Tracking indirect beneficiaries of complex development interventions in agriculture

Julius Okello / CIP SSA

APRIL 2019
Tracking indirect beneficiaries of complex development interventions in agriculture

Julius Okello
International Potato Center
Kampala, Uganda

April 2019
Acknowledgements

This work was undertaken as part of the CGIAR Research Program on Roots, Tubers and Bananas (RTB). Funding support for this study was provided by the UK Department for International Development (DFID) under the Scaling Up Sweet Potato for Agriculture and Nutrition (SUSTAIN) project. The opinions expressed here belong to the authors, and do not necessarily reflect those of RTB, CGIAR, DFID or the International Potato Center (CIP).

SUSTAIN is a six-year partnership (2013–18), led by CIP, to scale up the nutrition benefits of biofortified orange-fleshed sweetpotato (OFSP). The goal is to reach 1.2 million households with children under-5 years of age in Bangladesh, Kenya, Malawi, Mozambique and Rwanda. SUSTAIN supports integrated interventions in agriculture, nutrition, utilization and marketing to strengthen production and consumption of OFSP.

The author also wishes to thank all the CIP colleagues and other CGIAR center staff for their valuable comments, suggestions and feedback on this document and for the information and literature that have enriched the current version, particularly Norman Kwiriza for his assistance in gathering the initial literature that inspired the paper.
1. Introduction

Development interventions usually aim to improve the welfare of a target population. So, interventions that address poverty will be interested in reaching vulnerable households, defined in terms of assets or income, while nutrition interventions might target a certain group of undernourished children or a certain group of the population that is likely to be at risk of malnutrition. These targeted groups, however, normally co-exist in their geographical locales with the non-target group that are not of particular interest to the development project or program. The targeted beneficiaries are often a subset of the local economy, which is often defined as a village, community, county, district, province, or region, depending on the context.

Development planners and intervention implementers often intend to target the intended beneficiaries as precisely as possible, usually with the goal of effectively utilizing the project budget to reach the highest possible number of such beneficiaries. However, nontarget populations may still benefit from such interventions in ways that were not planned by the project implementers. For instance, nutrition messages intended for caregivers of a certain age-group of children could reach the caregivers of children in other age groups and influence their behavior in ways that generate general public benefit. Another example would be when improved varieties (e.g. orange-fleshed sweetpotato (OFSP) vines, seed potato, iron fortified bean seed, and Xanthomonas wilt tolerant banana) intended for targeted households, such as poor households or those with children under 5 years of age or with a pregnant mother, reach nontarget households.

This literature review is intended to outline how to track the fate or diffusion of a development intervention among the beneficiary households not directly targeted by a development program and/or its partners. For the purposes of this review, we define a direct beneficiary as a unit (individual or household) that is directly reached by the project and/or its partners and that meets the eligibility criteria set by the project. Such criteria can be defined in terms of poverty score, wealth index, a specific age group of children, geography, and/or pregnancy status. Consequently, an indirect beneficiary is any household, whether eligible or not, that receives the intervention from sources other than the project and its partners, that is informally from the direct beneficiary through household-household relationships or social networks. Such sources may include family, friends, and/or group and religious organizations to which they or their family and friends belong.

The rest of this review is organized as follows: Section 2 takes a broader look at the issue of spillover effects of a development intervention and defines different types of spillovers, while Section 3 briefly outlines the quantitative approaches used in tracking spillover effects. Section 4 presents some of methods that have been used to empirically track the indirect beneficiaries in the context where the population is not hidden, but the diffusion path of the intervention is unknown, such as most of the development projects implemented by the CGIAR centers. Section 5 then discusses practical guidelines on how to track specific types of beneficiaries in the context of the SUSTAIN project, whose implementation is currently underway in two eastern (Kenya and Rwanda) and two southern (Malawi and Mozambique) Africa countries. Section 6 concludes.
2. Definition of spillover effects

Indirect beneficiaries of a development program belong to a group of the population that receives benefits of a project or program in ways that were not initially intended by the project planners and implementers (Sveiby et al. 2009). Such benefits are often referred to in the development literature as spillover benefits/effects. In the example given earlier, poor households, i.e., the target group within a local economy, may be targeted by an intervention to remedy the situation. However, in the process of doing so, nonpoor households may also be indirectly affected through their social or economic interactions with those that are targeted, i.e., the treatment group. A spillover effect/benefit may therefore be defined as the spontaneous spread of an intervention, usually a technology such as improved seed/vine, through farmer-farmer or household’s social network in the absence of project/partner mediation (German et al. 2006).

Angelucci and Di Maro (2010) identify four broad types of spillover effects of a development intervention. These are: i) externalities, ii) general equilibrium effects iii) interactions, and iv) behavioral effects.

*Externalities* operate from treated to the untreated group/population, and are most common in a health domain. For example, participants in a school-children-deworming campaign not only benefit from being worm-free but they also confer benefits to non-participants by reducing contagious infections, and, again, to themselves from reduced re-infections (Miguel and Kremer 2004).

*General equilibrium* effects occur when a localized intervention benefits the entire population. In the context of agricultural intervention, these can occur when an intervention causes system-wide effects (Angelucci and Di Georgi 2009). For instance, an intervention that promotes the benefits of an improved variety, such as a biofortified crop, in a target population of a local economy can result in a general increase in the price of such crop to the entire population of the local economy. The increase in price of the crop arises from a surge in demand and affects the non-target farm households or individuals negatively who are often net buyers (Jayne et al. 2001).

*Interactions* occurs through social and economic networks. The treatment household or individual, through social and economic interactions, transfers some or all of the benefits of the intervention to nontarget households or individuals (Adamo 2001).

Lastly, *behavioral* effects can occur when an intervention affects the behavior or social norms in the locality in which it occurs. Avetabile and Di Mario (2009) discuss this kind of spillover effect in Mexico where nonparticipants of a project changed their cultural norms to let participating spouses comply with project conditionality.

These definitions are rather arbitrary but nonetheless emphasize that spillover effects can arise from several sources and influence nontarget populations in various ways. In agricultural contexts, and especially where adoption of a biofortified crop is promoted using nutrition-focused education, these effects can occur through interaction via social networks and behavioral influences (German et al. 2006). Moreover, they can result in systematic changes in the local economy through price effects or through external effects.

Tracking the households that are affected by these changes, and the ways in which they are affected, whether negatively, as in the case of price increases, or positively, through consumption of better/nutritious foods, is usually of major interest to project planners, project implementers and donors. Such households, due to the way in which they receive benefits from the development programs, are sometimes referred to in the literature as spillover beneficiaries.
3. Quantitative methods of measuring spillovers

Most development interventions start out with a very clear picture of the target beneficiaries. The intervention is therefore often directed to known intervention units for which the population is well defined. For example, an intervention targeting smallholder farmers with improved seed, e.g. of a biofortified crop, usually has all smallholder farmers as its target population. Similarly, agriculture-based nutrition projects such as the Scaling up Sweetpotato through Agriculture and Nutrition project (SUSTAIN) that target households with children under 5 years of age or pregnant women have all households with such members as the target population. This implies that direct beneficiaries are observable. Complications in tracking the fate of an intervention, however, start beyond this observable direct group of beneficiaries.

As the direct beneficiaries interact with the intervention and with nontarget population, it gets passed-on to the non-target population and/or other eligible members of the local economy hitherto unreached by the project or its partners. This passing-on of the project benefits can occur through voluntary and non-voluntary processes. It is not difficult to see that this process can repeat itself several times, in each case creating successive waves of beneficiaries. In the context of agriculture, the pace and rhythm with which the intervention spread depends on many factors, but key among them is the seasonality of farming. Specifically, for an agricultural technology, such as a biofortified seed, one can expect a new wave of beneficiaries to emerge every season or even more often\(^1\).

Although of much interest to project planners and implementers, development literature does not focus much on tracking the waves of project beneficiaries resulting from successive use and experimentation of an intervention. Instead, the literature currently has two main strands relating to measurement of unknown populations.

The first strand focuses on measuring the effects of these successive waves on the households or individuals that are not reached directly by a project and/or its partners. Specifically, this strand of the literature focuses on measuring the outcomes of an interaction on households that are not reached by the project with the technology. This literature deals extensively with the estimation of the effects and/or impacts of an intervention on some outcome variable of interest, usually using quantitative techniques. Some of the literature in this area include Caliendo and Kopeining (2008), Angelucci and Di Mario (2010), Ogutu et al. (2014), and Carrion-Yaguana et al. (2015).

The second strand focuses on sampling populations that are hidden or hard-to-reach, such as drug users, people living with HIV-AIDS or same-sex couples, for whom there is usually social stigma that prevents self-revelation (see for instance Salganik and Heckathorn 2004; Volz and Heckathorn 2008; Heckathorn 2011; Salganik, 2012).

Below, we briefly discuss these two approaches before turning to the issue of tracking beneficiaries in a development intervention in which the first set of direct beneficiaries is observable and hence there is a known population.

---

\(^1\) Under irrigation agriculture, new waves of beneficiaries can emerge much sooner resulting in faster diffusion of the intervention.
3.1 Measurement of spillover effects

The literature discusses two broad methods for measuring spillover effects of a development intervention: experimental and non-experimental techniques. Both methods rely heavily on quantitative measurements and sampling of the treatment and non-treated units is key.

In the former, a complex study design is laid out prior to the rollout of the intervention. The study areas are delineated and stratified into treatment and control groups, and the eligible units (households or individuals, depending on the nature of the intervention) in each of the treatment regimes (i.e. treated and nontreated) is systematically listed. This list of all the eligible treatment units provides the population of the treated units, i.e. the units that get to receive the intervention, from which the spillover effects can be measured. Obtaining a spillover effect requires that units (households or individuals) in the treatment area who do not directly received the treatment from the program and/or partners, but do so from other sources, are included among the treated. Similarly, residents of the treatment area who miss the treatment (non-treated) are excluded from the treatment group.

The non-experimental techniques differ from the experimental due to the lack of a predesigned control group. The analyst may, in this case, have information only on the units (i.e. households or individuals) that were reached by the intervention. The analyst must therefore rely on a carefully defined comparison group to assess the spillover effect of the development intervention. The comparison group can be obtained by interviewing a sample of units that live in the study area or in its neighborhood but were not directly reached by the intervention.

What is not apparent in the above discussion on the measurement of spillover effects is the process by which the treated and non-treated units who are compared are selected for interviews and eventual measurement of outcome variables. In both the experimental and non-experimental methods, the survey respondents are usually obtained through some form of random sampling. The decision to use random sampling is normally guided by the desire to work with a representative sample of the treated and nontreated households, which is, in turn, aimed at enabling the analyst to make inferences about the population from which the sample is drawn.

3.2 Tracking the hidden populations

There is an active and rapidly growing literature that focuses on how to sample and track hidden or hard-to-reach populations. This literature is motivated by the difficulty of compiling a credible population of a group of individuals or households that engage in activities that most communities consider socially undesirable. The literature discusses at least three broad methods of tracking such populations namely, snowball, respondent-driven, and random walk sampling techniques.

In all the three cases, a group of individuals belonging to the hidden population is interviewed and these individuals then identify fellow members of the hidden population, who are in turn interviewed. The process continues through a series of waves of interviews until the analyst reaches a desired sample size. The difference in these approaches is with regard to the representativeness of the attained sample. Indeed, an active debate in the literature revolves around whether or not the sample obtained using each of the methods can be used to make inferences about the entire hidden population (Goodman 2011; Heckathorn 2011; Heimer 2005).

---

2 This narrative assumes a simple treatment defined as receiving an intervention such as a biofortified seed. More complex treatments usually require splitting a composite treatment into several parts that enables the analyst to assess the effect of each treatment part on the treatment units.
Early usage of snowball sampling technique is associated with the work of Colman (1958) and Colman et al. (1958). It has more recently been used to sample role models of prominent investors by Barton (2001). The technique is sometimes referred to as chain-referral, link-tracing or network sampling. It was designed to study the social structure, network environments and/or network of friendships using a non-probabilistic sampling procedure. The findings arising from a pure snowball sample is non-representative and cannot be used to make statistical inferences about the population. Goodman (1961), however, developed a variant of the snowball technique that allows a statistically representative sample to be collected, but under stringent conditions. The technique requires that the initial sample is drawn “using a probability method on a known sample” (Handcock and Gile 2011). It further fixes the number of social links (i.e. people) followed from each participant and the number of waves of interviews that must be conducted.

Respondent-driven sampling (RDS) was developed by Heckathorn (1997) and has since been used widely to sample hidden populations that are HIV-AIDS infected and those involved in drug use (He et al. 2008; Ruan et al 2008; Rudolf et al. 2011). It is capable of producing unbiased samples that enables the analyst to draw statistically valid inferences about the population based on the sample (Aldana et al. 2008; Heckathorn 2011). Much of the usage of this technique has been focused on the study of aggregate characteristics of the issue under study, such as disease prevalence. Recent developments of the technique, however, now allows for the estimation of the hidden population (Handcock et al. 2014).

To implement RDS, the analyst selects an initial number of individuals of interest, known as seeds, such as drug users or members of a proscribed group, based on pre-existing contacts. These seeds form wave 0. They are interviewed and earn a participation payment for completing an interview. Each seed is then given $n$ unique vouchers that they can use to recruit their peers to the study. The recruiter receives bonus payment for each voucher used to successfully recruit an individual to participate in the study. The recruits of people in wave 0 become wave 1. Upon completion of screening and interview, each wave 1 recruit receives $n$ unique vouchers that are similarly used to recruit others. This process continues until the right number of respondents is attained. The choice of $n$ depends on the behavior of the recruits and the ease of recruitment, which, in turn, depend on the nature of activity the hidden population engages in and the intensity of social stigma associated with being known or associated with the activity. The major weakness of this technique is that recruiters can introduce unknown biases to the sample if they fail to recruit randomly from among their contacts (Salganik and Heckathorn 2004).

The last type of sampling for hidden population is the random walk technique, one of the chain-referral methods of sampling, attributed to Klovdahl (1989). In this technique, the respondents are asked to list all the people they know that belong to the population under study. This process generates a sampling frame. When the sampling frame is drawn, the analyst takes a random sample from it for the new wave of interviews. The process continues until the planned number of respondents is interviewed.
4. Tracking known populations with unknown diffusion path

In many of the development interventions, the eligibility criteria set by the project, as part of the development challenge to be addressed, enables the analyst to clearly delineate the population of direct beneficiaries in a local economy. As in the earlier example, an intervention addressing vitamin A deficiency will target a population comprising children under the age of 5 years, pregnant women or lactating mothers. These are the individuals whose need for vitamin A is elevated and that are at risk of suffering from its deficiency in the local economy. However, development projects could also be interested in addressing other development challenges such as low income, asset poverty, HIV-AIDS incidence, poor access to a technology (e.g. quality seed), among others. In all these cases, the project design will usually define the eligibility criteria for targeting beneficiaries, and the criteria allows the analyst to clearly define the population of direct beneficiaries.

The difficulty that typically arises is how to trace and record information about (or just the simple count of) households or individuals that benefit indirectly from the project, that is, households or individuals that receive the intervention without directly interacting with the project or its partners. There are different paths through which the intervention can diffuse to such indirect beneficiaries, but most of them occur through social networks including family, religious organizations and farmer and other community-based organizations.

There are two approaches that can be used to track the spillovers (indirect beneficiaries) in a project where the direct beneficiaries are known. These are i) tracking surveys and ii) snowball surveys. Below is a brief discussion of both approaches.

4.1. Tracking surveys

These surveys involve interviews with a representative sample of households or individuals (German et al. 2006). The respondents are selected randomly from a population containing i) direct beneficiaries who have been reached by the project or its partners, ii) indirect beneficiaries who benefitted through other channels, and iii) non-beneficiaries.

The survey is preceded by the generation of a sampling frame which is comprised of a complete listing of all the direct beneficiaries, and a complete listing of all the individuals or households that could potentially benefit from the intervention but were not directly reached by the project and/or its partners. They may not have been reached by the project for various reasons: they are not part of the targeted group, that is, they do not meet the eligibility criteria set by the project; resource constraints prevented the project from reaching them; or other barriers prevented them from being directly reached by the project. In other words, the analyst needs a stratified sampling frame that comprises direct and indirect beneficiaries and non-beneficiaries.

The list of direct beneficiaries is by far the easiest to obtain. Most project implementers and their partners will record the names of households or individuals they give the intervention to as part of the routine progress monitoring. The project leaders or their M&E officers and their partners are, therefore, the ideal source of this list. This list will often have all the necessary identifiers, such as administrative units (district, village, etc.) as well as gender of the household head or the individual.

Obtaining the list of indirect beneficiaries and non-beneficiaries is, however, usually more complicated. Since project implementers and partners usually lack complete knowledge of this group of the population, the list has to be drawn with the help of local leaders (administrative, group, opinion, etc.). Determining who should
be included in the list depends on the nature of project. For a seed technology project that is promoting the growing of a new variety of a crop, the candidate members of the list will be farmers who normally grow the crop. Hence, with the help of local leaders, all farm households or farmers that grow the crop are listed. As expected, this list will have farmers who have received the intervention and those who have not, and the task is to identify who the beneficiaries are and whether they are direct or indirect. The tracking survey can yield information on both the number of indirect beneficiaries and the structure of the network of beneficiaries, i.e. the location of beneficiary in the network of beneficiaries, as shown below.

![Diagram](image)

**Figure 1: Diffusion path of an intervention showing waves of beneficiaries**

The data collected can allow the analyst to tell how many respondents received the intervention from the direct beneficiaries, and also to learn about the nature of diffusion of the technology from first, second, third, and subsequent waves of indirect beneficiaries. One can therefore track the proportion of the project beneficiaries at the different levels of the diffusion stages. The other advantage of this kind of survey is that the analyst is able to track and compare some of the characteristics of the beneficiaries and the non-beneficiaries. Given a good sample size, and if sufficient number of control variables are collected, tracking studies can be used to assess quantitatively the drivers of decision to adopt the technology, especially among the indirect beneficiaries and non-beneficiaries since the latter get to receive the technology without the undue pressure from the project. A major downside of the tracking surveys is that they usually need a relatively large sample size, which will increase the costs: the time and financial costs of generating the sampling frame for the different waves of the indirect beneficiaries can be high.

### 4.2 Snowball sampling approach

This approach uses an adapted form of the snowball sampling technique. It differs from the actual snowball sampling procedure in that the population or sampling frame from which the initial sample of respondents (i.e. the direct beneficiaries (B₀)) is drawn is known and hence the sampling probability can be calculated with certainty. However, the subsequent sampling frames are unknown, and, therefore, the compilation of the next sampling frame (B₁) depends on the respondents drawn from initial population B₀. Similarly, the sampling frame B₂ depends on the respondents based on a sample from the population of B₁ households.

German et al (2006) used a variant of this method to track the social structure of technology sharing and the rate at which the technology flows in the community. They first interviewed 20% of the B₀ sample using a systematic random sampling procedure: every fifth direct beneficiary (B₀) was selected for interview until the pre-compiled list of direct beneficiaries was exhausted. Each of the B₀ respondents listed all the individuals...

---

3 It is, of course, the case that this kind of list will exclude those farmers who would have only just started growing the crop but are still unknown or who intend to grow it later during the lifetime of the project.
they shared the technology with, thus generating the \( B_1 \) sampling frame. This means that the \( B_1 \) sampling frame is based on the systematically randomly sampled respondents drawn from the direct beneficiaries. Not all the direct beneficiaries were interviewed in the wave 1 interview, hence, not all the indirect beneficiaries had the same probability of being selected or participating in the study. Specifically, some of the (\( B_1 \)) indirect beneficiaries who received the interventions from non-sampled individuals in \( B_0 \) have zero probability of being selected which implies that the \( B_1 \) sample is non-representative. The non-representativeness of the sample drawn from the \( B_1 \) sampling frame implies that the sample cannot be used to make inferences about the entire population of the indirect beneficiaries in wave \( B_1 \).

4.3. Empirical approaches for tracking indirect beneficiaries in the CGIAR centers

Different CGIAR projects and organizations are using different approaches to track indirect beneficiaries of their interventions. Here we document some of the known cases. They include: the survey, pass-on, and market network analysis approaches.

4.3.1 Survey approach

In general, the method used to track indirect beneficiaries under this approach is to include a series of questions that assess the sources of the intervention in a larger household survey module. Specifically, questions are included in the survey that explore whether the respondents have applied or used the intervention, and if so, whether they are indirect beneficiaries, i.e. received the intervention from sources other than the project, the year (and season) the intervention was received, and in some cases the relationship to the source. The survey respondents are therefore able to declare whether they are direct or indirect beneficiaries. Further probing questions, such as the relationship to the intervention source and year or season the intervention was received, can provide information on the structure of the network and the number of waves.

4.3.2 The pass-on approach

The pass-on approach seeks to track the indirect beneficiaries by asking farmers to record individuals or households in their social network with whom they share the intervention. In some projects, direct beneficiaries are obligated to pass-on the intervention to other farm households or farmers, as a way of expediting the diffusion of the intervention. A good example of this approach was in the Dissemination of New Agricultural Technology in Africa (DONATA\(^4\)) project implemented by CIP. The project required all the direct beneficiaries to pass-on vines of the OFSP to a specified number of households or farmers in their networks. The project kept a complete record of the beneficiaries reached directly, but did not require these beneficiaries to keep a record of those with whom they shared the intervention. In this case, therefore, the first wave of indirect beneficiaries can only be obtained by asking the direct beneficiaries to name these households (Kimeny and McEwan 2014). Indeed, this approach of ex-post listing of indirect beneficiaries was recently used by CIP’s Rwanda Superfoods project during its endline survey in 2014 (Sindi et al. 2015).

The Rwanda Superfoods project had a component that focused on organizing farmers into farmer groups and linking them to a sweetpotato root processor. The farmers were mobilized into groups by a partner, the Young Women Christian Association (YWCA). Indirect beneficiaries were defined as all the farmers who received vines without interacting directly with the project staff or partner, that is, farmers who received vines from

\(^4\) https://www.asareca.org/sites/default/files/publications/DONATA%20BOOK%20FINAL%20PDF%20%28for%20web%29_0.pdf
members of the farmer groups. In order to include these spillover farmers in the survey, the survey planners asked all the group leaders to compile a list of farmers who received vines indirectly through interaction with group members. The lists of spillover farmers from each of the groups in the study locations were then used to generate a sample from which spillover respondents were randomly selected, using probability proportionate to size sampling technique, until a desired quota of the spillover farmers in the overall survey was reached.

A more recent example of this pass-on approach used by HarvestPlus asked farmers not only to share the intervention (i.e. biofortified seeds = vines, beans) with other households or farmers but also mandated them to keep a record of farmers to whom they pass the intervention. The records are collected regularly, validated as part of the routine monitoring, and kept by the project office. This pass-on system has the advantage that the pass-on list acts as a good source of the information needed to generate the sampling frame for the first wave of indirect beneficiaries and is expected to generate a more reliable sample of these beneficiaries than the case where recall is used to generate the list of indirect beneficiaries and hence the sampling frame (Asare-Marfo et al. 2016).

**4.3.3 Market network approach**

The market network approach\(^5\) works from the upstream end of the value chain. It starts with traders of the commodity and attempts to work its way back to the farmers. An example of this approach was used by the DONATA project. The project approached sweetpotato traders in urban markets and obtained the names and contacts of their suppliers in the villages. As expected, the contact list was comprised of rural assemblers, large farmers, and rural traders. Using these contacts, the project visited the rural assemblers and rural traders and compiled lists of their farmer suppliers: when combined with the list of large farmers obtained from the urban traders, this comprises the list of direct and indirect beneficiaries of the DONATA project who participated in this value chain. With information from the project implementers, one could then disaggregate the direct and indirect beneficiaries of the project involved in this value chain.

This approach has two major disadvantages. First, the list of indirect beneficiaries obtained captures only farmers who participate in the sweetpotato root market. Second, because the list is based on actors’ recall, the majority of the farmers identified, and hence the indirect beneficiaries, are likely to be those that supply big volumes or who are major and long-term trading partners. Small, one-time suppliers are likely to be left out of the lists compiled by the actors. This implies that the list of indirect beneficiaries obtained from this approach may not be representative of the true population of indirect beneficiaries involved in this value chain.

**4.3.4: Interviewing the whole population in each wave**

One way of dealing with the problem of being unable to estimate the actual population of indirect beneficiaries by extrapolating from a non-representative or imperfect sampling frame of first and later waves of beneficiaries is to interview the entire populations of indirect beneficiaries at each wave. This approach was used recently by the International Maize and Wheat Improvement Center (CIMMYT) to estimate the number of indirect beneficiaries in its Sustainable Intensification of Maize-Legume Systems for Food Security in Eastern and Southern Africa (SIMLESA) project. The project worked with farmers organized into demonstration farmer groups, henceforth referred to as a demo groups, of around 30 farmers each.

---

\(^5\) Based on personal communication with Sarah Mayanja, CIP, Uganda.
The demo groups hosted an experiment that exposed them to SIMLESA technologies. To implement their snowball sampling approach, the list of all the farmers participating in the demo group (B0) was drawn up by the survey planners. All these B0 farmers in each demo group were interviewed and used to compile the list of farmers who obtained SIMLESA technologies from them, called technology users, or became aware of the SIMLESA technologies from them. This list of technology users formed the first wave of indirect beneficiaries (B1). As in the first case of B0, all the B1 farmers were also interviewed and used to generate the list of the second wave of indirect beneficiaries (B2), all of whom were similarly interviewed. In each wave the survey planners examined the list and discarded any duplicates although it is unclear how farmers who may have benefitted indirectly more than once from different neighbors in earlier waves were treated. The survey planners treated B2 as the last wave of indirect beneficiaries to report in the SIMLESA project because it had a 3-year lifespan.

While this approach overcomes the difficulties of extrapolation based on imperfect samples, it has one main disadvantage. It can only work where the project intervention targeted a small group of farmers. When the B0 farmers are many thousands, as is common in some agricultural projects, the cost of implementing the approach can be prohibitively high (Singla et al. 2014).

### 4.3.5 Social network analysis

This approach differs significantly from the first four above, and is mainly used to study the spread of a technology within a social network in communities. In this approach, the direct (B0) beneficiaries list the names of other actors that they passed the technology to or, in cases of awareness, those with whom they discussed the technology. The researchers then inspect the lists and noted the names that are most frequently mentioned, referred to as nodal farmers. The names of nodal farmers are then listed in order of frequency of mention and interviewers select respondents down the list until the desired quota is attained (Wasserman and Faust 1994).

Different variations of this approach exist. However, the approach relies on social network analysis rather than statistical estimation techniques to examine the diffusion of the technology within a closely-knit social network.

The CG Centers that have or are using this approach to study diffusion of technologies include CIAT, Bioversity, IITA, ICRISAT, and Africa Rice. IITA/CIAT applied the method in assessing the diffusion of technologies promoted by the Consortium for Improving Agriculture-based Livelihoods in Central Africa (CIALCA) project in DRC Congo (Hermans et al. 2017). In none of these cases was the goal to generate an estimate of the number of indirect beneficiaries.
5. A method for tracking indirect beneficiaries: applied to the SUSTAIN project

The SUSTAIN project is a five-year intervention being implemented by CIP in Kenya, Rwanda, Malawi, and Mozambique. The project has the goal of reaching 1.2 million farm household beneficiaries and reaching 400,000 rural and urban consumers over a period of five years, starting 2014. Its intervention is comprised of three components: agriculture, the dissemination of improved orange-fleshed sweetpotato (OFSP) vines combined with some agronomic information; nutrition training, education and counseling; and market, the manufacture and sale of commercial products from OFSP and farmer linkage to markets and processors. The intervention occurs through a wide array of partners. The project targets households with members that are most vulnerable to vitamin A deficiency, i.e. pregnant and lactating mothers, and children less than 5 years of age.

Among the 1.2 million beneficiary households, 200,000 are to be reached directly through project or partner interactions while one million are to be reached indirectly. At the same time, 200,000 of the direct beneficiary households are to be reached with nutrition and diversified use training. The project planned to measure and report on the number of these indirect beneficiaries on an annual basis. This section outlines an approach that can be used to track the indirect beneficiaries on an annual basis through interviews with samples of beneficiaries drawn each year.

The method outlined here is a variant of the snowball sampling technique developed by Colman (1958) and Colman et al. (1958) to study populations for which a sampling frame does not exist and are hard to reach. It extends the approach applied by German et al. (2006) to study the diffusion of technologies promoted under the African Highlands Initiative project by allowing for expansion of the different waves of technology diffusion. It differs from the approach used by CIMMYT recently to estimate indirect beneficiaries in the SIMLESA project, which interviewed entire populations, by relying on samples drawn from the different waves of indirect beneficiaries. The method also differs from the actual snowball technique because the initial sampling frame, which is comprised of the direct beneficiaries, is not hidden. Indeed, it is the population of all the households that received vines in earlier distributions because they met the SUSTAIN project eligibility criteria: having a child under the age of 2 years; having a child 2-5 years of age; and/or having a pregnant woman. We refer to this population as the wave zero beneficiaries (B₀). Since the implementation of the project follows planting seasons, this population should ideally be the same as the population of direct beneficiaries in season 1 and season 2 of year t (for t = 1, 2, 3, 4, 5) of the project implementation for countries with two planting seasons in a year, or just season 1 if there a single planting season.

As expected, some households who receive the intervention in season 1 of the first year of project implementation (in a two-season country) are likely to share it with their network of family and friends during season 2, thus giving rise to the first wave of indirect beneficiaries in season 2. This group of indirect beneficiaries are referred to as population B₁. During this second season, however, a new cohort of direct beneficiaries join the first cohort, but they too start sharing vines in season 3 (i.e. year 2, season 1), because of the lag in sharing of the intervention. This implies that in season three there are three groups of indirect beneficiaries: the first is comprised of households which receive vines from indirect beneficiaries, i.e. B₁

---

6 It is assumed that direct beneficiaries start sharing the intervention after their first season of growing the project vines. Hence, there is no sharing of vines by the direct beneficiaries in the season in which they received the vines from the project/sources, implying zero indirect beneficiaries in season 1 of year 1.
households. These households constitute the second wave of indirect beneficiaries, hence, using the same notation as above, are referred to as B2 households. The second are households which received the intervention from the first cohort of the direct beneficiaries and are therefore wave 1 population, but in year 2 hence denoted (B1). The last are households that receive the intervention from the second cohort of the direct beneficiaries, hence also wave 1 (B1) indirect beneficiaries, and so on for the subsequent years. As contained in the SUSTAIN logframe, the intervention is defined as the provision of improved OFSP planting materials (vines) and some agronomic information by the project and/or its partners.

Since the project and/or its partners has the records of the direct beneficiaries (B0) in each year, B0 forms the sampling frame from which a random sample (b0) used to study B1 will be drawn. Note that B0 is comprised of a listing of all households that the project provided the vines to within season 1 of year t. It is this random sample (b0) of B0 households that we use to get information about the indirect beneficiaries (B1) including the number and the characteristics of this network of beneficiaries, such as gender and age.

The random sample b0 will help list the households they shared the OFSP vines from the SUSTAIN project with, and also the gender of the head of that household as well as whether the beneficiary household has a child under 2 years of age, a child 2-5 years of age, and/or a pregnant mother. If the sampling frame B0 is truly representative of the direct beneficiaries, that is the list is complete and contains all the indirect beneficiaries reached via all the vine dissemination methods, and the sample proportion is known, then an estimate of the total number of wave 1 (B1) beneficiaries can be computed by simple extrapolation from the sample.

In year 2, it is recommended that the list of indirect beneficiaries be drawn from both cohort 1 and cohort 2 sources, described above. The cohort 1 source will give rise to both wave 1 and wave 2 indirect beneficiaries while cohort 2 gives rise to only wave 1 of indirect beneficiaries, due to the assumed lag in sharing of vines. However, since the random sample (b1) drawn from wave 1 indirect beneficiaries is a sample of a sample, and all the actual wave 1 indirect beneficiaries do not have the same probability of participating in the survey, it is no longer representative. Nonetheless, using the wave 0 (direct beneficiaries) population from cohort 1 and 2 sources in year 2, we will be able to estimate the number of wave 1 households that indirectly benefitted from the intervention arising directly from the two cohorts in year 2, if the sample representativeness is achieved. Achieving the sample representativeness will depend on the extent to which efforts are made to ensure that the direct beneficiaries, and their identifiers, in both the mass distribution and vine multiplier methods of vine distribution are carefully recorded. This process of sampling can be similarly repeated for year 3, 4, and 5 (See Figure 1).

Where more than one season of intervention has passed, there is always be the possibility that more than one beneficiary will share the intervention with the same household. This might happen when a household receives the intervention from another in season y, but is unable to maintain its own vines for the next season. The recipient can thus return to the same donor or go to a different donor for another round of vines, and thus could be listed more than once in the B1 population. Similarly, a direct beneficiary household in population B0 (a donor) can turn to a household in B1 (recipient) for vines, if it loses its crop in season 2 or a later season, and may thus become listed in population B1 as an indirect beneficiary. To avoid this double-counting, the household lists compiled should be carefully examined, and any household that is listed more than once

---

7 For now, we abstract from these reciprocal transactions so as to simplify the approach, but the project implementers can plan for and capture these transactions as part of the M&E
dropped prior to sampling. The same procedure should be followed when preparing the sampling frame based on subsequent waves of direct and indirect beneficiaries.

So, how do we obtain the random sample to be used at the end of year 1 of project implementation? We plan to use a combination of the two-stage cluster sampling technique and an adapted snowball sampling procedure. First, the lists of all the administrative units (clusters) within which the intervention occurred will be drawn, and clusters randomly selected using the probability proportionate to size (PPS) sampling technique. Next, in year 1, the list of all households that received vines from the project or its partners (the direct beneficiaries) in each of the selected administrative unit will be drawn and a random sample of the households (second cluster) drawn. The household member who managed the OFSP plot will be interviewed.

Using the case of Kenya, as an example, all the community units (CUs) in all the counties reached were listed and random sample of the CUs drawn. At this point, PPS sampling will be used in order to allow counties that have many CUs to have more CUs sampled from them, and vice versa.

In each of the selected CU, all the villages will be listed and random sample of villages drawn, again using PPS, so that CUs with many villages have more villages drawn from them, and vice versa. Lastly, all the direct beneficiaries that received vines, during the year in focus, from the project are listed and the list validated, especially to remove duplicate households. Wave B0 respondents will then be randomly selected from the validated lists using PPS. The sampling of households can be very easily done using the free online software, the Research Randomizer\(^8\) or other known tools. This process will be repeated in subsequent years, with adjustments based on how a beneficiary is defined. Specifically, from year 2 onwards, sampling will focus on both direct and indirect beneficiaries, with the latter applying the adapted snow-ball sampling approach described above. The indirect beneficiaries will be sampled from a list generated by the b0 sample, and is, hence, a snowball sample.

The size of the sample of direct beneficiaries to start off with will depend on the country and budget. German et al. (2006) drew 20% of the population of direct beneficiaries in order to study the extent to which direct beneficiaries passed on the technology to their network of family and friends. However, the decision on the exact number of direct beneficiaries to start with will need to be made by the country team leader and M&E officer. There are, however, are some useful points to consider while making this decision. First, the number of respondents interviewed each year will grow with the passing of time, if repeat interviews are conducted on previous samples (in addition to the new waves) in subsequent years. This growth is in line with the diffusion theory which posits that technology spread starts slowly but picks up pace as households learn about and observe it (Rogers 1976; Karahana et al. 1999). This means that one should start with a sample that will still be manageable after four rounds of surveys, as in the case of SUSTAIN. It is especially important to plan for the compilation of the snowball sampling frame for the indirect beneficiaries.

Second, it is important to aim for a minimum sample size that will still allow for the harnessing of the benefits of large sample properties, such as estimation and inference (Wooldridge 2012). Third, the data to be collected in this exercise can be limited to a one-page questionnaire and can, where possible, be done using phone interviews thus significantly reducing the costs. Where phone contacts of the sampled households are available, phone interviews can also gather the needed information at even lower costs.

Below is a suggested draft protocol for use during the interviews. The protocol is designed to allow the analyst to carry out the listing of the successive waves of indirect beneficiaries, and contains a few select identifiers

\(^8\) https://www.randomizer.org/
and information for tracking a few of the indicators in which the project may be interested. It is intended to facilitate easy and low-cost data capture, but can be expanded based on need. If this exercise is done as a stand-alone, it needs to be supplemented with tables that collect limited socioeconomic data that can be used to characterize the households. The amount and type of this supplemental data will depend on the country/context and the types of relationships the country leader wants to learn about.

Table 1: Suggested protocol for tracking indirect beneficiaries for follow up interviews, giving an example of how to complete it

<table>
<thead>
<tr>
<th>1 Name of indirect beneficiary (in full)</th>
<th>2 Season of pass on</th>
<th>3 Eligibility criteria met (codes)</th>
<th>4 OFSP variety passed (codes)</th>
<th>5 How vines were given (codes)</th>
<th>6 Gender of recipient 1=Male 0=Female</th>
<th>7 Relationship to recipient</th>
<th>8 Other pass on (codes)</th>
<th>9 Phone contact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**Eligibility criteria met:** 1=Child < 2 yrs; 2=Child 2-5 years; 3=Pregnant woman; 4=Not eligible 77=Don’t know

**OFSP vine passed on:** 1=Kabode; 2=VITAA; 3= Kabode and VITAA 77=Don’t know/can’t remember

**Other pass on:** 1=Nutrition 2=Child feeding 3=How to grow OFSP 4=Market 5=Value addition 5=Other..........................

**How vines were given:** 1=Bought; 2=Free/gift; 3=Exchange; 4= =Other....

**Relationship:** 1=Parents; 2=Sibling 3=Son/daughter; 4=Other family member; 5=Male neighbor; 6=Female neighbor; 7= Other.........
6. Concluding remarks

An increasing number of donor funded interventions are expecting project implementers to report not only the number of households or individuals reached directly by the project and/or its partners but also those who receive the intervention through indirect sources, known as indirect beneficiaries or spillovers. The development literature however, does not focus much on this issue. Instead, much attention is focused on the approaches for measuring the effects of interventions, with some examining the spillover effects of development interventions. Specifically, there are two strands of the literature on measurement of unknown populations. The first focuses on the quantitative methods for measuring spillover effects of an intervention and treats spillovers purely as influencers of specific outcome variables targeted by the project for change. The second and more active strand of literature is more specific to measuring the structure and size of hidden populations, and focuses heavily on sampling techniques that can allow an analyst to learn about these populations.

One aspect of the beneficiaries in agricultural interventions is that the first wave (direct) beneficiaries are often known because they can be well delineated using the eligibility criteria set by the project planners. However, the indirect beneficiaries are usually unknown. This implies that methods used to track hidden populations do not quite apply in the analysis of indirect beneficiaries in agriculture. At the same time, little is currently known about how best to capture the complexity of social interactions between an intervention and the target population, and the nature of reciprocal transactions that are likely to occur between and among different categories of beneficiaries. That is, the structure of diffusion of an agricultural intervention and how to measure it remains a challenge.

This brief provides a technique for measuring the extent of diffusion of an agricultural technique. The method proposed, however, focuses on the measurement of the initial wave of indirect beneficiaries for which a representative random sample of the development units can be drawn and used to make inferences about the population, and estimate the population size. Samples of additional waves of indirect beneficiaries are non-representative and can at best be used to study the structure of social networks of the diffusion of the agricultural intervention, but are not useful in estimating population sizes. The limitations of the methods presented in this brief therefore suggest the need for further studies to develop more credible techniques of tracking populations of indirect beneficiaries in complex agricultural interventions, such as those that combine nutrition and agriculture components.
References


Tracking indirect beneficiaries of complex development interventions in agriculture

Julius Okello / CIP SSA

APRIL 2019