Check-list to identify projects

Objective: Identify projects whose design is suitable for an impact evaluation with the aim of a) providing credible evaluations and policy relevant insights for IOM, b) answering questions of interest to IOM and c) contributing to academic publications for IPL.

The challenge of causal project evaluation

How can we evaluate and quantify the impact of a project using statistical methods?

Common methods:

- **Before-after approach**: compare the outcome before and after the project started
- **Comparison of treatment and control group**

Before-after approach can be problematic if other non-controllable factors affect the outcome and `confound' the analysis, e.g., the political or economic situation in the country changes during the project or another policy is implemented in the same time period.

Likewise, a simple treatment-control comparison might yield misleading results if the two groups are not comparable, e.g., because treated individuals actively decide (self-select) to be exposed to treatment. For example, some migrants are more willing to participate in an IOM program than others.

The impact is a comparison between:
- the outcome after the program started (what we observe; in yellow)
- the outcome if the program had not been introduced (the "counterfactual" which we don't observe; in blue)

Figure 1 - source JPAL training material

In an optimal world, we would compare the outcome of a treated individual to the counterfactual where that individual had not received the treatment. Unfortunately, we don't observe the counterfactual. Approximating the counterfactual is the challenge of causal impact evaluation.

Strategies for causal impact evaluation (project-based)

A wide range of strategies and statistical methods is available for causal impact evaluation:

- **RCT (Randomized Control Trial)**: randomized assignment of treatment and control group
- **Stepped wedge design**: comparing beneficiaries who receive the treatment at different points in time
- **RDD (Regression Discontinuity Design)**: exploits "arbitrary" policy thresholds
- Etc.

In the context of the IOM-IPL collaboration in LAC, for a project-linked survey, we should focus on projects which allow for causal impact evaluation. In order to evaluate the impact of a project, we need to survey both project beneficiaries (treatment group) and another group that has not benefitted from the project, but is comparable in terms of all other characteristics.

There are several options for credible impact evaluation. Below we describe three types of projects that are suitable foran impact evaluation and the necessary conditions. We also provide a checklist to identify relevant projects.

A. Randomized Control Trial (RCT)

We randomize treatment and control group. Then random assignment ensures that the two groups are comparable onall characteristics except for the treatment. A comparison of treatment and control groups in terms of the observed outcome of interest allows for a causal impact evaluation.

Conditions:

1. We advise to discuss the project design with the research team before the start of the implementation. The earlier the research team is involved, the easier it is to plan a project design suitable for the evaluation and feasible for IOM and its partners.
2. Sufficient number of respondents: at least 500 in the control group and 500 in the treatment group; ideally 1,000 in each.

Project examples:

Example 1: We want to evaluate two methods of financial literacy training. The new method emphasizes classic accounting principles, and the existing approach focuses on simple "rule of thumb" methods for decision making. From a pool of 1,200 beneficiaries who were planned to receive financial training, 600 are assigned to the usual training, the other 600 are assigned to the new type of training.

Example 2: Refugees have been selected for resettlement. Before they leave for their new country, they attend a pre-departure orientation training. The organization delivering the training wants to test a new component of the training to decide whether it should be adopted or not. 500 participants are assigned to the usual training while the other 500 receive the new training component. The difference in an observed outcome between treatment and control group (e.g. participant's average satisfaction) can be interpreted as the impact of the new component.

B. Stepped-wedge designs

In a stepped-wedge design, we randomize *the order* of the treatment. The entire sample eventually receives the treatment (benefit from the IOM program), but different groups of beneficiaries start the treatment at different points in time. We compare individuals who benefitted from the treatment with individuals who have not yet benefitted from the program and thus constitute the control group. The crucial difference to a standard RCT is that everyone will benefit from the program eventually.

Conditions:

1. Sufficient time gap between each group of beneficiaries

The minimum required time gap between treatment groups depends on the effect of the intervention and what outcomes we want to measure.

Minimum 3 month-time gap:
For example, with a psychological support intervention, usually the effects on the psychological well-being can be seen relatively quickly. In this case, a minimum of 3 month-intervals between groups is acceptable.

Or more:
Yet, in order to capture effects on more persistent outcomes such as the effect of the intervention on the beneficiaries' integration, the time gap between groups needs to be larger.

There is also the option to conduct follow-up interviews a year after implementation phase.

In general, in terms of the statistical analysis, a larger time gap between the groups of beneficiaries facilitates precise measurement of project impacts. From a practical point of view this can also be an advantage, since there are often delays in the implementation and/or the mission doesn't have the capacity to provide the intervention to all beneficiaries at once.

2. At least 500 respondents

Groups are compared among each other. Thus, there needs to be enough beneficiaries to be able to compare them to each other. Indeed, the ones receiving the intervention towards the end can only be used as a control group but cannot be included in the treatment group since there won't be anyone left to compare them with.

There should be at least 500 respondents. The project should preferably include more respondents but we consider 500who are 18+, migrants and who agree to participate in the study as the absolute minimum.

Sufficient variation in group assignment is also important for statistical precision. In other words, treatment groups should be of approximately equal size.

3. Contact details of respondents

As IOM will eventually need to collect the contact details of beneficiaries, ideally it has the contact details of all beneficiaries before they start the project so that we can interview them all once.

Project examples:

Example 1:
A project offers a 4-week psychotherapy session to refugees. The organization does not have the capacity to provide the session to everyone at once so there is a waiting list of several months. We conduct 3 surveys. Here are the main steps:
- We survey all the people eligible to the program before they start (interview #1)
- After a waiting period, some beneficiaries start receiving the psychotherapy session (intervention)
- 4 weeks after the intervention ended, we conduct interview #2
- 3 months after the intervention ended, we conduct interview #3
- 6 months after the intervention ended, we conduct interview #4

The time intervals between groups allow us to compare people who have received the treatment with people who have not (yet).

Example 2: The City of Zurich sent information letters to immigrants newly eligible for naturalization but groups of immigrants received it at different points in time. The first group (the treatment group) received the letter in October 2019, while the second group (the control group) received the letter in May 2020. This design allows us to precisely estimate the short and medium-term effects of the letter without depriving individuals eligible for naturalization of the letter.

	2021										2022		
	Jan	Feb	March	April	May	July	Aug	Sept	Nov	Dec	Jan	May	Sept
months	1	2	3	4	5	7	8	9	11	12	1	5	9
pilot group - 30 respondents	Interview #1	Intervention	Interview #2			Interview #3			Interview #4				
group 1 - 250			Interview #1	Intervention	Interview #2			Interview #3			Interview #4		
group 2 - 250			Interview #1				Intervention	Interview #2			Interview #3	Interview #4	
group 3 - 250			Interview #1							Intervention	Interview #2	Interview #3	Interview #4
					A		B				C		

	Control
	Treated
Intervention	Respondents benefit from the program
A	The 250 respondents of group 1 become the treatment group and can be compared to the other 500 respondents in groups 2 and 3 i.e. the control group
B	The 250 respondents of group 2 can be compared to the remaining 250 respondents in group 3 i.e. the control group
C	The 250 respondents of group 3 become treated and cannot be compared to any control group because there no longer is a control group for them.

4

C. Cutoffs in Eligibility: Regression Discontinuity Design (RDD)

Regression Discontinuity Design (RDD)In a stepped-wedge design, we randomize *the order* of the treatment. The entire sample eventually receives the treatment (benefit from the IOM program), but different groups of beneficiaries start the treatment at different points in time. We compare individuals who benefitted from the treatment with individuals who have not yet benefitted from the program and thus constitute the control group. The crucial difference to a standard RCT is that everyone will benefit from the program eventually.
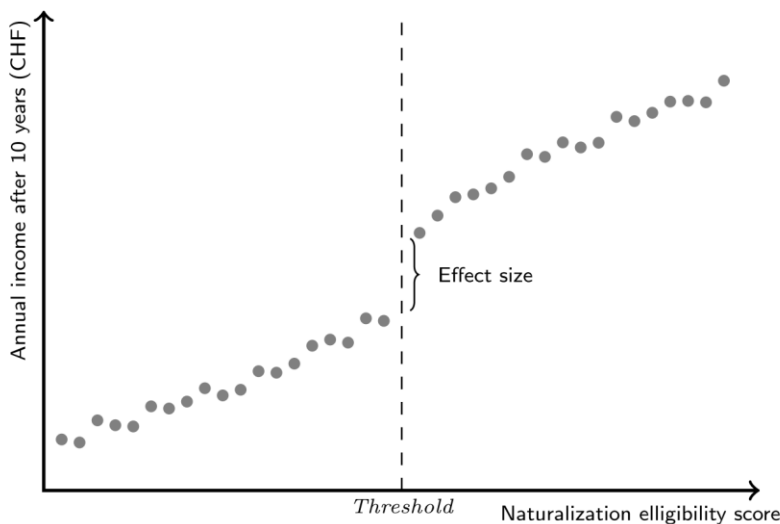

Figure 2 – Example of RDD

Conditions:

1. Contact Details

- IOM has access to the contact details of people eligible to the program (beneficiaries)
- IOM also has access to the contact details of people who were nearly selected to be beneficiaries of the project. (Practical implications: when doing the outreach campaign or selection of beneficiaries, IOM or its partners collect the contact information of all the applicants to the program.)
- IOM is allowed to reach out to and survey the two above mentioned groups

2. Well-defined selection criteria

The selection process of beneficiaries has a selection "rule" that determines who is assigned to treatment and control group. In order to be selected and become a beneficiary of the project, people have to meet a certain threshold.

3. At least 1,000 respondents

As a rule of thumb, if an intervention has a moderate to sizeable effect, there should be at least 500 respondents in the treatment group and 500 in the control group, hence a total of 1,000 respondents. Ideally, and when possible, there should be 1,000 in the treatment group and 1,000 in the control group, hence a total of 2,000 respondents. (The IOM project can include more beneficiaries but there should be at least 500 of them who are adult migrants who agree to participate in the study and 500 adult migrants who were barely eligible).

Project examples:

Example 1: We compare the integration trajectories of immigrants applying for Swiss citizenship who were narrowly rejected with applicants who were narrowly approved. In the first 5 years they have similar trajectories but in the long run we observe a difference in their annual income of CHF 5,000 on average.

Example 2: To benefit from a cash assistance program, people have to earn less than $500/month. In this scenario, the cash program beneficiaries are the treatment group, whereas people who are barely above the threshold (i.e., earn slightly more than $500/month), and were thus not selected, are the control group.

Conditions applicable to all project and research designs

The following criteria apply to any of the designs that you choose from the ones described above.

1. A sufficient number of respondents

The minimum size of respondents required depends on the effect of the intervention and on the research design. As a rule of thumb, if an intervention has a moderate to sizeable effect, there should be at the very least 500 respondents in the treatment group and 500 in the control group, hence a total of 1,000 respondents. Ideally, and when possible, there should be 1,000 in the treatment group and 1,000 in the control group, hence a total of 2,000 respondents.

In a stepped-wedge approach, since participants to the program serve both as treatment and control group depending on the time they start participating in the project, the number of respondents can be a bit lower. 500 may be enough but the more the better.

As a rule of thumb, the more respondents, the better. More respondents imply a larger number of observations that we can use in the statistical analysis. A larger sample size increases the precision of the effects size estimates and increases the statistical power, i.e. the ability to reliably detect effects of small to moderate size.

2. Selected among adult migrants who give informed consent

Adults

We do not interview minors so the respondents must be aged 18 years or older. This does not mean that a project with beneficiaries from all ages should be discarded, as long as enough adult beneficiaries could be interviewed (see above the minimum numbers of interviewees).

Literacy

The minimum level of literacy required depends on the survey mode. If the survey is self-administered, illiterate migrants cannot participate. If the survey is administered by enumerators, illiterate migrants can participate.

Legal status

In the context of the IOM-IPL LAC collaboration, only migrants can be included as respondents, not host country nationals. Migrants can be of different nationalities. This does not imply that projects which benefit both locals and migrants should be discarded. As long as enough adult migrant beneficiaries could be interviewed (see the minimum numbers of interviewees). From the IPL perspective, the legal status of respondents does not matter, they can have entered the country legally or illegally, have a residency permit or not etc.

Informed consent

Respondents provide informed consent after being introduced to the study. They will be informed of the implications in a language they understand. This step can happen at the start of the interview.

Summary checklist for project evaluation:

This document provides a check-list to help you to identify which of your projects could be eligible for implementing an impact evaluation in the context of the IOM-IPL collaboration in LAC.

We kindly ask you to conduct an inventory of your ongoing and upcoming projects. The check list below should help you to identify projects that fulfill the minimum criteria. We can then identify together which projects are better suited for an evaluation and are of interest to the missions and IOM HQ.

	RCT (A)	Stepped-wedge (B)	RDD (cutoff) (C)
Well-defined selection criteria	NA	NA	√
IOM has contact details of beneficiaries	√	√	√
IOM has contact details of people barely eligible for the program	NA	NA	√
IOM has authorization to survey beneficiaries	√	√	√
IOM has authorization to survey people barely eligible for the program	NA	NA	√
Number of respondents (adult migrants who agree to participate)	≥ 1,000	≥ 500	≥ 1,000
Minimum 3 months between groups of beneficiaries starting the program	NA	√	NA

Population-based survey

If no project fits the criteria outlined above and if this is of interest to IOM, we could also survey a well-defined population of migrants:

1. Representative sample

Representative samples are designed to learn about characteristics of a whole population, for example the population of migrants in a country, when it is not feasible to collect data on every member of that population (due to practical/financial constraints). For example, a representative sample of Venezuelan migrants in Peru would enable us to draw conclusions about the full population of Venezuelan migrants in the country.

Conditions:

In order to build a representative sample of the whole migrant population in a country, we would need to know the composition of the population of migrants in the country. We currently do not have sufficient data on this. If at a later stage, IOM has more information on the prevalence of migrants <u>and</u> COVID restrictions are lifted allowing enumerators to travel through the country and conduct face to face interviews, it may be possible to build a representative sample. Note that this is an expensive exercise. The types of data that would be needed are: age distribution, gender distribution, origin country distribution, distribution of residency (when they arrived), education distribution.

Project example:

Example 1: we conducted stratified random sampling in order to recruit a representative sample of Syrians in Lebanon. First, we randomly selected localities based on the prevalence of Syrian refugees using UNHCR data and the sectarian component of the localities using data from the 2018 elections in Lebanon. In each locality that was selected, we interviewed a local leader (such as the mayor) to find locations where Syrians are concentrated and then conducted a random walk sampling strategy in those areas to identify Syrian households to be interviewed.

2. Use of existing databases

Since building a representative sample is not feasible in our context, we could, if this is of interest to IOM, build a representative sample of migrants who are IOM beneficiaries. While we could **not** draw conclusions about all migrants, this approach would allow us to draw conclusions that are valid for the *sub*-population of IOM beneficiaries in a country.

Another option is to survey a population of IOM beneficiaries and non-IOM beneficiaries. This, however, has the caveat that the sample would not be representative for the whole population of IOM beneficiaries.

Conditions:

1. A sufficient number of respondents

At least 800-1,000 respondents (adult migrants who agree to participate in the study).

We advise against merging databases across the 3 countries as people who migrated to Dominican Republic may be fundamentally different from people who migrated to Brazil. Similarly, the conditions in each country will differ.

Once IOM has identified the subgroups of interest e.g. women/men, single head of households, with or without legal status etc, we need to make sure that there at least 100 (ideally more) of each subgroup in the sample.

2. Have access to the database
3. The database contains updated contact details