Element 5: Gathering data and information

Assessing data needs and available data and planning the phasing of future data collection efforts are an important part of PSIA. Identification of data needs will benefit from the prior identification of policy issues, stakeholders, and likely transmission channels, as outlined above. Four discrete steps are suggested: mapping out desirable data for PSIA; taking stock of available data and analysis; coping with PSIA data limitations up front; and addressing PSIA data limitations today so they do not limit PSIA in the future.

Mapping desirable data for PSIA

Analysis of the poverty and social impacts of policy can be extremely data-intensive. Specific data requirements will, of course, depend on the nature of the reform being analyzed and the analytical tool or technique being employed. In approaching data and methods, it is useful to distinguish among data collection instruments (close-ended or open-ended); data type (numeric or non-numeric); and associated methods of data analysis (quantitative or qualitative). Traditionally, analytical approaches have been either quantitative in nature and based on numeric data collected using close-ended data collection methods, or qualitative in nature and based largely on non-numeric data collected using open-ended data collection methods. "Mixed methods" are increasingly being employed and are extremely useful for PSIA.

The approach based on quantitative analysis, numeric data, and close-ended data collection offers certain advantages. Analyzing the poverty and distributional impacts of policy on welfare indicators will require linking data at the macro or sectoral level (generally corresponding to the level of policy intervention) to disaggregated household-level data that capture the welfare measure of interest (usually an income/expenditure aggregate, but possibly other welfare measures such as literacy or infant mortality) and other behavioral variables (such as access). Close-ended surveys have generally been used to collect such data. For analysis to be generalizable, data should be derived from a random sample. When the reform is expected to impact only a discrete group (for example, laid-off mine workers) or a geographic subregion, purposive sampling of just that group or subregion may be more appropriate and economical than a nationally representative survey. Numeric data can be used to undertake statistical and multivariate analysis to test hypotheses and determine relationships (see table 1).

Likewise, the approach based on qualitative analysis and open-ended data collection has particular strengths. A variety of open-ended data collection methods can be used to collect non-numeric information relevant to PSIA. Qualitative and contextual data can be collected through participatory appraisals, asset mapping, and structured interviewing of individuals, communities, or focus groups. This information can be used to undertake stakeholder analysis (discussed above), participatory poverty assessment, beneficiary assessment, institutional analysis, and risk analysis (discussed below). Open-ended data collection methods such as those described in table 1 permit an interactive analytical process—one in which research questions can be formulated, answered, and analyzed iteratively in the field. The open-ended approach allows subjects to articulate the research problem and question. This interactive analytical process could enable quicker turnaround and a shorter time lapse between questionnaire design and analysis than close-ended data collection methods and associated statistical analyses.[7] Open-ended data collection methods may also be undertaken using a random sample or a purposive sample and may also be quantified to tabulate and analyze information.[8]

In undertaking PSIA there is much benefit to mixing and, where possible, matching elements of the above approaches.[9] This includes drawing on different types of data collected by different techniques for multidisciplinary analysis. It is important to be aware that economic analysis is not limited to quantitative analysis. Close-ended and/or open-ended data collection techniques can be used to generate numeric and/or non-numeric data, for analysis using quantitative and/or qualitative techniques and approaches. Moreover, analytical methods can be mixed sequentially or in parallel over time. Mixed methods can leverage the benefits of both quantitative and qualitative analysis. Qualitative analysis can inform the design of close-ended questionnaires or the specification of an econometric model and generate hypotheses to be tested

Table 1. Data Collection Methods

Aspect	*Close-ended*	*Open-ended*
Data collection instrument	• Structured, formal, predesigned questionnaires, such as living standards measurement study, social impact assessment survey[a], willingness-to-pay survey, client satisfaction survey, citizen report card.	• In-depth, open-ended, or semi-structured interviews, such as key informant interviews and case histories, focus group interviews, community interviews, mini-surveys. • Ethnographic observation. • Systematic (or directed) consultation, such as beneficiary assessment. • Participatory data collection methods, such as participatory action research, participatory rural appraisal, participatory public expenditure review. • Focus group discussion. • Community and institutional surveys. • Written documents (for example, program records, process documentation, media reports). • Participatory visual exercises.
Analytic method	• Predominantly statistical analysis. • Deductive reasoning.	• Inductive reasoning. • Interactive analytical process: research questions formulated, answered, and analyzed iteratively, e.g. in stakeholder analysis, participatory poverty assessment, scenario analysis. • Methods tailored to social context.
Advantages	• Findings can be generalized. • Can quantitatively estimate size and distribution of impacts. • Explains statistical correlations.	• Able to analyze behavioral responses, explore new hypothesis, or recognize previously undiscovered phenomena. • More effective in capturing intra-household features and non-income dimensions of poverty. • Can identify particularly vulnerable subgroups. • Allows respondents to articulate their own views.
Disadvantages	• Results not available for long period of time. • Limited types of information can be gathered. • Can sometimes be expensive and time-consuming.	• Findings difficult to generalize, and difficult to aggregate and compare systematically. • Fieldwork requires greater research skills than for quantitative enumeration.

Note: This table is intended to provide an indicative distinction between these methods and not a comprehensive description of individual techniques.
a. Social impact assessment adopts a more eclectic approach to data collection, choosing among open-ended, semi-structured, and close-ended instruments to fill information gaps for mixed-method analysis.
Sources: Adapted from Carvalho and White 1997; Baker 2000; and World Bank 2002a.

further through quantitative research. Hypotheses generated by qualitative analyses can be tested for generalizability using quantitative approaches. The results of quantitative analysis can be further examined using open-ended data collection methods to develop a richer understanding of the impacts of policy on different subsets of the population, and to analyze counter-intuitive results that might otherwise be dismissed as spurious. And a successful mixture can elucidate history, context, process, and identification of transmission channels and differential impacts. While mixed methods can involve higher costs, requiring more complex skills and coordination with multidisciplinary teams, the benefits in some cases outweigh the costs. As the work of Amartya Sen and others demonstrate, economics has contributed a great deal to, and made liberal use of, qualitative analyses.

Taking stock of available data and analysis

The first element of the stocktaking is to ascertain the existence of key data. This will allow identification of data gaps that need to be filled or taken into account when choosing an analytical approach. Household survey data are generally pivotal to undertaking quantitative poverty and distributional analysis.[10] An important consideration for poverty and social impact analysis is whether, in addition to a welfare (e.g. income/expenditure) aggregate, there is information in the survey that provides the variable (or the computation of such a variable) related to the policy lever in question—for example, household expenses on transport, or specifically public bus transport, if bus tariffs are to be increased; or purchases of maize at subsidized prices, if the subsidy is to be removed. Other important sources of data include sector studies—which may

include administrative data, household survey data, and qualitative information—and information on the macroeconomic situation, including national accounts. In analyzing policy reform, it is very useful, where possible, to test the robustness of conclusions by matching data from different sources. This is often referred to as "triangulation," the practice of validating results among three different sources. For example, in Armenia three different sources were used to compile and compare information on consumption of, and expenditure on, utilities (using household survey data, utility accounts data, and focus groups). Similarly, for particularly controversial issues, participants in discussion groups may have an incentive to exaggerate or minimize certain impacts. Matching or triangulating results is particularly important to validate such results.

Second, after identifying the availability of relevant primary data, ascertaining the existence of analysis and secondary data on the policy issue at hand is an obvious next step. In many instances, burning policy issues have been the subject of analysis and debate in the past; it is useful to draw on whatever analysis already exists, and whatever public debate has already occurred. Project and program documentation, as well as data and analyses from other development agencies, are invaluable. For sectoral reforms, information from existing sector analysis, including administrative, household survey, and qualitative data, can strengthen PSIA. Academic research and theses can also yield in-depth insights not normally available in official reports.

Third, it is useful to ascertain and build the capacity of local agencies involved in data collection and analysis (such as national statistical offices, ministries, universities, research organizations, consulting firms, NGOs, and so forth) to collect and analyze data.

Coping with PSIA data limitations

In many countries there are severe data limitations to conducting poverty and social impact analysis. Some or many of the desired data outlined above may simply not be available. In this case, policymakers and analysts will need to consider several options, outlined below.

First, they can adapt the analytical approach to data currently available. If the urgency of policy action severely limits the time available to gather further data, expeditious analysis using the limited available data may be required. Some tools and approaches to poverty and social impact analysis are far less data-intensive than others. Adapting the analytical approach to the available data, such as using time-use data or focus group data to construct a simple household model, might be the best course of action. While any analysis entails making assumptions, taking shortcuts generally means making more assumptions in order to proceed. The analysis should be honest and transparent in stating these assumptions. Qualitative techniques, such as individual, community, or focus group interviews, can be used to validate assumptions and inform the design of quantitative surveys.

A second option is to collect more data. If critical data gaps have been identified, it may be useful to gather the data needed—whether administrative or survey data. In the interest of building national capacity and enhancing ownership of the data and analysis, where possible these data collection efforts should be undertaken through national institutions, such as the statistical agency, ministries, universities, or other research organizations. A national household survey is a large undertaking; it can take months to plan and implement such a survey and analyze the resulting data. Where possible, it is useful to identify planned household surveys that are to be fielded imminently and to add key questions relevant to the policy issue at hand. These questions can leverage a wealth of analytical possibilities in the context of a full-fledged household survey.

Alternatively, there are now several "off-the-shelf" survey instruments that can be used to quickly collect, enter, and analyze data (for example, the Core Welfare Indicator Questionnaire, or CWIQ, survey). Social impact assessment surveys, based on purposive sampling, can often be turned around in a shorter time than a representative national household survey. Likewise, depending on the reform issue at hand, quantitative surveys can be employed using a purposive sample (for example, among workers of a firm that is to be downsized).[11] When possible, use of mixed methods, combining qualitative and quantitative analytical approaches to triangulate results, helps to generate richer and more robust findings. The use of data from

a non-representative sample to estimate parameters may sometimes be required, and the "borrowing" of parameters from other countries may also be needed. Again, clearly stating assumptions (for example, that these elasticities apply to the population at hand) will be important in these instances. Care should be taken when generalizing from such a purposive sample.[12]

Third, policymakers can rethink the policy decision or the sequencing and pace of reform. One option is to postpone the policy decision until adequate data can be collected and appropriate analysis conducted. If this course is taken, the costs of delaying reform (a policy decision in itself) will need to be considered. Other possibilities are to pilot or phase the reform, so that progress can be monitored before a final decision is made to implement a national program.

In the end, a tactical judgment will have to be made as to how to proceed based on these considerations. This judgment will be influenced by the time and resources at one's disposal, which in turn will depend critically on political and economic pressure for action. In most cases, decisionmakers will not want to embark on a major policy change without a sound understanding of the poverty and social implications of a policy action, particularly if such action is aimed at reducing poverty. In some instances, however, political or economic imperatives (as in a crisis situation) may lead policymakers to take quick action. Where this happens, it will be important to undertake PSIA as soon as feasible and to consider measures to protect the poor from adverse impacts and vulnerability to significant risks (see section on compensatory measures, below).

Addressing PSIA data limitations today so that they do not limit future PSIA

When circumstances dictate that a policy decision needs to be made without adequate data, it is important that steps be taken to improve the information set over time. Since PSIA is necessarily a dynamic process of formulating and adjusting policy based on increased knowledge, it would also be important to put into place a strategy to gather the necessary data to enhance the basis for further and future (ex-ante and ex-post) analysis of the poverty and social impacts of policy. Such a strategy can be designed in a manner that builds national capacity for data collection and analysis. Where possible, a strategy for data collection should be linked to the timetable for policy formulation, or for policy review and reformulation. In other words, the reason for developing a strategy for future data collection is not solely to permit ex-post monitoring and evaluation of a current policy decision, but also to lay the groundwork for future ex-ante analysis. Developing such a strategy is an integral part of PSIA.