End-line impact assessment of the 2017-2021 DGD-funded programme implemented by Rikolto

Farmer survey data analysis – Lessons learnt

March 2022

*This document synthetizes the main lessons learned based on ADE's analysis of the Farmer Survey (FS) data collected by Rikolto at baseline (2017), midterm (2019) and endline (2021). These remarks aim at helping further **improve the capacity of Rikolto to create, implement and analyze its FS.***

1. Sampling strategy

Sampling is the process of drawing units from a population of interest to estimate the characteristics of that population. Sampling is necessary, as typically it is not possible to directly observe and measure outcomes for the entire population of interest.

Sample size

Sample sizes must be big enough to detect a change in relevant indicators. *Power calculations* can be used to estimate the minimum size of the sample size needed to detect a change in an evaluation. If you wish to assess the impact of Rikolto's intervention for a specific commodity in a country, the sample size of respondents in this country producing this commodity must also be large enough.

The power (or statistical power) of an impact evaluation is the probability that it will detect a difference between the treatment and comparison groups, when in fact one exists. Thus, drawing on large samples makes it less likely that we will conclude that the program has had no impact, when in fact it has had an impact. We recommend overestimating the sample size at baseline to account for potential attrition (see below).

Moreover, if it is in the interest of Rikolto to find out whether its intervention affects groups differently (i.e., **females vs males and younger vs older producers**), the sample sizes to observe any change should be balanced and large enough.

Selection of a treatment group

Before selecting the units (individual or cluster) of the treatment group, the population of interest needs to be clearly defined. This requires accurately specifying the unit within the population of interest for which outcomes will be measured, and clearly defining the geographic coverage or any other relevant attributes that characterize the population of interest. Stratified random sampling can then be used to select the respondents.

Stratified random sampling: The population is divided into groups (for example, male and female), and random sampling is performed within each group. As a result, every unit in each group (or stratum) has the same probability of being drawn. Provided that each group is large enough, stratified sampling makes it possible to draw inferences about outcomes not only at the level of the population but also within each group. Stratified sampling is useful when you would like to oversample subgroups in the population that are small (like minorities) in order to study them more carefully. Stratification is essential for evaluations that aim to compare program impacts between such subgroups.

Selection of a comparison group

To be able to estimate the impact of a program on outcomes, any impact evaluation method chosen must estimate the so-called *counterfactual*: that is, what the outcome would have been for program participants if they had not participated in the program. In practice, impact evaluation requires that the evaluation team finds a **comparison group** to estimate what would have happened to the program participants without the program, then make comparisons with the treatment group that has received the program. A comparison group (sometimes called the *control group*) is then the group that remains unaffected by the program, and allows to estimate the counterfactual outcome. Therefore, the importance of having a comparison is that if the two groups are identical, with the sole exception that one group participates in the program and the other does not, then we can be sure that any difference in outcomes must be due to the program itself.

In this sense, a valid comparison group must be the same as the treatment group in at least three ways. First, it has the same characteristics, on average, as the treatment group in the absence of the program. Second, it should not be affected by the program either directly or indirectly. Third, it would react to the program in the same way as the treatment group, if given the program. Stratified random sampling can also be used to select the comparison.

To observe any change, in addition to have a balanced and large enough sample size, it is also important to ask the same questions to the treatment and control group across survey waves. The importance relies in that it is very difficult or not possible to compare outcomes between group if each outcome is measured differently.

Attrition

Attrition is a type of selection bias that occurs when parts of the sample disappear over time, and researchers are not able to find all initial sample members of the treatment and comparison groups in follow-up surveys or data. Differences between individuals who leave a program and those who continue, particularly between study groups, can be the reason for any observed effect and not the intervention itself. Therefore, it is important to try to follow the same number of individuals in each group across survey waves to minimize the bias coming from attrition.

If we find attrition during a follow-up survey, the following two tests can help you assess the extent of the problem. First, we can check whether the baseline characteristics of the units that dropped out of the sample are statistically equal to baseline characteristics of the units that were successfully resurveyed. As long as the baseline characteristics of both groups are not statistically different, our new sample should continue to represent the population of interest.

Second, we can check whether the attrition rate in the treatment group is similar to the attrition rate in the comparison group. If the attrition rates are significantly different, then there is a concern that our sample is no longer valid, and we may need to use various statistical techniques to try to correct this. One common method is inverse probability weighting, a method that statistically reweights the data (in this case, the follow-up data) so as to correct for the fact that a portion of the original respondents is missing. The method reweighs the follow-up sample, so it looks similar to the baseline sample.

On the other hand, non-statistical measures to minimize attrition can be applied as well**: prevention and persistence**. Some of the actions to prevent attrition are to employ various data collection strategies (e.g., in person, by phone); make shorter questionnaires which allow us to have a higher response rate; be strategic in deciding when to carry out the follow-up survey (e.g., no during holidays). The other measure is to decide early in the preparation stage of the evaluation if we will try to find all individuals who participated in the evaluation or a representative sample of those who dropped put; nonetheless, this strategy may represent higher data collection costs.

It is also important to collect data on the reasons of attrition, in order to better avoid it in the future.

2. Questionnaire and ODK tool – good practices

KoboToolBox allows for a lot of different functionalities (for example connecting the questionnaire to existing datasets). We recommend exploring and testing these different functionalities to make use of the tool in the most efficient way. This is an investment that will save much time in the future.

Variable names and labels

Variable names and labels are usually a personal preference, so there is no formal convention. Nonetheless, here are some guidelines: (i) all variables should have labels, and all multiple-choice variables have value labels; (ii) the labeling system should be internally consistent; and (iii) it should be easy to connect the variable in the dataset with the question on the questionnaire since most analysis should be done with the questionnaire in hand.

Variable options

For data analysis, the assignment of numeric values to alphanumeric data is very important. Therefore, in the elaboration of questionnaires in any data collection software tool, we recommend the assignment of numeric values to alphanumeric answers options. For example, while for a binary answer question related access to drinking water, 1 and 0 can be assigned to yes and no respectively; for a multiple-choice question related to level of education categories, 0, 1 …, 7 can be assigned to no education, primary school and doctorate degree respectively.

Another recommendation is to restrict the range of values in the questions. For example, if maximum possible value for *the share of production commercialized through the farmer organization* is 100%, restrict the answer value of that question to 100%. This is to avoid the computation of *mean* with values that overpass the maximum possible value.

Finally, for questions in which the respondents do not select or include an answer, we recommend assigning a dot "." which represents a missing value to the pre-determined answers *not answered "N/A"* when they are not answered. This can be done before applying the survey to avoid doing it in the data cleaning phase. If this is not possible to do in the survey tool before its application, the "N/A" can and should be replaced in the data cleaning phase by ".".

Unique identifiers

A unique identifier is an attribute (usually a number, code or piece of information) which is guaranteed to be unique among all of the used identifiers. In data collection practices, the identifier is linked to a specific individual. The importance of the use of unique and consistent identifiers across survey waves is that they allow to identify easily and follow the same individuals. This saves us the work of re-identifying individuals in each survey wave in which they participate.

Once the unique identifier is assigned to survey participants, it is important to collect the same personal information/credentials across survey waves. For example, to collect their age, gender, level of education, phone number, place of residence, and GPS location of farm (if needed) in all survey waves in which they participate. For survey-based data analysis, a unique and consistent identifier that has all these personal information elements for each survey participant allows to (i) easily identify the individual, (ii) avoid duplicate individuals, (iii) and observe changes across periods for the same respondent.

In line with this, we highly recommend making personal information or credentials as a required field to fill out in the surveys. Similarly, we recommend surveying the same person of the household so that any observed change can be directed to that person and not another person from the household.

Revision of questions

We recommend keeping variable names and its labels consistent across survey waves to avoid any identification problems in the data analysis. Moreover, it is indispensable to indicate when new questions are added or reformulated from survey to survey.

3. Data cleaning and analysis

Using STATA

To facilitate the cleaning and consolidation of survey-based datasets, the use of the user-friendly statistical software package STATA is an option. STATA allows to keep track of all changes made to the raw datasets in a file called "do-file" where all commands to execute any specific task for the cleaning or merging of the data can be stored. Similarly, the software package highly reduces issues of reliability of the data to be used or presented since all changes made to it can be tracked if stored in a do-file. Therefore, we highly recommend the use of STATA or any other statistical software package.

If Excel were to be used for data cleaning or merging, we strongly recommended to document and explain in a separate file (e.g., Word) all the tasks or changes made to the raw datasets and why they were made. In this sense, the changes made to the datasets can be traced back if needed. Nonetheless, to avoid all this manual process of looking for and eliminating values that may not make sense or copy-pasting values in excel, we suggest the use of STATA.

In terms of data analysis, we also recommend the use of STATA or other statistical software for the productionn of quantitative analysis, including descriptive, summary statistics, and econometric analysis, if working with an excel file obtained from a data collection software package (e.g., Kobo Toolbox). The (raw) excel file can be exported to STATA, and in there, we can select a subset of countries, commodities, and variables to work with. In STATA, we can also observe, modify and/or exclude any value from that data set or data analysis that is not consistent with the units or expected sizes of a variable. For example, outliers – data points that differ significantly from other observations – like very high values in the income distribution can be excluded from the analysis if spotted so that they do not produce abrupt changes to *mean* values.

From ODK to STATA

Data extraction of ODK software through Stata can be done by using the *Kobo2stata* package.