COSA

Reflections from the Field:

Notes on a Small-Sample Size Contribution Analysis

Abstract

Evaluative research on the effects of agricultural interventions frequently encounters small sample size (small n) situations where there are only a limited number of “units of assignment” for analyzing impacts. In these cases, the sample size is too small to allow statistical inference. Unlike experiments that involve a large sample size (large n), there is little consensus among impact evaluators on how to assess attribution. This paper sets out to provide some insight into the practicalities and value of Contribution Analysis for small sample size impact assessments by drawing on lessons learned and challenges faced during a real-world evaluation. Utilizing mixed methods, the Committee on Sustainability Assessment (COSA) has incorporated Contribution Analysis into an approach that minimizes the risk of potential sources of bias.

By Keith Child and Mica Bennett

Introduction

For over a decade, the Committee on Sustainability Assessment (COSA)[1] has conducted experimental and quasi-experimental impact assessments in countries around the world. COSA continuously endeavors to improve processes and to employ the most rigorous methodologies available to understand better the social, economic, and environmental impacts of agriculture in the belief that credible evidence of what works and what does not will lead to more sustainable practices. One of the ways that COSA pursues its goals is by conducting large n surveys intended to allow researchers to estimate statistical significance and reasonably establish attribution. Sometimes, however, large n surveys are simply not feasible or should be conducted in conjunction with other more qualitative methods. In these instances, for small n cases where only a few units of assignment exist, COSA has developed a mixed-methods approach built upon a contribution analysis, pioneered by John Mayne.[2] This COSA Paper sets out to provide insight into the practicalities and value of Contribution Analysis for small n impact assessments by drawing on lessons learned and challenges faced during a real-world evaluation[3]. The hope is that our hard-earned experience and practical suggestions will be useful to researchers facing conditions for which a small n impact assessment using Contribution Analysis is an appropriate option for answering their research questions.[4]

---

1 COSA is a non-profit, independent global consortium of partners dedicated to accelerating agricultural sustainability through robust information systems. It is financed in part by leading agencies, including the Swiss Government (SECO), the Ford Foundation, and the InterAmerican Development Bank.

2 Mayne 2001, 2008, 2011, 2012.

3 COSA also usefully employs small n modelling-based approaches in some situations. Our focus in this paper, however, is on using Contribution Analysis in small n evaluation situations. We regard this as an indirect response to Howard White, who has listed the need to make progress with small n impact evaluations as one of five major challenges facing the impact evaluation community (White 2011).

4 Following Bamerger (2012), a mixed method is understood as an attempt to integrate "predominantly quantitative and predominantly qualitative approaches to theory, data collection, data analysis and interpretation."

Large and Small n Research

Here, **n** refers to the number of units of assignment available for the survey sample – that is, the units assigned to a treatment (i.e., the sample size). Small n research is often associated with behavioral science, education, and policy studies but can occur in any field of study where there are simply too few units to allow statistically significant comparisons between treatment and control groups. As opposed to large n research, which seeks to discover a causal inference from a large number of observations, small n research seeks to do the same by studying a limited number of cases, but in greater depth.

Large n analyses can employ various strategies to conduct experiments and quasi-experiments that can validly attribute results to specific activities, including statistical matching such as propensity score matching, difference-in-difference (DID), regressions, and so on. In the realm of large n analysis, prominent advocates for experimental methods like the Bill & Melinda Gates Foundation, the International Initiative for Impact Evaluation (3ie), and the Jameel Poverty Action Lab have successfully promoted Randomized Control Trials (RCTs) as a "gold standard" against which the rigor of other methods is judged.[5]

For the most part, large n research is well understood and the methodology is taught, rather than debated. RCTs limit bias and generate an internally valid estimate by randomly selecting appropriate units of assignment and then placing them in either a treatment group (those who receive a treatment) or a control group (those who do not). By comparing the difference in outcomes between the two groups, it is possible to measure what impact a treatment (program or project) had in a particular place and at a particular time. The random assignment to either treatment or control group is assumed to ensure that the two groups are identical in every way, except for the cause and effect, so no bias is introduced.

5 Publication of the Center for Global Development's (CGD) paper, When Will We Ever Learn (CGD, 2006), helped to usher in a gradual shift to more rigorous evaluations. Here, the method of preference is the RCT, which frequently places high in various evidence ranking systems. There are many evidence ranking systems and organizations in existence. Some of the most well-known are: GRADE, created by the Grading of Recommendations Assessment, Development and Evaluation Working Group, the Scottish Intercollegiate Guidelines Network; or the Oxford Centre for Evidence-Based Medicine. Interestingly, the intent of the CGD paper was to encourage researchers to harness the rigor of RCTs to help inform evidence-based policy, though today even some of the leading advocates of "rigorous evaluation" concede that turning results into policy impact is neither "automatic nor easy" (Dhaliwal, 2011).

However, evaluators of development interventions frequently face settings with too few units of assignment available to supply the power needed to find significant differences between treatment and control groups (when such differences exist in the population) or when other limiting factors make inappropriate purely large n methodologies[6]. Some objects of evaluation are simply too complex for an RCT even when the evaluation question is not. For example, policy interventions, agricultural innovations, and value chain development projects can combine some type of innovation with an observed change at the individual, farm, community or national levels. These types of projects typical rely on multiple inputs or interventions deployed over a substantive period of time to willing participants, and are conducted in partnership with multiple stakeholders. In short, attempting to attribute an observed change to a project that is in practice a "package" of actions and inputs affecting a wide group of actors, over a prolonged period of time, and in potentially different and complex ways may not be appropriate using an RCT.[7]

When the number of assignment units is too small to conduct tests of statistical significance, then a small n approach is preferable. In their assessment of small n research methodologies, White and Phillips[8] identify four situations when a small n approach is most desirable:

When the overall size of the entire population from which a sample might be drawn is very small. Take, for example, policy research that often focuses on those exceptional outliers that suggest an innovative new approach to success. The focus of research is not on establishing a trend line but investigating what makes a policy (or pilot study) work.

When there is so much diversity in the treatment population, the treatment itself or the context of the intervention that divining sub-groups for analysis results in groups that are too small to provide statistically meaningful results. In real-world evaluations, this type of situation arises more often than is generally acknowledged because the intervention, the treatment group or the wider socio-political context changes to such an extent that treatment effects can no longer be separated from other possible effects.

When the treatment is for an entire country. In this case n=1.

When there are budgetary or political constraints that prevent large n research. Again, in the real world of evaluation, both are potentially decisive far more often than is desirable or acknowledged.

---

6 Sample size is an important factor in determining statistical power, or simply power. Power refers to how confident we are in detecting a change in outcome attributed to a particular intervention. It is not the intention of this paper to discuss power in detail. More information on power can be found in J-PAL's excellent briefing note: The Danger of Underpowered Evaluations.

7 There are a number of methodological challenges to conducting an impact evaluation of this type, particularly in the realm of agriculture (Nelson and Maredia 2001; Campbell et al., 2003; Mayne and Stern 2013; Kidoido and Child 2014; de Janvry, Dustan, and Sadoulet 2011; Hawkins 2016). The Overseas Development Institute (Jones et al., 2011) outlines more broadly when and where experimental methods should be used.

8 White and Phillips 2012.

It is easy to imagine additional situations where small n methodologies might be most appropriate. In their examination of methods for ex-post impact assessments related to agricultural technologies, de Janvry, Dustan, and Sadoulet (2011) list other considerations that may guide the choice of research methods toward a small n approach, such as the type of intervention (simple vs. complex), the duration between treatment and effect, and the transferability of results (or representativeness of the experimental environment relative to the ultimate adoption domain).

While settings sometimes allow researchers options for selecting the unit of assignment, COSA and many other organizations doing agricultural impact assessments increasingly find that the unit of assignment is given, particularly where interventions are disseminated to farmers through intermediary groups. For example, many of COSA's assessments have looked at projects implemented with farmers through producer groups, NGOs or trading partners. In these cases, the groups—not the farmers—are the units of assignment since the intervention is assigned at the group level. The situation is analogous to testing a new teaching method. While student surveys properly measure performance, classrooms (or even schools) are the units of assignment if the method was assigned at the classroom level. Intuitively, teacher quality and specifics of class make-up could have a large impact on the results of the new method that could affect overall means of performance.[9]

The research situation for COSA is complex because often too few suitable groups for statistical power are located within regions offering similar agro-ecological and socio-demographic conditions needed for good counterfactuals. Figure 1 shows the situation in regard to Producer Organizations (PO)s in western Kenya, with many factors exerting influence on the group, which then mediates these influences among its farmers. Thus, in this case and many others, the setting dictates the unit of assignment rather than allowing the researcher to freely select a unit of assignment that would allow a larger n.10 Adding further complication, selection bias is frequently a factor. Even when the opportunity for a baseline exists, specific groups may have already been chosen for the treatment, usually because the group has characteristics that could increase the likelihood of success for the intervention. Given all of the above, small n research techniques are an important component to our skill set.

9 We acknowledge that power issues associated with a small sample size might be overcome in some situations by randomizing at a lower level or through stratification. For an excellent discussion on overcoming challenges to randomization, see Heard et al. (2017).

10 There is much ambiguity over the number of clusters needed to quantitatively establish attribution. For a summary of the debate, see Ozler (2012).

In cases where a small n methodology is required, producing a quality impact assessment can be difficult even when time and money are not barriers to success. At COSA, attaining a high level of research quality is a priority. The World Bank has set three minimum criteria for a quality impact assessment[11]. They are:

A set of indicators that can measure inputs, implementation processes, outputs, intended outcomes, and impacts.

A counterfactual that persuasively indicates that observed outcomes are the result of a treatment, rather than serendipitous factors like good weather, a thriving economy or other similar, overlapping projects.

Analysis, using accepted procedures, that the treatment has benefited a significant number of the intended beneficiaries.

Regardless of the methodology employed, these simple criteria offer guidelines for quality work and form part of the COSA criteria in its work, including small n research. Contribution Analysis is an approach that helps us meet these criteria and fulfill our commitment to quality. Below, we outline our approach and how it meets each of these criteria.

11 Additional optional criteria are also cited. See: Bamberger, Michael. Conducting quality impact evaluations under budget, time and data constraints. World Bank, Independent Evaluation Group/ Poverty Analysis, Monitoring and Impact Evaluation Thematic Group, PREM Network, 2006.

Contribution Analysis

In recent years, Contribution Analysis has gained a considerable following because of its methodological rigor, applicability for a wide number of real-world development applications, and promise to address questions about why an intervention worked—making it both an evaluation and learning activity. Mayne (2012) captures the essence of Contribution Analysis (CA) as follows:

CA is based on the existence of, or more usually, the development of a postulated theory of change for the intervention being examined. The analysis examines and tests this theory against logic and the evidence available from results observed and the various assumptions behind the theory of change, and examines other influencing factors. The analysis either confirms—verifies—the postulated theory of change or suggests revisions in the theory where the reality appears otherwise. The overall aim is to reduce uncertainty about the contribution an intervention is making to observed results through an increased understanding of why results did or did not occur and the roles played by the intervention and other influencing factors.

One of the strengths of CA is that it follows a reasonably clear and structured set of methodological steps[12]. This is a significant refinement from other small n evaluation methodologies that can be more aptly characterized as mere “approaches” or “ways of thinking[13].”

Six Steps to Contribution Analysis

Step 1: Set out the attribution problem to be addressed

Step 2: Develop a theory of change and risks to it

Step 3: Gather the existing evidence on the theory of change

Step 4: Assemble and assess the contribution story, and challenges to it

Step 5: Seek out additional evidence

Step 6: Revise and strengthen the contribution story (Steps 4 through 6 are iterative until the contribution or the challenges are confirmed)

12 Each step is explained in more detail in Mayne, 2012. Also see, Better Evaluation, http://www.betterevaluation.org/en/plan/approach/contribution_analysis

13 Rick Davies points to this issue in his discussion of recent developments in Realist Evaluation methods. See: http://mande.co.uk/2016/uncategorized/two-useful-papers-on-the-practicalities-of-doing-realist-evaluation/

Despite the appeal of CA, there remains little explicit guidance on how to avoid different types of bias. Bias is a potential problem for both large and small n research, but as a recent International Initiative for Impact Evaluation (3ie) publication points out, for CA, a particularly pernicious form of bias may arise from the "systematic tendency to either under- or over-estimate the strength of a causal relationship" (White, et al., 2012)[14].

14 Recently, there has been a growing interest in how to assess the level of confidence underpinning a contribution claim. For example, Befani and Stedman-Bryce (2017) have developed a method for theory-based approaches, coined "Contribution Tracing," which combines the principles of Process Tracing and Bayesian Updating to increase the level of confidence in a contribution claim, as well as reduce the risk of bias. While this approach appears promising, we do not discuss it here.

The COSA Contribution Analysis Framework

COSA has adopted a disciplined contribution analysis framework for impact evaluation in small n situations, incorporating mixed methods. This framework combines quantitative and qualitative tools that enhance the individual strength of each methodology. The approach has grown out of our experience working with agricultural POs and their farmers on issues related to sustainability, particularly when the research context involves a small number of producer organizations and a large number of small- and medium-scale farmers within a subnational area (often within one or a limited number of agro-ecological zones). It takes selection bias into account through the explicit step of presenting it as an alternate explanation, or challenge to, the contribution story for any changes observed following the intervention. Our approach is summarized below.

Steps 1 and 2: Set out the attribution problem to be addressed; develop a theory of change and risks to it.

The COSA approach begins by clearly defining the change pathway, underlying assumptions, and necessary support factors from inputs to the results that are being sought. By making our causal hypothesis explicit, discussing it with our stakeholders, and thinking through alternative competing hypotheses, our draft Theory of Change (ToC) is a fundamental first step in building a case for reasonably inferring causality.[15]

As an initial step in developing our ToC, the COSA mixed-method approach involves a qualitative scoping study that helps establish our hypothesis regarding how change happens. The scoping study draws on secondary data sources and key informant interviews with farmers and others to help us understand our research setting and to help identify likely causal pathways. We also use this component of our research to collect data from other key stakeholders, often in the form of data-rich questionnaires (e.g., although producer organizations are few in number, nevertheless considerable data must be collected in order to understand the scope of their functions and effects). This scoping step serves many functions, two of the most important being: (1) production of a draft version of the ToC and (2) the collection of context information that we can use to adapt our surveys so that they are appropriate for our intended research subjects (e.g., correct indicators, sample strategy, data collection, cultural sensitivities, and so on).

15 Theories of Change are typically composed of a change diagram and narrative explanation of the diagram. For communications purposes with donors or the general public, a very simplified ToC may be sufficient to convey the basic logic of the intervention. To be useful for an evaluation, however, cause and effect change statements need to be made as clear as possible without becoming too burdensome for practical use. For this reason, many interventions choose to develop two expressions of their theory of change: one that conveys the general logic of the intervention in a simplified format in which explanatory emphasis is placed on a graphic visualization, and a second, more robust ToC for evaluation and learning purposes. The ToC developed as part of an impact assessment necessarily falls into this latter category.

Step 3: Gather the existing evidence on the theory of change.

As a critical component of the evidence base, COSA also employs a quasi-experimental design using panel data models following the differences-in-differences (DID) approach.[16] While contribution analysis often uses qualitative data to provide evidence about changes, our approach benefits from solid quantitative data and sound econometric methods to minimize selection bias and power issues in the estimation. In this way, COSA can account for the fundamental importance of context to better understand change pathways and provide quantitative data arrived at through rigorous analysis.

16 Though the quantitative research component of our method is important in its own right, for the purpose of this paper, we will focus primarily on our Contribution Analysis.

Steps 4 through 6: Determining and confirming (or not) the contribution story.

A final component of our baseline study is an "Insights" set of activities that involve more structured key informant interviews, along with focus group discussions. We use this set of activities to help validate the ToC in light of new information that we have gathered during the research process. This may imply only minor changes to the ToC or, when new and unexpected information warrants, a more thorough re-draft. This sort of planned reflection is critical for providing both 1) the needed assurance that the ToC reflects a close proximity to reality; and 2) accounting for disagreements among stakeholders in endline conclusions about contribution. By engaging stakeholders in a meaningful dialogue, we can be more confident in the accuracy of causal claims and our understanding of them.[17]

---

17 Planned reflection through a participatory process can sometimes help managers and stakeholders reach a shared understanding of how positive change is meant to occur, though in complex programs in which there are many different interests at stake, this can be a challenging task.

Western Kenya Case Study

Using our method in practice provided considerable insight into some of the challenges often faced by researchers who employ the CA approach.[18] For example, COSA was selected to conduct an impact assessment of an initiative to certify coffee farmers in western Kenya to multiple standards. The sample consisted of two POs already selected for certification and four control POs. The four control POs represented the only ones in the region to be similar enough to the treatment POs to act as counterfactuals. Thus, we faced both a small n and selection bias situation.

Table 1 shows the high-level approach used in western Kenya, mixing qualitative and quantitative methods for achieving different steps of the contribution analysis. As is typical, our ToC (see Figure 1) emerged from a wide-ranging stakeholder consultation. We conducted the baseline for this work in 2015 and will conduct a follow-up study in 2018.

Table 1: Contribution Analysis research plan for Western Kenya using mixed methods

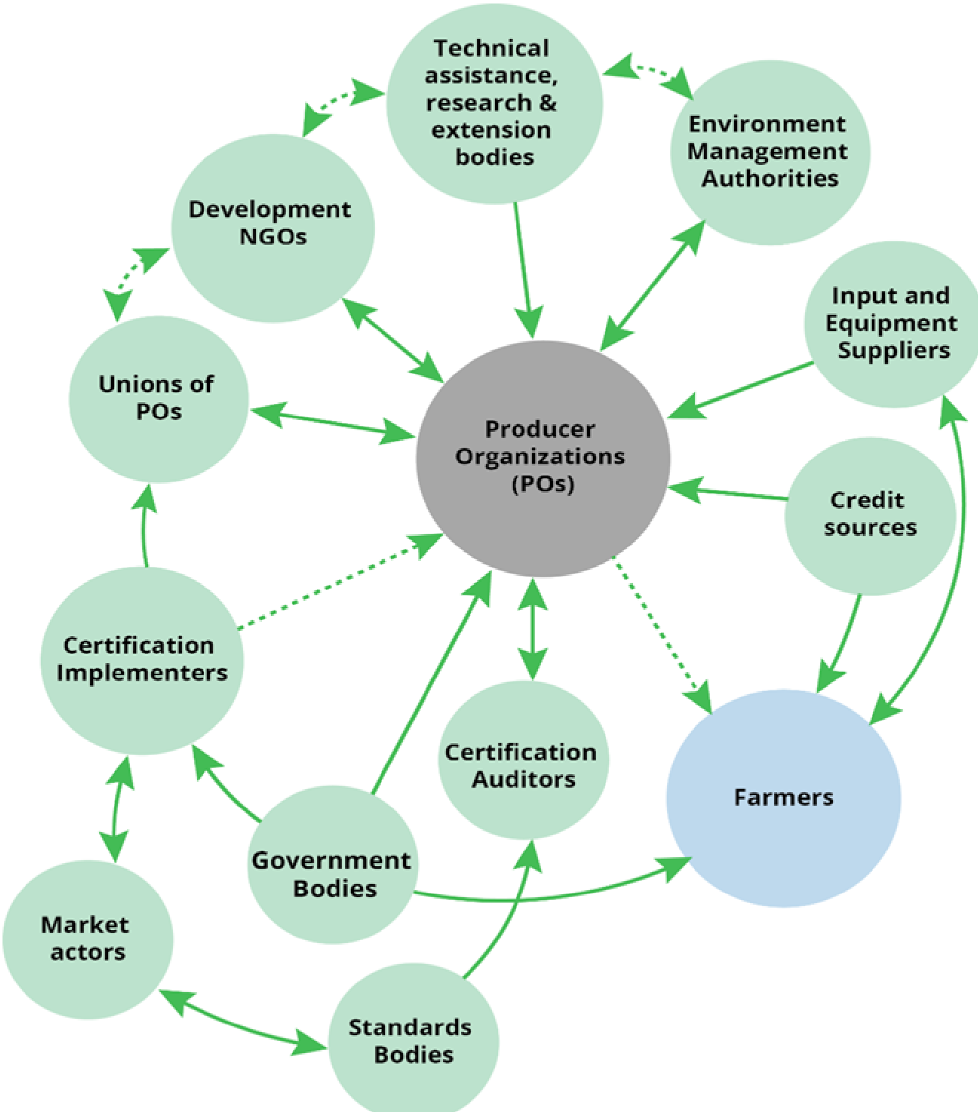
Contribution analysis step	Method
Baseline phase	
Steps 1 & 2: Set out the attribution problem to be addressed; Develop a theory of change and risks to it.	Secondary data sources and key informant interviews
	Interviews to key actors (POs, CMS)
	Participatory Rural Appraisals
Endline phase	
Step 3: Collecting evidence of change from baseline using difference in differences design (quantitative)	Farm-household survey
	PO survey
Steps 4-6: Final confirmation (or rejection) of contribution story (qualitative)	Structured key informant interviews
	Focus groups

18 The complete baseline report can be download at: http://www.isealalliance.org/sites/default/files/ISEAL%20DIPI%20Kenya%20baseline%20study%20report.pdf

For the baseline work, the quantitative survey measured the differences in the farmer performance for each of the PO sample groups. We then used qualitative methods to understand the reasons for differences in average performance for key indicators among the groups. With this information, we can build a more accurate appraisal of whether changes in performance are actually in evidence at endline, and if the certifications contributed to the differences or if they existed beforehand. For example, coffee yields differed on average among the sample groups according to the quantitative data.

We conducted focus groups with the farmers and knowledgeable informant interviews with PO board members, researchers, coffee buyers, and others. We learned that within every sample group a few farmers obtained very high yields while most others obtained poor yields. Farmers attributed the good yields to specific practices which poorer producing farmers did not use. The qualitative work suggested that factors influencing practice adoption included financial resources, low prices, and lack of trust in PO management. At endline, to build and assess a contribution story, we will investigate change in performance indicated by the quantitative work and whether the certification led to greater practice adoption, but also whether it successfully targeted improving any of the influencing factors. For some performance indicators, stakeholders did not know why a difference was observed. In these cases, the quantitative data can help to fill in gaps, particularly in information about initial endowments, such as family size, education, and other variables.

Figure 1: Key Stakeholders



Who should be included as a stakeholder for a producer organization? Through participatory processes, we were introduced to a surprisingly diverse group of actors, many of whom we might not otherwise have included.

Identifying and Interviewing Stakeholders

Following best practices, we aimed to ensure that our interview sample was representative of all of the key stakeholders involved in the project. There is already a vast literature on conducting interviews for qualitative research that we will not rehash here. Our experience, on the other hand, suggests two particularly useful points worth emphasizing:

1. To capture both the visible and invisible (e.g., informal, cultural, ideological, etc.) causes of change and their expected and unexpected effects, it is important that stakeholders are adequately identified and mapped into the research plan. Ideally, stakeholders are identified during the scoping phase of the research, but it is not uncommon to add previously marginalized stakeholder groups as they emerge during the course of the research. COSA field researchers are trained to consider whether to modify or expand their research plan as required to achieve a truly representative sample.

2. Figure 1 makes clear the large number of stakeholders in the intervention. For our goals at the baseline, we included stakeholders with direct or wide perspectives on the farmers and POs – such as the farmers themselves and PO management, as well as government and research roles. For exploring contribution at endline, we will interview these same stakeholders and then follow a snowball approach in which we interview each of the stakeholders mentioned by others as having significant insights for understanding reasons for changes in performance. We will follow up with mentioned stakeholders until we see that a point of saturation has been reached (e.g., once we stop hearing new information). This is, of course, a subjective decision and one that is undoubtedly easy for an external audience to second guess, but one that becomes reasonably clear in practice (Patricia and Ness 2015). It is possible that after gleaning the same information from a great many people that the next person might add something new, but this is unlikely and at some stage, a pragmatic decision needs to be made considering budget, time, and other constraints.

Confirmation Bias

Confirmation bias occurs when a researcher inadvertently influences the supply or interpretation of the information they collect so that it better matches their own preconceived understanding of cause and effect. COSA researchers are critically aware of this bias, which is probably one of the most necessary steps to avoiding it. Nevertheless, experience has taught us that through a few simple preventative measures, we can minimize this threat to internal validity.

1. Communicate in the language of the interviewee. Violating this simple rule of thumb is the most obvious threat to conducting quality research and yet it is an enticingly easy mistake to make. Field researchers generally assume that they are effective communicators because they have been professionally trained to ask questions. But, it does not matter how simply and eloquently a question is asked if it is posed in a language that is not fully familiar or native to the interviewee. Merely because a person can respond in, for example, English or Spanish does not mean that they can correctly understand the question or are willing to give the same answer that they would give if the exchange occurred in their native language.

2. In consulting stakeholders, always present a number of competing cause and effect hypotheses for any changes observed from the baseline to the endline without showing a preference for one over the other. Many people prefer to make intuitive leaps of logic based on convention or the advice of an authority figure. COSA researchers have learned that even the appearance of favoring one answer over another is a fetter to the free expression of opinion. By examining multiple ToCs without bias, we can explore counterfactual hypotheses that might otherwise have received little or no consideration because they were unfashionable or taboo.

Making Sense of Data

Making sense of qualitative data without interjecting researcher bias is always challenging. From experience, COSA employs three types of safeguards to avoid this common error:

1. Triangulation of data is one of the most effective techniques we use to ensure that data is accurate. Our mixed-method approach allows us to question, make sense of, and validate what we learn at different stages of our research and from different sources or perspectives. For example, we use information gathered from the Insight stage of our analysis to help validate our farmer survey results. Likewise, we use our farmer surveys to fact check the information that we have gathered from Scoping phase questionnaires. We may, for example, try to corroborate the level of service provided by a producer organization with a related question from our farmer surveys. Regardless of the specific example, in all cases, finding multiple sources of evidence for cause and effect relations is critical to identifying the existence of a relationship and understanding why and how it works.

2. Coding and categorizing qualitative data is a sometimes painstaking but necessary component of our research. We have found that while generally desirable, transcribing interviews in their entirety can be overly burdensome when the volume of interviews is substantial. Nevertheless, qualitative data needs to be recorded in some form in order to assess and make sense of it. As a compromise, COSA researchers use the Cornell Method for note taking, in which qualitative data is recorded for significant information and then immediately categorized according to a tangible change process—including changes that are ultimately excluded from the ToC due to lack of evidence or general acceptance. Over the course of the research project, COSA analysts are then able to weigh the preponderance of change-specific qualitative evidence.

3. Choosing the right mix of diverse research skills and staff is another way COSA tries to minimize the potential for bias. A central tenet of the COSA research approach is to partner with local, science-based organizations. Our research teams are always comprised of individuals with local, context-specific knowledge paired with foreign-based subject experts. Furthermore, COSA has learned that certain topics such as gender perspective can present an important variable; therefore our research teams typically include both men and women in leading roles. The failure to account for environmental, social or business emphases can similarly color the findings. Bringing to bear a diverse range of expertise and perspectives to the analysis helps to avoid blinkered thinking.

Concluding Remarks

Contribution Analysis plays a significant role in helping us move beyond mere statistical correlations to a nuanced understanding. While quantitative data provides us with a breadth of information on key indicators, it is the in-depth analysis of qualitative information through CA that allows us to make better sense of it. In the case highlighted here, where producer organizations are relatively few but are composed of a large number of farmers, the ability to engage complementary large n and small n research of this kind is appropriate, particularly when due diligence has been taken to ensure a high level of research quality.

Box 1: Contribution vs. Attribution

Scholars have spent considerable effort to distinguish “attribution” from “contribution.”[19] Vaessen distinguishes the two as follows: “Attribution emphasizes the issues of whether or not and how much of a particular change can be attributed to an intervention. Contribution emphasizes the confluence of multiple causal factors to a particular change and emphasizes the issue of whether or not and how an intervention contributes to the change.”

RCTs aim to establish attribution, which is generally regarded as more rigorous than contribution. As Cartwright points out, RCTs are so attractive because they tell us what caused an outcome without the need to understand how the outcome came about.[20] But in most small n research settings (e.g., policy studies) a mere statistical inference is not sufficient in itself. Researchers want to know why something worked or did not. Questions around transferability and scalability are important: If a treatment worked in one place and time, will it work in another? To answer such questions, it is necessary to know what other context factors are important (or support factors in Cartwright’s terminology). It may be that X leads to Y in one context, but not another.

On the other hand, causality is established in CA not through statistical correlations but through an understanding of how cause and effect are related. It is underpinned by sufficient evidence to warrant the understanding.[21] CA thus starts from (1) the creation of an explicit and reasoned theory of change and is supported by (2) an evidence base that corroborates the identified cause and effect relationships. By building a credible evidence base through the collection of data that supports each step in the ToC, the resulting “contribution story” is sufficiently robust that “a reasonable person, knowing what has occurred in the program and that the intended outcomes actually occurred, agrees that the program contributed to these outcomes” (Mayne 2011). As Mayne argues, a contribution claim can ultimately be expressed as:

Contribution claim = verified theory of change + other key influencing factors accounted for.

It is important to reflect that counterfactual reasoning is basic to both RCTs and CA, though in the latter, the counterfactual is established by considering alternative theories of change or rival explanations that are ultimately discarded due to a lack of evidence.

19 ‘ Stern et al. 2012, 2013; Patton 2008, Mayne 2012

20 Cartwright and Hardie 2012.

21 Mayne has defined a contributory cause as something neither necessary nor sufficient, but still a necessary component to a package of causes that is sufficient to produce an outcome (Mayne 2012); also see Mayne 2012a.

References

Cartwright, N., & Hardie, J. (2012). Evidence-based policy: a practical guide to doing it better. Oxford University Press.

Child, K., Kidoido, M. M., & Kanyuuru, C. (2015). Using the Theories of Change approach to monitor, evaluate and learn in the CGIAR research program on Livestock and Fish.

Contribution Analysis, Better Evaluation. (2017). Betterevaluation.org. Retrieved 24 April 2017, from http://www.betterevaluation.org/en/plan/approach/contribution_analysis.

De Janvry, A., Dustan, A., & Sadoulet, E. (2011). Recent advances in impact analysis methods for ex-post impact assessments of agricultural technology: options for the CGIAR. Unpublished working paper, University of California-Berkeley.

Dhaliwal, I. (2011). Conditional cash transfers and health: unpacking the causal chain (http://nonie2011.org/?q=node/88).

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. The Qualitative Report, 20(9), 1408.

Hawkins, A. J. (2016). Realist evaluation and randomised controlled trials for testing program theory in complex social systems. Evaluation, 22(3), 270-285.

Heard, K., et al. (2017). Real-World Challenges to Randomization and Their Solutions. JPAL.

Jones, N. et al. (2009) 'Improving Impact Evaluation Coordination and Use'. A Scoping Study commissioned by the DFID Evaluation Department on behalf of NONIE (www.odi.org.uk/resources/ download/3177.pdf).

Kidoido, M., & Child, K. (2014). Evaluating value chain interventions: A review of recent evidence (Vol. 26). ILRI (aka ILCA and ILRAD).

Koehler, J., Gosztonyi, K., Feda, B., & Child, K. (2015). Toward mixed-methods impact evaluation: Making stabilization assessments work for development cooperation. The Economics of Peace and Security Journal, 10(2).

Mayne, J. (2001). Addressing attribution through contribution analysis: using performance measures sensibly. The Canadian journal of program evaluation, 16(1), 1.

Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect.

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. Evaluating the complex New Brunswick. New Jersey: Transactions Publishers, 53.

Mayne, J. (2012). Contribution analysis: Coming of age? Evaluation, 18(3), 270-280.

Mayne, J. (2012a). Making causal claims. ILAC Brief, 26.

Mayne, J., & Stern, E. (2013). Impact evaluation of natural resource management research programs: a broader view. Canberra: Australian Centre for International Agricultural Research.

Ozler, Berk. “Beware of Studies with a Small Number of Clusters.” Development Impact. World Bank, 21 June 2012. Web. 25 Apr. 2017.

Patton, Q. (2008). Advocacy impact evaluation. Journal of Multidisciplinary Evaluation.

Ramalingam, B. (2011). Learning how to learn: eight lessons for impact evaluations that make a difference. ODI Background Note. April.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development. DFID: Department for International Development.

White, H., (2011). Five challenges facing impact evaluation, Monitoring and Evaluation News (http://mande.co.uk/2011/uncategorized/five-challenges-facing-impact-evaluation/).

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. New Delhi: International Initiative for Impact Evaluation.