QuIP
Qualitative Impact Assessment Protocol

Briefing note: case selection for QUIP studies

Introduction

The QuIP approach places a strong emphasis on the rigour of good **purposive case selection** in qualitative data collection, compared with the approach taken to representative sample sizes in quantitative studies.

Sample size and selection strategy for a QuIP study are highly dependent upon contextual factors which should help to define a reasonable scope for the study. Key questions which will determine the number, geographical location and type of respondents include:

- What is it that you really want to know about your project/area?
- If you have monitoring data, what is this telling you about variation, including positive or negative deviance?
- Are there particular groups, locations or projects that it would be helpful to understand more about?

The QuIP is an opportunity to do a 'deep dive' into a selected group, and sample selection should be based on expected saturation within a defined group or location.

As a guide we suggest using a minimum sample of 24 individual respondents, which can be complemented with focus groups. You should expect that most respondents within this group have experienced broadly similar outcomes based on your knowledge of their profile (e.g. sex, age, location) and circumstances (e.g. wealth, exposure to intervention). It's often a good idea to split the 24 individual respondents between two contrasting localities (e.g. one thought to be doing well and one badly), and the 12 selected from each can also be quota sampled to draw out other important differences (e.g. by dividing it equally between men and women). But it is generally dangerous to generalise about differences between sub-categories based on less than six cases within each. If a commissioner is seeking evidence about more fine-grained differences then they should be encouraged to increase the overall sample size

Similarly, where there is a high degree of variance, there is often a case for conducting two or more QuIP studies, for example in contrasting geographical areas (rural/urban), in areas where an intervention has been delivered differently, or where results show significant differences in outcomes between groups.

There is no universal best practice method for sample selection for a QUIP study, but this paper explores in brief the following key factors to consider:

a) the main purpose of the study

b) availability of relevant data about variation in the characteristics of expected gainers and losers from the project

c) availability of relevant data about variation in their exposure to project activities

d) time and resource constraints

Factors affecting sample selection

(a) Main purpose of the study

Deciding who to interview, how many people to interview, and how best to select them requires clarity about what information is being sought, by whom and why. Neglecting this not only leads to poor practice but also misunderstanding about the quality of a study. For example, sample bias is not a problem for a QUIP study that deliberately set out to identify drivers of successful outcomes by interviewing what Atul Gawande refers to as "positive deviants." Deliberately selective (hence biased) sampling is in this instance fit for purpose!

More generally, differences in sampling strategy arise from whether the priority is to confirm and quantify the overall impact of a completed project on a defined population in relation to a predetermined set of measurable indicators, or to identify and explore what is happening in a more open-ended way – to improve implementation of an ongoing project, for example. The QUIP is a relatively open-ended approach. Its primary purpose is to gather evidence of causal processes at play, not to quantify them.[1] Deciding on the number of interviews and focus groups to conduct depends less on reducing sample bias than on assessing at what point the extra insight into causal processes gained from more data no longer justifies the extra cost.[2]

(b) Contextual variation

Random selection of respondents across the entire population affected by the project is an option for sampling in a QUIP study, but there are risks associated with this which can be avoided if some thought goes into any known contextual variation amongst intended beneficiaries. If we expect causal processes to be different for different sub-groups, and we have data that enables us to identify those sub-groups prior to sample selection then there is a case for stratified random sampling. For example, we might choose to ensure the QUIP study includes a minimum quota of people living in urban and rural areas. Stratification of the sample on these grounds is an art not a science that depends on prior thinking about what contextual factors are most likely to be a source of variation in project outcomes.

It also depends on the quality of monitoring data available. For example, it is good to stratify on the basis of baseline income or wealth indicators. Better stratification might also incorporate data on observed change in income or wealth income over the project period. Hence a simple design might quota sample four groups: richer and improving; richer but declining; poorer but improving; poorer and getting worse.

[1] If the primary purpose is to quantify specific causal effects then there are two options. The first is to use an appropriate experimental or quasi-experimental approach instead. The second is to build a simulation model, using both QUIP data to identify the main causal factors, and quantitative monitoring data to calibrate their magnitude. The first is more precise, the second potentially more flexible.

[2] To do this formally would not entail estimating statistical sampling errors but a Bayesian process of assigning confidence parameters to prior expectations and assessing how these change with each extra observation.

(c) Exposure or 'treatment' variation

This refers to variation in how project activities affect different people, including those who are direct beneficiaries of different packages of goods and services. In addition there are those who may only be affected indirectly: because their neighbours are affected and may share things with them, for example. If data is available on variation in who directly received what and when, and it is expected that these differences will have different causal effects, then there is a case for stratifying the sample to ensure it reflects the full range of such exposure. This is particularly the case if part of the purpose of the study is to aid decisions about which of a range of project activities or packages to expand or to stop. Impact assessment using the QUIP do not require a control group of people completely unaffected by the project. There may nevertheless be an argument for interviewing some people unaffected by the project, but similar to those affected by it in order to explore whether they volunteer different or additional drivers of change.

(d) Time and resource constraints

A third reason for departing from pure randomization in sample selection is to cluster respondents geographically in order to reduce the time and cost of data collection. One way to do this is to adopt two stage random sampling, with the first stage based on geographical units (e.g. villages, districts or census areas). However, there is often a strong case for using contextual information (e.g. about agro-ecological zones) to purposefully select or at least stratify area selection. The rationale for this is precisely analogous to stratification based on contextual data at the household level as already discussed under (b).

Ultimately, budget constraints may also limit the total number of interviews and focus groups that the QUIP study can cover. There may also be a case for staggering studies – i.e. conducting two smaller studies a few months apart rather than doing a single larger study. This can help to build understanding of project impact lags, pathways and cumulative processes, as well as those of other drivers of change. Sampling strategy for repeat studies can also be informed by lessons from earlier studies; credibility of findings builds incrementally with the addition of each extra piece of evidence.