glewwe, P., Kremer, M. and Moulin, S. (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. American Economic Journal, pp. 112–135.

Glewwe, P., Kremer, M. and Moulin, S. and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. Journal of Development Economics, 74 (1), 251–268.

Hussam, R., Kelley, E. M., Lane, G. and Zahra, F. (forthcoming). The Psychosocial Value of Employment. Journal of Development Economics.

Laudati, A., Mvukiyehe, E. and van der Windt, P. (2018a). Participatory Development in Fragile and Conflict-affected Contexts: An Impact Evaluation of the Tuungane 1 Program in the Democratic Republic of the Congo. Available at: https://www.rescue.org/report/participatory-development-fragile-and-conflict-affected-contexts-impact-evaluation-tuungane-1.

Laudati, A., Mvukiyehe, E. and van der Windt, P. (2018b). Participatory Development in Fragile and Conflict-affected Contexts: An Impact Evaluation of the Tuungane 2 Program in the Democratic Republic of the Congo. Available at: https://www.rescue.org/report/participatory-development-fragile-and-conflict-affected-contexts-impact-evaluation-tuungane-2.

Metzler, J., Diaconu, K., Hermosilla, S., Kaijuka, R., Ebulu, G., Savage, K. and Ager, A. (2019). Short- and longer-term impacts of Child Friendly Space Interventions in Rwamwanja Refugee Settlement, Uganda. J Child Psychol Psychiatry. Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. Econometrica, 72 (1), 159–217.

Phadera, L., Sharma, D. and Wai-Poi, M. G. (2020). Iraq's Universal Public Distribution System : Utilization and Impacts During Displacement (English). Policy Research working paper, no. WPS 9155 Washington, D.C. : World Bank Group., available at: http://documents.worldbank.org/curated/en/239031582135436157/Iraqs-Universal-Public-Distribution-System-Utilization-and-Impacts-During-Displacement.

Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y. and Brück, T. (2015). What Methods May Be Used in Impact Evaluations of Humanitarian Assistance? IZA DP No. 8755 Discussion Paper Series. —, —, —, — and — (2017). Can rigorous impact evaluations improve humanitarian assistance? Journal of Development Effectiveness, pp. 519–542.

Quattrochi, J., Aker, J. C., van der Windt, P. and Voors, M. (2020a). Contributions of experimental approaches to development and poverty alleviation: 30 Field experiments and humanitarian assistance. World Development, 127, available at: https://doi.org/10.1016/j.worlddev.2019.104830.

Quattrochi, J., Bisimwa, G., van der Windt, P. and Voors, M. (2020b). Effect of an Economic Transfer Program on Mental Health of Displaced Persons and Host Populations in Democratic Republic of Congo: A Randomised Controlled Trial. The Lancet Preprints, available at: SSRN: https://ssrn.com/abstract=3706054 or http://dx.doi.org/10.2139/ ssrn.3706054.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. Journal of Development Economics, 74, 199–250, available at: https: //www.sciencedirect.com/science/article/pii/S0304387803001858.

Shamsuddin, M., Acosta, P. A., Battaglin Schwengber, R., Fix, J. R. and Pirani, N. (2021). Economic and Fiscal Impacts of Venezuelan Refugees and Migrants in Brazil. Policy Research working paper, no. WPS 9797 Washington, D.C. : World Bank Group., available at: http://documents.worldbank.org/curated/en/865871633611414287/ Economic-and-Fiscal-Impacts-of-Venezuelan-Refugees-and-Migrants-in-Brazil.

Tappis, H. and Doocy, S. (2018). The effectiveness and value for money of cash-based humanitarian assistance: a systematic review. Journal of Development Effectiveness, 10 (1), 121–144.

UNICEF – The United Nations Children Fund and IDMC – Internal Displacement Monitoring Center (2019). Equitable access to quality education for internally displaced children.

Tol, W. A., Barbui, C., Galappatti, A., Silove, D., Betancourt, T. S., Souza, R., Golaz, A. and van Ommeren, M. (2011). Mental health and psychosocial support in humanitarian settings: linking practice and research. The Lancet, 378 (9802), 1581–1591, available at: https://www.sciencedirect.com/science/article/ pii/S0140673611610945.

Tol, W. A., Leku, M. R., Lakin, D. P., Carswell, K., Augustinavicius, J., Adaku, A., Au, T. M., Brown, F. L., Bryant, R. A., Garcia-Moreno, C., Musci, R. J., Ventevogel, P., White, R. G. and van Ommeren, M. (2020). Guided self-help to reduce psychological distress in South Sudanese female refugees in Uganda: a cluster randomised trial. The Lancet Global Health.

UNHCR – United Nations High Commissioner for Refugees (2019). Stepping up: Refugee education in crisis. Available at: https://www.unhcr.org/steppingup/.

UNHCR – United Nations High Commissioner for Refugees (2021). Global Trends in Forced Displacement – 2020. Available at: https://www.unhcr.org/ 60b638e37/unhcr-global-trends-2020.

UN – United Nations (2018). Report of the United Nations High Commissioner for Refugees — Part II — Global compact on refugees.

Vermeersch, C. and Kremer, M. (2005). School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. Policy Research Working Paper; No. 3523. World Bank, Washington, DC.

Woodruff, C. (2018). Addressing constraints to small and growing businesses. IGCANDE-Argidius Working paper. Available at: https://www.theigc.org/wp-content/ uploads/2018/11/IGC_ANDE-review-paper_final-revised.pdf.

Appendix 1 – Selection bias

Let's imagine that a government is interested in piloting a vocational training program to better prepare youth to join the workforce. The pilot is initially planned for 1,000 youth. The program is advertised, and 1,000 youth apply and participate in the training. Two years later, a survey is conducted to measure employment outcomes. The economy in the country has been faring well in the past two years, so the government is aware that a pre-post evaluation (looking at income before the program and after the program) will give a biased estimate of the training's impact. So, the government decides to compare outcomes for youth that participated in the program to outcomes of youth who didn't. To find youths that didn't attend the program, the consulting firm asks each youth for the name of a friend that wasn't in the program.

- Surveys are conducted with both groups and it is determined that average income is lower for youth that participated in the program compared to youth that didn't. These results are quite the opposite of what we would expect as in the worst case, we would expect the training to have little or no impact. **Selection bias** may be at the root cause of these findings. The two groups of youth that were compared (those that attended the training and the friends who didn't) may be different. Youth that applied to the program may have been faring worse than youth that didn't attend it because they were not interested in the first place (they may have been more educated and already employed). The valid comparison group would have been a group of youth with similar unemployment rates. Our impact estimates in this case are likely **biased downwards**.

- Similarly, positive impact results should not be trusted in this study design. Youth that participated to the training may be more motivated – unemployed but keeping an eye out for opportunities – or they may have been recommended (so already be earning their lives well before participating into the training because they come from a well-off family). Hence, the impact estimate results are likely **biased upwards**.

Randomizing youth into the training program eliminates selection bias by ensuring that the study is comparing two groups of individuals that are, on average, similar. The typical way to introduce randomization for a training program is to screen applicants that qualify and randomly offer a seat in the training to the ones randomly selected (and "use" the ones that applied and qualified but were not trained as a comparison group to estimate the training's impact).

Appendix 2 – Pre-post evaluations

Pre-post (or before-after) impact evaluations are commonly used by agencies in their M&E processes and consist of comparing post-intervention outcomes to pre-intervention outcomes, for a specific group of individuals, typically those eligible to receive an intervention. This method has an important limitation as it relies on the assumption that there are no outside factors affecting outcomes of interest alongside the program or intervention being evaluated (as if the program operated in a vacuum). This concern is particularly relevant in forced displacement contexts, where the environment in which a project is implemented can be rapidly changing. The box below illustrates how this limitation – the inability to disentangle the effect of the program from the effect of external factors – may lead to failure to accurately measure the impact of a program/intervention.

Box - Pre-post evaluations

Imagine conducting a pre-post evaluation of a financial literacy program targeting poor households that took place in the United States in January 2020, before the COVID-19 crisis hit. Results from the baseline survey conducted in December 2019 show that households saved an average of $300 a month in the 12 months preceding the survey. If a follow-up survey was done now, in the middle of the COVID-19 crisis, likely, there would be a lower average monthly savings than in pre-COVID times, for example $150 on average per month. Did the training program cause such decrease in savings? Not likely but instead the decrease in savings is caused by the pandemic hardship. There may be some positive effects of the program, but a pre-post IE method doesn't allow to measure them.

If there had been a comparison group, it would have been possible to measure average monthly savings, finding that that savings were on average equal to $100. The difference between average monthly savings for the group that attended the training ($150) and the group who didn't ($100) is the correct estimate of the program impact: it helps households that participated save on average $50 more compared to a comparison group.

Appendix 3 – Impact evaluation examples (FCDO-UNHCR-World Bank *Building the Evidence on Forced Displacement* research program)

Impact evaluation of a Community-Driven Development project in refugee-hosting areas of Ethiopia

Context: The Development Response to Displacement Impacts Project (DRDIP) is a Community Driven Development (CDD) project implemented in refugee-hosting areas in four countries in the Horn of Africa (Djibouti, Ethiopia, Kenya, and Uganda), with the goal of enhancing the quality of health and education services available to both host and refugee populations. World Bank researchers are working with World Bank operational teams and government counterparts to conduct impact evaluations of the project in Ethiopia and Kenya.

Identification strategy: The research team has adopted a difference-in-differences method with statistical matching to identify suitable neighbor control villages and individuals for the impact evaluation, among the sites that were not covered by the pilot project.

Research questions: Does the program improve utilization and quality of health and education services in treatment communities as compared to comparison communities? Does the program improve social cohesion between host and refugee households? Does the program have an impact on gender norms?

Source: Ongoing study implemented by the World Bank Social Development Global Practice in partnership with the governments of Ethiopia and Kenya.

Impact evaluation of the UNHCR cash reintegration assistance program for Afghan returnees

Context: Between 2016 and 2018, more than 800,000 Afghan refugees returned from Pakistan to Afghanistan, following the deterioration of refugee policies in the host country. A data and analytics partnership between World Bank, the UNHCR and the government of Pakistan made it possible to collect and analyze data to understand the socio-economic conditions of returnees from Pakistan and inform programming needs. In the context of this existing partnership, an opportunity was available to rigorously evaluate the impact of the UNHCR cash reintegration program with an impact evaluation.

Identification strategy: In April 2017, UNHCR had to reduce the amount of the reintegration assistance it provided to returning refugees from USD 350 to USD 150 because of budget cuts. Afghan returnees were not aware of this policy change nor could they know when it was going to be introduced. The sudden and unexpected nature of the policy change "naturally" created two similar groups of Afghan returnees, one group returning to Afghanistan with an allowance of $350 and the other group returning with an allowance of $150. The situation lent itself to an impact evaluation using a regression discontinuity design (RDD), where the "discontinuity" that defines the treatment and comparison groups is defined by the day the cash assistance amount changed.

Preliminary findings: Returnees who received the higher cash assistance amount were more likely to purchase long-term assets and were more likely to have legal documentation for all household members. Instead, different cash reintegration amounts had no different effects on employment outcomes nor enrollment rates of children into schools.

Source: More is Better: Evaluating the Impact of a Variation in Cash Assistance on the Reintegration Outcomes of Returning Afghan Refugees (forthcoming study implemented by the World Bank Poverty and Equity Global practice in partnership with the UNHCR).

Impact evaluation of a new training curriculum for coaches

Context: The Syrian refugee crisis has led to a significant increase in enrollment in public schools in Lebanon, with a doubling of the student population in just over four years (in mid-2019 Syrian refugee children represented 45% of all public-school students in the country). In 2013, the Government of Lebanon, with support from the international community, launched the *Reaching all Children with Education (RACE)* initiative, with the objective of maintaining high quality education despite the

challenges brought by the displacement crisis. As part of the initiative, an enhanced training curriculum for coaches will be introduced as a pilot and scaled-up nationally if the evaluation finds the new curriculum is effective.

Identification strategy: This impact evaluation will employ a randomized controlled trial (RCT) method where half of the coaches will continue with the current curriculum (comparison group) and half will receive training on the new curriculum.

Research questions: The randomized evaluation will generate evidence on the following research questions: 1) What is the impact of the training on a) coaches' knowledge? b) teachers' received dosage of coaching? c) teachers' perceptions of coaching session quality? d) teacher satisfaction, e) classroom instruction and (f) students' perception of their teachers and their academic engagement? Additionally, the RCT will generate findings on the specific mechanisms driving impact, whether impacts vary by teacher, coach and/or student characteristics and other data of interest to the Ministry of Education.

Source: Ongoing study implemented by the World Bank Education Global Practice in partnership with the Lebanese Ministry of Education and Higher Education.

Impact evaluations of socioemotional learning and pedagogical interventions in Jordan

Context: In the context of classroom environments where half of the students report that classes are disrupted by noise and disorder, the *Providing Opportunities with Education for Refugees and Jordanians (POWER-J)* project in Jordan will introduce and rigorously test two interventions (socioemotional and pedagogical) aimed at improving classroom environments and ultimately enable better teaching and learning for grade 7 students.

Identification strategy: The IE will employ a randomized controlled trial (RCT) method with schools assigned to different interventions (either to receive the pedagogical intervention only, or the pedagogical intervention and different dosages of the behavioral intervention) or to a comparison group.

Research questions: What is the impact of the socioemotional skills intervention on students' socio-emotional skills? What is the impact of the pedagogical intervention on teacher's classroom management skills? What is the impact of both interventions on school attendance?

Source: Ongoing study implemented by the World Bank Education Global Practice in partnership with the Jordan Ministry of Education.

Randomized evaluation to test a 'mental imagery' intervention for refugees in Ethiopia

Context: Ethiopia has committed to the ambitious goal of granting the out-of-camp policy to 10% of the total refugee population (70,000 – 90,000 people) and rights to work to up 30,000 individuals. There are concerns however that these policy commitments alone may not suffice to improve the refugees' quality of life and that employment support services are needed to address the unique challenges refugees face. An impact evaluation will test a low-cost and scalable mental imagery intervention that could help refugees tap into the full potential of the newly granted right to work.

Identification strategy: A randomized controlled trial impact evaluation design will be employed to test this intervention. Researchers will work with implementing partners to randomly draw from the pool of job seekers. Randomization is possible given that it is anticipated that there will be at first more job seekers than jobs available.

Research questions: 1) Does the imagery approach change future-oriented behavior, including investments into micro-enterprises or job search, hours worked, seeking training opportunities, and children's school attendance? 2) Does the imagery approach improve labor market outcomes, including employment status and income?

Source: Ongoing study implemented by the World Bank Social Protection and Jobs Global Practice in partnership with external researchers, the UNHCR and the Government of Ethiopia.

Randomized evaluation to measure the impact of a photography training intervention for Rohingya adolescent girls

Context: Adolescent girls make up a large portion of Rohingya refugees and are especially vulnerable to experiencing long-term trauma. The intervention seeks to improve their socio-emotional wellbeing through a photography-based art therapeutical training that teaches photography while at the same time building a sense of 'home' and belonging to a community. The intervention will be benchmarked against the same intervention, but the latter will focus on teaching photography skills.

Identification strategy: The study will employ a phase-in approach, where the implementing NGO will provide a register of all eligible adolescents. The first group receiving the intervention will be randomly selected.

Research questions: The study will measure the impact of the photography program on respondents' sense of belonging, sense of security, happiness, feelings of nostalgia, and psycho-social wellbeing.

Source: Ongoing study implemented by the World Bank Development Impact Evaluation unit (DIME).

Impact evaluation of a refugee-led microenterprise program in an urban setting: focus on social cohesion

Context: As one of the largest host of refugees in Africa and the world, Uganda is well recognized globally for its inclusive refugee policy. Most refugees in Uganda reside in government-designated settlements in rural areas and receive monthly food assistance, shelter, and a plot of land to farm. About 78,500 registered refugees (but likely many more) reside in Kampala and engage in business activities similar to those of the nationals, and so become their competitors. The evaluation seeks to understand whether an intervention to support small Ugandan businesses that makes explicit the link to Uganda's inclusive refugee policy (the intervention is led by a refugee-led organization) can be effective in building the political support for Uganda's inclusive refugee policies and improve social cohesion.

Identification strategy: The impact evaluation uses a randomized controlled trial method with factorial design: business owners are assigned to either: 1) receive cash and information about Uganda's inclusive refugee policy; 2) receive cash only; 3) receive only information about Uganda's inclusive refugee policy; 4) receive mentorship by a refugee and 5) receive mentorship by a Ugandan.

Research questions: Can aid-sharing generate political support for refugee integration?

Findings: All program variants are effective in increasing support for inclusive policies like refugees' right to work and allowing more refugees into the country. The cash and information program outperforms the other program variants.

Source: Can Aid Change Attitudes toward Refugees? Experimental Evidence from Microentrepreneurs in Urban Uganda (forthcoming study implemented by the World Bank Social Development Global Practice in partnership with external researchers and a refugee-led NGO in Uganda).

Impact evaluation of the Targeting the Ultra-Poor (TUP) graduation program in Afghanistan

Context: While global poverty has declined markedly over the past 30 years, gains have not been equally distributed. In Afghanistan, the share of people living below the national poverty line increased from 38% in 2011/12 to 55% in 2016/17. The ultra-poor in Afghanistan face multiple, simultaneous constraints that can trap them in persistent poverty. The situation is particularly acute for women: in the study areas in Balkh, less than 4% of primary women in the household can read and write, two in three of these women experience depression, and just over half of eligible girls attend school.

Identification strategy: The Microfinance Investment Support Facility for Afghanistan (MISFA) collaborated with the World Bank to address these challenges through the Targeting the Ultra-Poor (TUP) graduation program, as part of the IDA-funded Access to Finance program. By providing a time-limited, sequenced, "big-push" package that combines a large investment in a productive asset, access to savings accounts, temporary cash support, skills training, coaching, and other complementary services related to education and health, the TUP aims to lift ultra-poor households out of poverty. Addressing multiple

constraints simultaneously can offer an effective strategy for lifting households out of poverty in these settings if these constraints reinforce each other.

Findings: One year after the intervention ended, consumption had increased by 30% and poverty decreased by 20% (from 82% in the comparison group). The intervention improved women's labor participation in a context where gender gaps in access to assets and inputs, and discrimination in paid employment, are the norm. The program is cost-effective, with an estimated 26% internal rate of return. These results are the first positive and significant impacts of a TUP program in a conflict setting, with larger consumption increases than those reported in TUP programs in more stable settings (Banerjee et al. 2015; Bandiera et al. 2017).

Source: Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M. and Shapiro, J. P. (2019). No Household Left Behind: Afghanistan Targeting the Ultra Poor Impact Evaluation. Policy Research working paper, no. WPS 8877 Washington, D.C. : World Bank Group., available at: http://documents.worldbank.org/curated/en/855831560172245349/ No-Household-Left-Behind-Afghanistan-Targeting-the-Ultra-Poor-Impact-Evaluation.

The Iraqi Public Distribution System and its impact on displaced households

Context: The Public Distribution System (PDS) is a universal food distribution program instituted in Iraq in 1990, in response to food shortages engendered by UN-imposed sanctions. The program is universal and in 2018 accounted for 2.3% of the total government budget spending. An impact evaluation of the PDS and its role in mitigating the displacement impacts was conducted with the goal of informing the ongoing policy dialogue concerning the program's reform.

Identification strategy: The impact evaluation leveraged data from the 2017-18 Rapid Welfare Monitoring Survey (SWIFT) and used a propensity score matching technique to match displaced households who didn't have access to the PDS to displaced households with access. The two groups of displaced households were matched based on a set of observable characteristics (household size, number of children between 7 and 14, age of household head, current place of residence is rural, education, date living in current governorate, and number of shocks since 2014).

Findings: Among displaced households, access to PDS plays an important role in mitigating – but only partially – the welfare loss due to displacement. Displaced households with access to PDS rations have higher food and non-food expenditures, spend 20% more on education, have a higher caloric intake, and lower levels of vulnerability to poverty. Additionally, those that retained access to PDS benefits consumed significantly fewer calories from non-ration food and spent more on non-food items. However, access to PDS did not seem to affect households' overall perceived economic outlook.

Source: *Phadera, Lokendra; Sharma, Dhiraj; Wai-Poi, Matthew Grant. Iraq's Universal Public Distribution System: Utilization and Impacts During Displacement (English). Policy Research working paper, no. WPS 9155, Impact Evaluation series Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/239031582135436157/Iraqs-Universal-Public-Distribution-System-Utilization-and-Impacts-During-Displacement*

Impact evaluation of an entrepreneurship support program (in the context of a cash transfer program)

Context: The Boko Haram crisis and instability in neighboring Mali have led to Niger hosting over 300,000 displaced people comprising of 166,000 refugees, 129,000 internally displaced people (IDPs) and 15,000 Nigerien nationals who have returned from Nigeria (UNHCR, 2017). In this context, Niger is implementing the *Niger Refugees and Host Communities Support Project*, a comprehensive program including infrastructure rehabilitation and provision of short-term employment in the form of cash-for-work and entrepreneurship support to stimulate local enterprise.

Identification strategy: The IE will employ a randomized controlled trial method with factorial design. This design will allow to study the differential impacts of the intervention on forcibly displaced people, hosts, women and youth.

Research questions: What are the social, economic, and psychological impacts of the entrepreneurship support interventions in a context of displacement?

Source: Ongoing study implemented by the World Bank Development Impact Evaluation unit, the Social Protection and Jobs Global Practice and the government of Niger.

Impact evaluation of an intervention to reduce gender-based violence (in the context of a cash transfer program)

Context: According to the 2011 Demographic and Health Survey, nearly half of all Cameroonian women were victims of emotional, physical and/or sexual violence from their partners. The Cameroonian government is implementing a Social Safety Net project to establish a basic national safety net system, that includes monthly cash transfers to targeted vulnerable households and accompanying measures. The project is scaling up to refugee-hosting communes.

Identification strategy: The World Bank Gender Innovation Lab will be working alongside the World Bank operational team and government Project Implementation Unit to rigorously assess, with a randomized impact evaluation design, two different interventions to prevent gender-based violence. These will be administered to randomly identified households.

Research questions: What is the impact of adding couples' training to a cash transfer program on intimate partner violence, cooperation within the household, and women's economic empowerment? What is the impact on these same outcomes of adding edutainment, in the form radionovela listening groups, to a cash transfer program? Is there a differential impact of either intervention on refugee vs. host households? Is there a differential impact of either intervention on monogamous vs. polygamous households? How cost-effective is each intervention?

Source: Ongoing study implemented by the World Bank Gender Innovation Lab, the Social Protection and Jobs Global Practice and the government of Cameroon.

Appendix 4 – Engaging in an impact evaluation for implementing teams

This Appendix describes how impact evaluations (and especially randomized evaluations) and their success depend on the effective collaborative between researchers and practitioners.

The implementing organization is very involved in designing the evaluation. This exercise can take various months (if not years) of careful planning. During this process, implementing teams need to ensure that the researcher understands how the program is typically carried out and the operational questions that are important for implementation. The researcher needs to guide the implementing agency through the trade-offs between possible impact evaluation methods and designs until an impact evaluation plan is developed which includes agreement on:

- *Timeline and sequencing of program and IE activities*. IE activities may include village listing, household sampling, baseline survey, randomization, and one or more rounds of follow-up surveys. In the field this translates into careful joint planning of the sequencing of impact evaluation and program activities. The box below contains an example of how research and program activities are linked in the impact evaluation of a graduation program.

Box – Activities in the IE of a graduation program

Month 1	Month 2	Month 2	Month 2-12	Month 14	Month 28	Month 32
Identification of target villages	Household targeting	Village randomization	Project implementation	Follow-up survey 1	Follow-up survey 2	Project implementation in control villages
Project team	Project team	Research team	Project team	Research team	Research team	Project team

- *Research questions*. Research questions are designed based on a program's theory of change and drive the data needs for the impact evaluation. It is important that implementing teams and researchers share a common understanding of how the program operates—including assumptions and possible mechanisms intended to improve intermediate and final outcomes.

- *Data needs.* After deciding on the key research questions that should be the focus of evaluation, teams will need to determine whether data is available within their administrative or M&E datasets or whether new ad-hoc data collection is required.

- *Cost.* Primary data collection needs are key cost drivers for an impact evaluation. In a forced displacement context, data may be limited (for instance in the case of refugees, data may be limited to key information UNHCR collects as part of their registration processes). Some data collection efforts are ongoing in forced displacement contexts that we anticipate will be useful for impact evaluation, such as the collection of cross-sectional and panel surveys that are representative of both host and displaced populations[38].

- *Key technical design elements for the study*. These are 1) the level of randomization, which is chosen so that the possibility of treatment "spillovers" is minimized[39], 2) the number of clusters (villages, schools, health centers, markets), 3) the number of individuals to survey in each cluster[40], 4) the number of treatment groups, and 5) whether or not to have a 'pure' comparison group ("a group that is only targeted and surveyed but does not receive the program").

- *Timing of the randomization in the program life cycle and how to "message" it to study participants*. Determining a suitable randomization strategy is a particularly sensitive aspect of a randomized evaluation and includes making choices on timing (*When alongside the project lifecycle should we randomize?*) and the communication and awareness plan for participants (*How do we inform study participants that they may or may not be offered the intervention without harming the reputation of the implementing organization?*).

The implementing organization remains engaged in the research throughout its implementation. Continuous engagement of the implementing agency is critical to the successful implementation of the evaluation. In the following paragraphs describe the various ways the implementing team remains engaged throughout the course of the IE:

- **Rigor in program implementation**. Because conducting a randomized evaluation typically involves significant staff and financial effort, it is important that researchers and practitioners be confident in the results and learn from them. If the results of the experiment suggest that the program had little or no impact (no observed difference in outcomes between project participants and non-participants), it is important to know if this was due to poor implementation. For this reason, the types of programs that lend themselves best to an impact evaluation are those that can be implemented in a standardized way.

- **Strict adherence to the impact evaluation design**. Implementing a program using randomized evaluation often changes considerably the standard operating procedures for program implementation that field teams are used to. To minimize the likelihood that deviations from the impact evaluation design occur in the field (such as teams swapping a "control" village for a "treatment" village for instance), it is important to ensure field team understand the value of the research and are committed to making it a success.

---

[38] The World Bank and the UNHCR have launched several initiatives to collect data representative of both displaced and host communities. For example, in Uganda, the Uganda Refugee and Host Communities 2018 Household Survey is a survey representative of both host and refugee households. In Nigeria, the Profile of Internally Displaced Persons in North-East Nigeria 2018 is a representative survey of IDPs living in camps, host communities and host households. In Bangladesh, the WB and IPA have collected a survey representative of Rohingya refuges and host communities (see Cox's Bazar Panel Survey). Furthermore, the World Bank and the UNHCR have established the Joint Data Center (JDC), a dedicated center to improve the quality of data available on forced displacement.

[39] Treatment spillovers may lead to underestimate program impact: intuitively if some of the individuals that are part of the comparison group in fact receive the treatment, when we compare outcomes between the two groups, we will underestimate impact.

[40] There is a cap on the # of individuals that we need information about within the same cluster. Since we expect individuals living in the same cluster (a village for example) to be similar (for example engage in similar income-generating activities), we don't need to survey too many from the same cluster.

- **Documentation of program implementation.** In practice, and especially in a context of displacement, program activities can change during implementation. In the context of an impact evaluation, it is important to document the changes so that researchers can be aware and consider them when analyzing the data. For this reason, in the context of an impact evaluation, the implementing organization is asked to keep a log of the implementation issues that arise. For example, field teams may have a hard time implementing in a specific village because they face reluctance to engage from the village leader: this information should be shared with the researcher to be considered in the analysis.

- **Minimizing "attrition".** Attrition refers to the inability to interview individuals that were surveyed at baseline during follow-up survey(s). Attrition that could be caused by the program – "systematic attrition" – is problematic in the context of an impact evaluation as it could bias the estimation of the causal effects of the program. For example, in the case of the evaluation of a business training program, it is possible that participation in the program improved business revenues for some business owners leading them to seek opportunities in the city, making them unavailable to take the follow-up survey. The inability to survey these individuals can lead to underestimating the impact of the program as the population of business owners who stayed behind is on average less successful than the population that left. Program implementation teams who are more often in contact with program participants can help locate households for the research team.