Chapter 6

Measuring household wealth through surveys

This chapter discusses the use of surveys for producing reliable and relevant wealth statistics for households. Sample surveys have been employed for many decades for measuring household income and expenditure, but regular and broad use for the collection of wealth information is more recent. The challenges for practical implementation are discussed.

While the collection of surveys on household income and expenditure is part of the regular production of National Statistical Offices for many decades, the measurement of wealth distribution through surveys is more recent (see Figure 6.1 for a sketch of the typical information cycle for a household survey).

Figure 6.1. **Information cycle for a household survey**
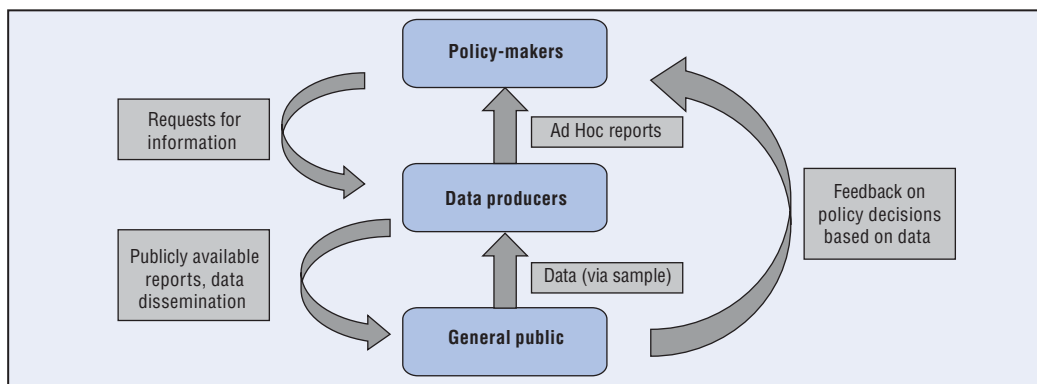
In defining the questionnaire, the sampling design, the interview procedures and the indicators appropriate to this task, and a variety of associated tasks, designers of a wealth survey face particular challenges. They need not only to address the potential problems raised by the highly asymmetrical distribution of wealth, but also to present complex questions to respondents in ways that minimise cognitive difficulties, while persuading the respondents to part with information that most people regard as very sensitive.

The next section provides general background on survey design, emphasising the dimensions that matter most for studies focusing on economic variables. The following section points out the core challenges related to the measurement of household wealth. The remaining sections go into deeper detail, offering a conceptual summary and some practical suggestions related to each phase in the development of the wealth survey.

6.1. General measurement issues

Sample surveys serve the purpose of estimating the value of certain parameters for a population of interest, e.g. the median wealth or the average mortgage debt of households, in a cost-effective way, i.e. collecting data only from an appropriate subset of the population. When designing a survey, data producers must keep in mind that their goal is to achieve the best possible estimates for the outcome measures of interest, subject to a budget constraint (Figure 6.2).

While several possible metrics exist to define what "best" means in this context, they all rely on the same sequence of steps that support the desired estimates. *First*, an instrument is constructed to obtain the information sought, generally in the form of a questionnaire. *Second*, a random sample of the population – sometimes called a *theoretical sample* – is selected. *Third*,

Figure 6.2. **The role of data producers in sample surveys**
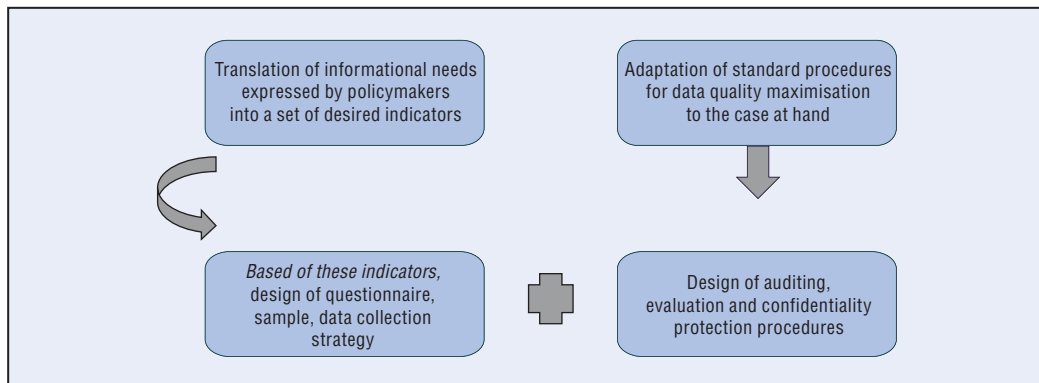
fieldwork activities (generally, attempts to contact or interview sampled households) are conducted and data from the *final sample* are obtained. *Fourth*, the data are processed and weights are constructed. *Finally*, the target estimates are computed. There may be various choices of method at each of these steps, and often such choices have different implications for aspects of the quality of the ultimate results and the overall cost of the work.

The information provided by respondents, sometimes called *raw data*, may be inaccurate or incomplete. Data producers must work on correcting inaccuracies through data-editing and dealing with missing information, typically through imputation. The resulting data set is sometimes referred to as a *validated* or *final* data set.

The final sample differs from the theoretical sample when the former is affected by unit non-response, that is, by the failure to obtain an interview with the desired sample element. If unit non-response rates are high and/or concentrated in specific sectors of the population, the final sample might look quite different from the theoretical sample. *Estimation weights* must be computed for each observation in order to account for any disproportion in the initial probabilities of the selection of sample units, to adjust for differential propensities to unit non-response and to align the final sample composition with that of the target population.

The desired *estimates* are obtained by applying mathematical formulae called *estimators* to the final data and weights. Each estimator has the statistical properties of a distribution. Survey designers generally aim at minimising a version of its *mean square error*, i.e. the sum of the square of its *bias* (distance between the expected value of the estimator and the true population parameter) and its *variance* (a measure of the variation of the estimate that would be expected as a result of repeated execution of sampling and all other steps toward the construction of the estimate). In other words, the distribution of a good estimate is tightly centred around the true parameter. The key steps in designing any sample survey are summarised in Figure 6.2.

The first input tends to come from researchers or policy makers, and it is typically expressed in general terms, e.g. "there is a need for more information on household wealth" or "it is urgent to know who the highly indebted individuals are". Data producers need to translate this policy demand into a clearly defined set of key indicators: for example, median net wealth, average debt-to-income ratio, shares of indebted individuals by employment status, etc. Subsequently, categories must be defined and sequences of questions designed to obtain such information for individual sample elements (see Box 6.1 for an example). Very often, there may be a desire for relatively broad information that can be used to address research or policy questions that are unknown at the time a survey is

Box 6.1. **Measuring household financial vulnerability**

The subprime crisis that hit the United States and, subsequently, the rest of the world in 2007-08 was triggered by the inability of a cluster of low- and middle-income households to repay their mortgages. Events such as this, which depend on the concentration of a given phenomenon in a specific segment of a population, cannot be predicted based on aggregate statistics. An increasing trend in aggregate household debt, or even the average debt-to-income ratio, does not necessarily signal increasing systemic vulnerability; this could also emerge during periods of solid economic expansion.

Sample surveys produce a tool for estimating the probability of financial difficulties at the micro level and the possible economy-wide effects that they may trigger. They allow reconstructing household budgets individually, while also controlling for characteristics such as education and employment history, which help in determining earning potential. They give a fuller picture of each debtor's situation and default risk. For this reason, after the crisis policy makers have expressed a growing demand for survey-based statistics to assess financial vulnerability. Data producers are key to this in that they have to translate this generic demand into a set of target estimates, and then devise optimal strategies for the collection of data, the production of the estimates and the communication of the results. The questions and possible answers involved in this process can be sketched as follows:

- What is "financial vulnerability"? The idea is clearly related to the likelihood of incurring financial difficulties, but measurement requires a clear definition, both in terms of content and in terms of reference unit. In turn, this implies a number of choices. At the time of writing, no international standard existed for this concept, but several countries have defined it as a binary indicator, valued positively if the amount of debt-related payments (capital and interest, summed over all existing debts) at the household level exceeds a certain share of aggregate household income in a given year. Some data producers look only at mortgage debt, while others estimate vulnerability at the individual level. Fine-grained versions of the indicator may also be produced, taking into account the depth of vulnerability.

Once a definition has been decided upon, and assuming a survey framework already exists, target variables must be selected. What is the essential information set? Should it be complemented by auxiliary variables and, if so, which ones? In the case of the most widely adopted definition outlined above, households need to provide at least an estimate of each debt-related payment or set of payments over the course of the reference period, along with an estimate of income. It may also be useful to collect additional information on each debt, in terms of stock (e.g. outstanding principal), the incoming flows of funds (e.g. any refinancing during the year), interest rate, mode of collateralisation and so on. While these items are not strictly necessary to estimate vulnerability in terms of a ratio between outgoing flows and income, they are instrumental in giving a fuller representation of each household's debt situation, which might be of help to policy makers. Since a balance must be struck between respondent burden and information completeness, any additional variables that go beyond what is essential to the original request should be chosen parsimoniously and, if possible, through a bilateral clarification process between the data producers and policy makers.

A measurement strategy should then be determined for each of the target variables. In the following, we forego issues related to the measurement of income and focus on debt. Different types of households may recall debt-related information with varying degrees of difficulty: for example, those who operate under a strict budget constraint might be more aware of the exact amount of each payment, while affluent respondents might not be

Box 6.1. **Measuring household financial vulnerability** (cont.)

equally attentive and might even fail to recall some outflows, such as small-amount payments for consumer durables debited automatically every month on a credit card or bank account. One possible strategy to improve accuracy entails an initial set of Yes/No filter questions, i.e. asking households whether they hold a certain type of debt (mortgage on primary residence, mortgage on other real estate, consumer credit for vehicles, consumer credit for other durables, credit card debt, bank overdraft, informal debts towards friends and relatives and so on). For each debt identified by a positive answer, details are then requested. Another strategy, used in some broader-scope surveys, consists in asking how each type of asset is or was financed, and then investigating details whenever debt is mentioned as a form of financing. Additional questions are then needed to cover loans that do not go directly toward a specific asset, including the reason why they were taken out. Compared to the former measurement strategy, this one has the advantage of giving a clearer picture of how households plan and carry out the acquisition of assets; however, it generally entails a larger response effort.

Data producers should also envision in advance whether respondents may need help in answering certain questions; if yes, they should predispose cognitive aids for respondents such as cards and glossaries, and integrate information on using them in interviewer training sessions. For example, in the case of Yes/No questions covering different types of debt, it may be useful to provide interviewers with a standard definition of concepts such as revolving credit or bank overdraft.

Once the data has been collected, it must be checked, validated and, where necessary, subjected to imputation procedures before it is fit for the production of estimates. Choices have to be made on editing rules, treatment of outliers, and computation of variability in results. Generally speaking, these choices should be made beforehand for the whole survey, and not on a variable-by-variable basis, in order to achieve methodological consistency.

Finally, the results have to be presented to policy makers and, in some cases, to the general public. Population-level statistics, such as the total share of financially vulnerable households, should generally be accompanied by meaningful information on the distribution of the phenomenon. Depending on the variables available in the surveys and on any external information pointing to problematic population segments, breakdowns by age, gender, education level, household size, employment status and/or sector, etc., and any combination thereof, can be offered to users.

constructed. In the case of wealth measurement, this desire argues for binding the approach to question design as closely as feasible to a general accounting framework, such as that described in Chapter 3.

When establishing a survey that will be carried out regularly, as opposed to a one-off study, data producers should choose the frequency based on the characteristics of the target concept. An additional consideration is whether a repeated survey should be executed as a repeated cross-section or as a sequence of interviews with a fixed panel, possibly supplemented with additional elements to compensate for population changes since the formation of an initial panel. Repeated cross-sections can provide good estimates of changes in characteristics of population groups over time. In contrast, a panel (longitudinal) component may be desirable if changes over time at the level of individual households figure importantly in the desired estimates, or if other statistical concerns motivate repeated observation. Obtaining estimates that are representative both of a panel

and of the population in periods after the initial panel formation typically requires supplementing the panel observations with elements that were either not present at the time of the panel formation or were present but in a different proportion in the population.

The sample should always be selected according to a probabilistic scheme, i.e. each unit in the population should have a known *ex ante* probability of being selected. Only in this case will the survey estimates have good statistical properties. Because such properties are undefined for non-probabilistic samples, it is usually not possible to describe scientifically what estimates based on such samples represent, or to provide meaningful measures of precision for those estimates.

A tolerable level of error for each of the key estimates should be agreed upon with the researchers or policy makers requesting the information, subject to any cost constraints. Survey error is a consequence of both *sampling error* and *non-sampling error*. Sampling error is a consequence of making estimates on the basis of a sample, rather than on the entire population. Non-sampling error is a consequence of non-response, conceptual error, reporting error and processing error. Once an error tolerance has been set, the minimum sample size compatible with it and with cost constraints must be computed, exploring various possibilities until an optimal sampling plan has been determined. Particular care should be taken in making realistic assumptions about the response process and the full range of survey costs. If a sample design cannot satisfy both the desired error tolerance and the budget constraint, the project might have to be reconsidered: narrowing the scope of the survey, for example, might be preferable to delivering a large quantity of inaccurate results.

A *main sample* should be drawn, with a size equal to the target size, supplemented by a *reserve sample* large enough to substitute non-responding units based on reasonable assumptions on response rates. For example, if the target size of the sample is 1 000 households and a response rate of 50% is anticipated, the total sample should comprise 2 000 households. Some variations on this approach are dealt with later in this chapter.

The quality of estimates starts with the quality of the raw data. Questionnaire design and implementation, interview mode, interviewer selection and training, economic incentives offered to participants, and real-time quality control methods are critical contributors to data quality, and each should be considered carefully.

Audit activities should be carried out both during the fieldwork phase and after its conclusion. In all cases, data producers should have a clear monitoring scheme covering contact activities, refusals, substitutions, the contents of completed interviews, and any data manipulation taking place prior to transmission to the agency sponsoring the survey. If data collection is not outsourced, a third-party auditor should be involved in the process. As a part of audit activities, a share of the sampled households should be re-contacted in order to verify the truthfulness of interviewer statements.

When the results are released, measures of variability should be published, accompanied by a non-technical explanation of what these measures mean. If a micro-level data set is released for research or public use, it should contain information that allows users to compute the variability of their own estimates.

As an ethical requirement and sometimes a legal requirement, a clear programme for protecting the confidentiality of the data collected must be developed and implemented. In some cases, a plan must be put in place to further restrict the use of the data; for example, the data might be allowed to be used only for non-commercial purposes.

A thorough and continuing programme of evaluation of all steps in the survey should be instituted. Systematic evaluation enables quality improvements as well as the detection of changes in the behaviour or opportunities available to the population.

## 6.2. Measurement issues specific to wealth surveys

Three broad categories of issues are of particular concern to survey designers interested in the measurement of household wealth, apart from the more detailed issues discussed in the following sections of this chapter.

● *Distribution-related issues*. Wealth inequality differs across countries, but where it has been measured, it exceeds inequality in the distributions of income and expenditure (Figures 6.3 and 6.4). To obtain meaningful estimates of many wealth statistics, it is important to have an accurate representation of the entire spectrum of wealth. In practice, the greatest difficulties are in obtaining sufficient observations of the two extremes of the wealth distribution. Households with very low wealth sometimes see little relevance in participating in a survey about wealth. Although under-representation of these households may have little effect on estimates of totals, it would result in bias to many other estimates, particularly those related to inequality or credit use. Effort should be devoted to instructing interviewers and respondents about the importance of interviewing these households. At the other end of the spectrum, very wealthy households may be extremely difficult to contact and when contacted they may be difficult to persuade of the value of participation in a survey or that information collected in a survey could not be associated with them. Although such households are small in number, they own a large share of total wealth. Thus, under-representation of these households would have detrimental effects on many wealth-related estimates. Great persistence and other special efforts may be needed to reach this part of the population successfully, and such efforts are most often expensive. Effort should be devoted to developing measures of effort applied to all parts of a survey sample and to ensure that these measures are used to learn both about potential biases in the final sample and the most cost-effective means of reaching and persuading respondents, particularly those at the two extremes of the wealth distribution. Evaluation efforts toward this end should pervade the entire measurement process.

● *Data quality issues*. Respondent error may be particularly high in wealth surveys on account of several factors: i) *deliberate misreporting of assets or liabilities*, out of security concerns about the use of the data or social desirability considerations; ii) *misreporting as a consequence of* cognitive *difficulties in recalling information,* such as recall or framing bias; iii) *reporting incomplete or outdated information*, particularly when a respondent is answering questions about another member of the household; iv) *lack of clarity for the respondent in the questions asked or in any instructions given*; v) *unwillingness of the respondent to consult records*, often owing to time constraints or mistrust of interviewers; or vi) *respondent fatigue*, particularly near the end of the interview or in sections with a high degree of technical complexity. The failure of interviewers to follow instructions or to probe for clarity in ambiguous or obviously incorrect responses may also contribute to reduced data quality; commonly, interviewers may not press respondents for clarification, feeling they might refuse to complete the interview. Thorough testing of the questionnaire, provision of automated data evaluation during the interview, rigorous interviewer training and evaluation, and efforts to build trust with respondents may all serve to offset some of these sources of error. Careful review and evaluation of the raw data and supporting

procedures may provide insights into how to improve the procedures for subsequent surveys, in addition to possibly identifying errors in the data.

● *Privacy and confidentiality issues.* For many people, wealth information is considered very private and often not even to be shared among members of a household. Experience suggests that such data are typically considered more sensitive than information on income or sexual behaviour. In addition, some respondents may be so conscious of the possibility of identification that they may not even want to provide demographic information. Given the special sensitivity of wealth information, survey designers should begin with a credible plan for protecting the confidentiality of respondents' data. Such a plan is helpful not only in persuading respondents to co-operate, but also in convincing interviewers that they are not betraying the people they interview. A strategy should be developed to address respondents' initial concerns about legitimacy and confidentiality directly and clearly and to reassure respondents as necessary throughout the interview and beyond. Once the survey is completed, it is essential that the survey designers rigorously execute their plan for protecting the data. Even if only tabular data are released, there may still be important data confidentiality concerns to be addressed.

Figure 6.3. **Income and wealth distribution in the United States, 2007**

Percentages held
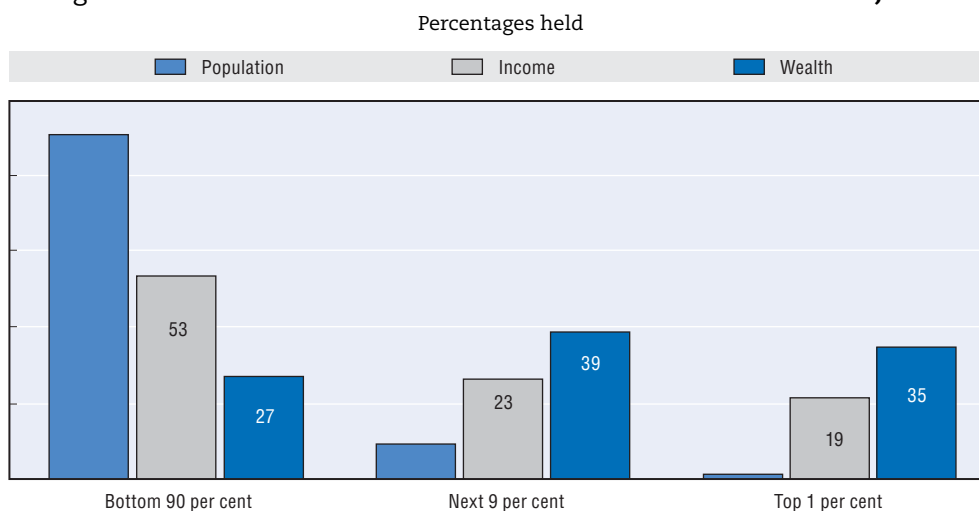
Source: Wolff, E.N. (2010), "Recent Trends in Household Wealth in the United States: Rising Debt and the Middle-Class Squeeze – An Update to 2007", *Levy Economics Institute Working Papers Series,* No. 159; graphics by the Economic Policy Institution, 2010.

6.3. Survey development and data collection

The development of a household wealth survey and the collection of data based on it requires addressing a number of different issues, such as the choice of survey scope and , cross-sectional or panel, content, frequency, sample design, execution protocols, etc. Guidance on each of step is provided below.

6.3.1. *Survey scope*

The scope of a survey depends on the needs of the sponsor. Colleting a variety of variables other than wealth enables a broader range of analysis of the effects of household characteristics on wealth, and on the effects of wealth on other household characteristics. However, there are limits to the willingness of most respondents to answer survey questions.

Figure 6.4. **Income and wealth distribution in Italy, 2008**
Estimated densities
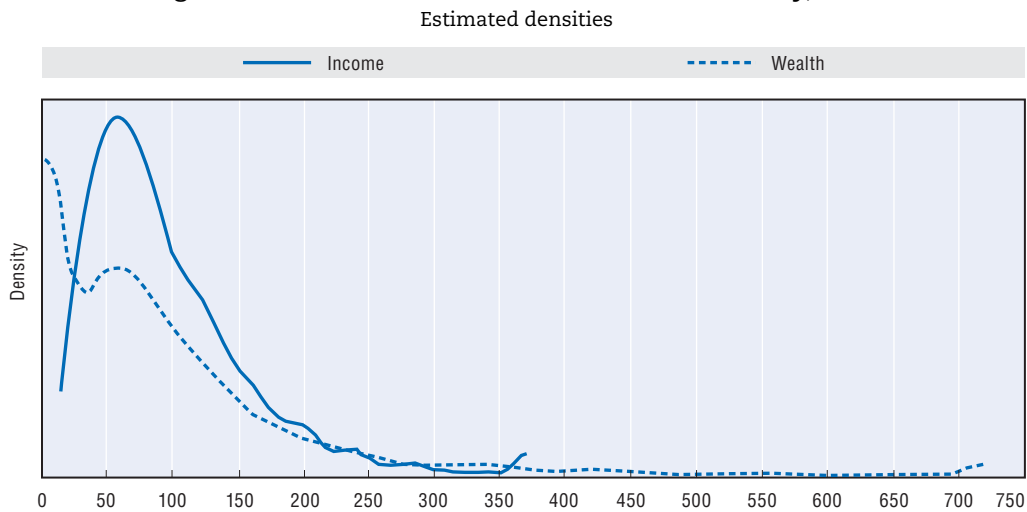
Source: Own calculations based on data from the Bank of Italy's Survey of Household Income and Wealth (2008).

Many data producers run multi-purpose surveys on the economic behaviour of households, with separate modules covering demographic characteristics, income, employment and wealth; in some cases, expenditure questions are also asked. This solution is chosen frequently in an attempt to strike a balance between cost and serving a variety of informational needs. Within multi-purpose surveys, the relative weight of different topics may vary a lot: for example, wealth is the predominant concern in the US Survey of Consumer Finances, while it takes up only about one-fifth of the questionnaire in the Dutch DNB Household Panel, alongside other topics such as health conditions, psychological attitudes and income. A higher focus on wealth allows for more intensive efforts to gear sampling plans and fieldwork efforts toward achieving a balanced representation of the entire wealth distribution by oversampling wealthy households to support the investigation of narrowly held items and to improve the precision of overall wealth measures.

6.3.2. *Cross-sectional or panel observations and survey frequency*

Cross-sectional and panel surveys provide very different estimates. Repeated cross-sectional surveys provide good information for groups in existence at the time of each survey. However, because some groups may change composition over time, it is only possible to talk about changes for members of groups in different periods. Panel surveys make it possible to characterise changes at the level of individual households, which may be aggregated into groups using the characteristics of any given period. Panels also allow for more complex treatments of error structures in formal models. However, because the household population changes over time – through immigration, emigration, births and deaths in households – estimates from a panel for a given period may not represent the state of the full population at that period. Some survey designers attempt to approximately "refresh" their panel samples with new cross-sectional observations that are followed in subsequent panel waves. Another potential problem with panels is that respondents may lose interest and learn how to avoid being asked various types of questions.

A survey might be conducted only once or be repeated over time (as a panel, cross section or mixed type of design). A one-time survey may be adequate for assessing a particular situation – for example, ownership of hedge funds. However, when a survey is

executed only once, there is generally no adequate basis for comparing survey estimates. In addition, there is much learning-by-doing in any survey, and that knowledge would be discarded in a one-time survey. Multi-wave surveys of any sort provide a series of comparisons and allow progressive learning about successful strategies in the survey. The time interval between waves of a repeated cross-sectional or panel survey should be a function of the expected change of the items that are measured, the desired statistics from the survey, and the expected minimum detectable level of change. The sample size and design determine the sampling error of the estimates, where non-sampling error is likely to inflate that amount. Typically, wealth moves relatively slowly overall, but the rate of change at a more disaggregated level may be more pronounced, and there may be substantial and abrupt changes, for example when an asset bubble bursts. The relationships among variables may also change significantly, even while some statistics might not change much (see Box 6.2). The researchers or policy makers requiring the data must be given the information necessary to determine the trade-off between cost and the likelihood of meaningful change and use that information to set a survey frequency. Setting a regular frequency has the advantage of allowing both straightforward comparisons over time and a principled approach to evaluation, production and analysis.

Box 6.2. **Choosing the frequency for a survey of household participation in financial markets**

While several statistics on financial assets, both at the macro and the micro level, are available through surveys of financial institutions, the additional information (e.g. on household size and composition, education, income, employment) afforded by a household survey is relevant to tasks such as efficient planning of taxation, estimation of the impact of market events on the economic condition of households, and the selection of appropriate targets for financial literacy campaigns.

Survey modules on financial assets tend to induce relevant costs: even when they are limited to basic "Yes/No" questions on ownership of certain financial instruments, they require a high degree of technical preparation on the part of interviewers, and a considerable response effort on the part of certain categories of respondents. For example, in some countries, middle-income households frequently buy financial "packages" from their bank, i.e. bundles of assorted assets with a given expected yield; when the contents of such bundles are described in the contract signed with the bank, it is often forgotten by household members after a few weeks. If the survey questions go beyond simple ownership and probe the value of each asset, numerous evaluation issues emerge (see Chapter 3); cognitive difficulties also may come into play, ranging from a lack of information about current market conditions to recall bias.

Based on these considerations, and given that data producers normally work on a tight budget, surveys of household participation in financial markets are not carried out frequently. Most happen every two or three years, and as modules in broader-ranging surveys. This may be a problem, however, considering that markets can be volatile; in countries where investments in financial assets represent a significant part of household saving, failing to observe the effects of a market boom or crash for months or years may result in severe misunderstanding of the economic conditions of households.

One way to strike a balance is represented by the deployment of special survey modules whenever a major event happens, while keeping a low frequency in ordinary times: for example, the recent financial crisis was monitored by the Federal Reserve through a one-off

---

Box 6.2. **Choosing the frequency for a survey of household participation in financial markets** (cont.)

panel wave of the Survey of Consumer Finances, an operation that was costly for both data producers and respondents, but necessary in the light of the magnitude of market fluctuations and their effects. Another possibility consists in interpolating regular survey waves with low-cost intermediate modules, administered only to a part of the sample, and limited to qualitative questions that are relatively easy to answer and aimed at updating information on the type of assets held and at understanding whether the value of each asset increased, decreased or stayed roughly the same compared to the previous observation. Also, models can be built based on a combination of household survey data with contemporaneous data from financial institutions; as the latter are typically more frequent, on account of supervisory requirements, they can constitute a basis for estimating some household-level variables in intervals between two survey waves. This operation should, however, be performed with great care, considering that periods of financial turmoil are often accompanied by changes in the structure of relationships between variables.

As noted earlier, a survey must start with a set of objectives, from which nearly every element of the survey process should flow. The connection is particularly obvious in the case of the survey content. When elaborating the detail associated with the objectives, data producers should balance meeting the analytical goals, keeping within budget, and making the survey experience sufficiently tolerable to respondents. Measurement of wealth requires the use of a balance-sheet framework, such as that described in Chapter 3. Particular problems may arise when trying to obtain information about assets or liabilities that can be classified in a variety of ways. For example, in many countries, people may contribute to tax-deferred retirement accounts, and those accounts may be invested in a variety of assets; although a complete balance-sheet classification could be made for all items held in such accounts, if there are dominant investment modes – e.g. mutual funds – it may be sufficient to ask briefly about the direction of investment. Knowledge of additional attributes of balance-sheet items is often quite useful, but care should be taken not to reach a level of detail that would be too thinly supported by the data. In general, an additional question is generally worth asking only if it adds information or enables the calculation of more complex estimates that would otherwise not be available. For example, asking households for the model year of each car owned might be redundant if a public register of cars already exists in the country. But if there were a specific analytical interest in such data – e.g. to track and project household inventories of vehicles – or the data were useful in estimating other values – e.g. to match to auction prices of used vehicles or to use in imputation models – then it may be useful to collect such data. Similarly, including a question about financial derivatives unknown to the large majority of the population would require asking every household about ownership of such items, yet risk not having enough observations of respondents who own such assets for their answers to be statistically meaningful. Often, rarely held items can be collected either as part of a higher-order aggregate or in questions designed to capture any items not explicitly enumerated.

For many analytical purposes, it is not sufficient to have only balance-sheet data in a wealth survey. At a minimum, a variety of household characteristics are likely to be important. Because work and income are usually very central to understanding patterns of wealth, variables describing at least the basic outlines of these topics should also be included.

Questions about attitudes, expectations, institutional relationships, indicators of financial sophistication, past or expected inheritances and other factors may also be relevant.

6.3.3. *Design and development of the survey questionnaire*

To design a questionnaire, a list of output variables should be defined and those variables should then be translated into preliminary versions of the survey questions. The questionnaire must be clear and parsimonious, so as to minimise the cognitive effort required by respondents and to save time. Key to this effort is the sequencing and framing of sections and individual questions to exploit the range of understanding and order that respondents would bring to the interview. For example, as mortgages are typically tied to the purchase of real estate, it may make sense from a respondent's perspective to ask about such items together. In general, the order should minimise the extent to which a respondent is likely to feel that the same subject is being addressed in more than one place, as switching can lead to respondent fatigue or to error.

The same questions must generally be used for a broad set of respondents. Thus, it is important to develop a questionnaire that is no more technical in content that is required to approach logical coherence. Because there is sometimes a trade-off between logical coherence and other sources of measurement error, it is important to minimise total error for the survey objectives, rather than only a local source of error at the cost of introducing or exacerbating others. *Cognitive testing* is conducted by a moderator (sometimes a cognitive psychologist) with one person to shed light on both the understanding survey participants have of questions and the thought mechanisms behind the answers they give. Where sufficient resources are available, cognitive testing may be used to refine the design of sequences of questions.

Ideally, the reference time for the items reported in a wealth survey might be a fixed time, such as the end of the previous calendar year. For a variety of reasons, there is no unanimous agreement on this point; some believe that use of a fixed time as reference period leads to better data quality, because of a lower recall bias. Wealth holdings often have a high degree of persistence, but their value may change abruptly. Because fieldwork activities may continue for a substantial period, values at the time of recording would reflect different economic conditions at different interview times, though it might be possible to adjust values for analytical purposes using market indices. A potential disadvantage of a fixed reference time is that this time point may not be salient to respondents if it is many months in the past; in that case, additional measurement error might be introduced for respondents who knew the approximate current values but not the value as of the reference time. However, for co-operative respondents with complete records who are also willing to consult those records, use of a fixed reference time would not be problematic. Ultimately, the survey designer must consider the trade-off between variation due to a variable reference period and variation due to additional measurement error for respondents who must estimate their answers.

Particular care should be taken to allow variant text depending on previously provided answers; for example, if the respondent is a married female and her spouse is a male, questions about the spouse's employment should use the correctly gendered pronouns or other language adaptations that may be relevant. Introduction of variant text is more straightforward with the use of computer-assisted interviewing (CAI), including computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), computer-assisted web interviewing (CAWI) and related technologies. Where it is known in

advance that certain questions may arise in the interview, an effort should be made to place relevant answers or instructions directly into the questionnaire for use by the interviewer at the appropriate moment. References to any supporting materials aimed at interviewers (such as a glossary) or at households (such as a supplemental card used in defining appropriate answers) should be included in the question text or clearly spelled out alongside the relevant question. For CAI, conditional routing instructions should be encoded directly in the software; for paper-based questionnaires, such instructions should be made clearly visible through the use of bold typeface, arrows, colours, etc.

Although the goal of an interview is to gain complete information from each respondent, this is not always feasible. In some cases, the respondent may have only partial information, but that information may still be helpful either in the analysis or imputation of missing data. Thought should thus be given to providing alternative means of reporting some types of information. For example, a respondent may have no information about the specific types of mutual funds the household owns, but still have a reliable value for the total amount they own. A common problem in wealth surveys is the inability or unwillingness of respondents to provide an exact answer to a question denominated in a monetary amount. Thought should be given to allowing respondents to report range information – using their own range, a value from a range card, a value determined by progressing through a decision tree, or a combination of these options.

It is generally the case that not every possible category can be included for questions that have a categorical answer. Including too many answer categories may confuse both the interviewer and the respondent and decrease the reliability of the answers. However, including too few categories may lead interviewers or respondents to classify some responses inappropriately, unless such responses are far outside the categories included. Except where there is a relatively small and exhaustive set of answers – e.g. *"yes"* or *"no"* questions – a residual *other* category should be included. Where resources are available, that residual code should be specified by the respondent, and the answer entered verbatim to be recoded after the interview.

It is desirable to push efforts to enforce data quality into the interview process to the extent that doing so does not create serious offsetting problems. Although this process is more straightforward with CAI, it is possible to take similar steps with a paper questionnaire. In the case of CAI, when programming the computer interface data producers may implement three checks on the information input.

Hard checks, *soft checks* and *confirmations*. Hard checks may be programmed in the form of acceptable ranges for a given variable, acceptable ratios between two variables, or a variety of other logical tests. For example, a respondent in 2010 might say he is 30 years old, but he first worked for wages in 1970. Failing a hard check should cause the computer to generate an error message to the interviewer and to reject the value first entered. In principle, hard checks should eliminate all values that are deemed impossible. However, care should be used in instituting such checks, since there may be very unusual occasions where the value might be correct and the respondent (and interviewers) may become alienated as a result of disallowing the answer. Soft checks may be used alert interviewers to values that are very unlikely, but not impossible (Box 6.3). In practice, the computer might simply generate a warning to the interviewer or require a short comment before allowing the questioning to proceed.

Box 6.3. **Soft checks for real assets**

Measuring the current value of real assets can sometimes be difficult, especially in the case of households who own only their primary residence and have either acquired it in the distant past or inherited it. Such households may not be aware of market conditions, and they might attempt to answer questions based on outdated prices. The implementation of soft checks in a CAPI program can be very useful for these cases, and different levels of sophistication can be achieved based on information available to data producers. Surveys of real assets normally include a simple question on the surface area of each dwelling; this is a solid base for checks because it is not as prone to measurement error as other indicators. Most households have an idea of their home's size and, if they don't, interviewers can easily walk them through the production of a rough estimate. Once a reliable measurement of size is achieved, a question on current value can be asked: computing the ratio between the answer given and the dwelling size yields a price per surface unit (square meter or square foot), which can then be compared to average prices for the area. In some countries real estate transactions are a matter of public record, which means that data producers can assemble a detailed database of reference prices for each geographical subdivision and use it to benchmark responses, instructing the CAPI program to generate a warning message every time the ratio falls outside a reasonable interval, e.g. between the 5th and the 95th percentile of the distribution of prices for the area. If such a possibility does not exist, other sources – such as private databases produced by real estate agents or classified ads in specialised publications and/or the generic press – can be used to construct a reference database.

● Depending on the complexity of the survey, checks can be even more fine-grained, taking into account the quality of the building as observed by the interviewer, the number of bathrooms, the type of heating system, the presence of certain amenities, etc. Once checks have identified a value as potentially unrealistic, interviewers should proceed to confirm the number reported with the household in case a reporting error has been made.

● These checks should allow for specific situations where there is a reason for anomalous values, such as proximity to a source of loud noise or pollution, or recent changes in the area that have not been discounted in recent transactions yet.

The confirmation of values may lead to higher data quality for some variables. Interviewers may routinely repeat what they have heard the respondent say, but computer-driven confirmation ensures that what is entered into the computer is what the respondent intended. For example, the respondent may provide a monetary answer such as the purchase price of the household's home but using a currency that is no longer used (such as the Italian lira), rather than the current unit of account (such as the euro); the computer can be used to generate the words implied by the numbers entered and ask the interviewer to repeat the figure to the respondent, framing the number as an amount in the current unit of account. This approach may also be useful in confirming values computed from other values, such as the respondent's age given that person's birth date.

Where possible, it may be helpful to introduce a facility in the interview for the interviewer to record comments made during the course of the interview. Similarly, it may be helpful to provide a place for interviewers to make a systematic record after leaving the respondent of any case-specific problems, whether with respondents or the survey instrument. This information may assist both in reconciling values in individual cases and in improving subsequent waves of the survey.

The questionnaire, whether CAI or paper, should be extensively tested by developers and interviewers, and then by a small sample of the general public selected for a *pilot survey*. The pilot survey does not serve the purpose of collecting data, but rather of getting feedback from interviewers and respondents on any difficulties arising with specific questions. The questionnaire should then be modified to take this feedback into account and subjected to another pilot survey if the changes are substantial and time and resources permit.

6.3.4. Sampling design: General principles

Precise knowledge of population characteristics can be gained only if data are collected from all units in the population and in the absence of measurement error. However, reaching every unit (household, firm, bank, etc.) with a survey may be expensive, and gathering complex information from such a large set of respondents in this way would also be extremely time-consuming. For these reasons, in most countries a full census is carried out infrequently and limited to basic demographic variables. In some places, administrative records make it possible to observe important data for all or a sufficiently large part of the population that can substitute for surveys to a degree.[1] The utility of survey statistics rests on the fact that meaningful information can be obtained by applying scientific procedures to select an appropriate subset of the population, a *sample*, and by collecting data only from this subset. While such information is not exact, statistical procedures can be used to characterise and manage the level of uncertainty associated with sample-based estimates.

Effective sampling design entails a sequence of activities that lead to the selection of an optimal sample, i.e. the population subset that yields the best results for a given budget, or that minimises survey costs for a given degree of variability of the estimates. As argued above, data producers should start with a clear definition of three elements: *i)* the key target variables; *ii)* the acceptable level of uncertainty; and *iii)* the budget available. Statements should be formulated clearly, such as, "The sample must be selected in order to estimate average net wealth, average income and average expenditure at the regional level, with a 3 per cent standard error, within a USD 500 000 budget". As a general principle, respondents should be chosen in such a way that they parsimoniously and accurately represent the range of variation of the target variable(s) within the population.

The problem with this criterion is that in most cases the survey is run precisely because there is only limited unit-level information on target variables. Moreover, because it is usually expensive to construct a sample frame, surveys often use general-purpose frames that may contain little information directly related to the objectives of the survey. In such circumstances, data producers may have to rely on external sources in order to get an idea of what easily observable factors may be roughly related to variations in the target variable(s) across the population. Similar surveys carried out in the past are one potential source for defining meaningful subgroups. The available information should be used in order to plan the *sample stratification*, a partitioning of the population into groups (*strata*) that have the highest possible *variance between groups*, (i.e. they are as different as possible from each other) and the lowest possible *variance within groups* (i.e. they are internally homogeneous). Stratification variables should be discrete, free from measurement error, simple, and compatible with the sampling frame. Further, they should identify cells that are not too small.

Geographical indicators for large sub-national areas (e.g. states or counties in the United States, NUTS-1 regions in Europe) are often an acceptable choice in defining strata: in most countries regional variations exist in living conditions. By the same token,

municipality size may also be an acceptable choice. Other variables could be suitable in some countries but not in others, depending on the relevance to the survey topic, e.g. immigration status, linguistic group, birth cohort, gender, average level of house prices in the region/province/cluster of municipalities, average number of small businesses in the region, etc.). Sometimes a sample may blend more than one approach; for example, in the United States many surveys partition the sample universe into municipalities and groups defined in terms of other geographic units.

In very small countries, stratification might not be necessary or feasible; in that case, units may be chosen through simple *random sampling*, a direct draw from a complete listing of relevant population elements. Conversely, because it is often infeasible to send interviewers everywhere in large countries, a *multi-stage* design may be used to select a sample from a limited number of areas: first a certain number of primary sampling units, generally corresponding to geographical locations (municipalities or districts), is randomly chosen for inclusion in the sample; then households are drawn within each location. The general principles of stratification apply but are adapted to take the number of stages into account. Once strata have been defined, data producers can see whether their design is compatible with the budget and the estimated quality constraints.

Standard formulae are available to determine the optimal allocation of units to strata and to derive either the minimum number of units for a given variability of the survey estimates or the minimum variability for a fixed number of units. Any information on stratum-specific expected response rates *should* be incorporated in the calculations performed to determine stratum size. When an acceptable sample design has been chosen, specific units should be chosen randomly from each stratum. Where a survey must use a predefined set of sample areas or subgroups, as might be the case when using an existing sample frame, the only option available in sampling is to vary the rate of selection from the various predefined domains.

As noted above, allowance must be made for the expected level of unit non-response in setting the total sample size for a survey. If the achievable level of response is known *a priori*, the most straightforward option would be to draw a main sample and reserve sample of a size large enough to achieve the target number of interviews. If the response rate is not known *a priori*, another option is to divide the reserve sample into *sample replicates*, where each replicate is a smaller version of the main sample. As the rate of non-response becomes clearer during the field period of a survey, replicates can be introduced individually up to the level needed. The replicate approach helps to control costs and to assure that all observations in the sample are exposed to approximately the same level of effort. In some cases, a *substitution sample* is selected, where the substitutes are defined at the case level (i.e. if a particular case does not participate – typically after a fixed maximum number of attempts – a designated substitute is used; that substitute may be determined by a fixed unit specified in advance or a by mechanism specified in advance). Substitutes can be helpful in achieving the desired number of interviews and in controlling costs, but they introduce an element of uncertainty about what the achieved sample represents. Presumably, there are several reasons why some households participate in a survey and others do not; if there are systematic components to non-response, then using substitutes introduces an element of unobservable selectivity into the final set of respondents. In some situations, it is possible to match original sample units and substitutes based on a large range of key attributes, so that the potential selectivity may be tolerable. Whenever substitutes are used, it is recommended that the data creators make every effort to identify

systematic elements introduced by the substitutes; one way of doing this might be to compare weighted estimates made from the set of respondents excluding the substitutes with those from the full set of respondents. The use of substitutes should be fully documented in the data, and proper account should be taken in reporting response rates.

6.3.5. *Sampling design: Wealth surveys*

When designing a sample to measure household wealth, the specific challenges described above should be kept in mind. Sufficient representation of households at all wealth levels should support the desired estimates. Care should be taken that the least wealthy part of the population is properly represented. To obtain reliable measures of the upper tail of the wealth distribution, it is likely to be necessary to *oversample* wealthy households, except perhaps where it is feasible to have a very large sample. Relatively wealthy households account for a disproportionate share of the total wealth, and existing evidence suggests that the likelihood that they will not complete interviews when included in a sample is disproportionately high. Thus, there are potentially both bias and variance implications stemming from the treatment of wealthy households. Standard designs used when measuring income or expenditure might not be adequate for measuring wealth. If external information about the distribution of wealth exists, it might be employed to stratify the sample. Constructing a list of wealthy households may be easy in countries with accessible administrative records on wealth or income from wealth, though it might be necessary to study both the degree to which administrative definitions correspond to the desired ones and the accuracy of the administrative data.[2] However, only a minority of countries have such data available for use in sampling. Other possibilities include the assessed value or taxes for real estate, vehicle registrations, electricity consumption, etc. However, such information may be only loosely related to overall wealth. In cases where information is available only at a higher level of aggregation, for example property taxes by area, there may be so much variance within areas that disproportional sampling of high-tax areas would increase the estimation variance. There is often no information sufficiently related to wealth. Where only weak proxies are available or where no such information is available, attention should be given to estimating the effects on key estimates of potential misrepresentation of the wealthy.

In some cases where there is partial information in more than one sample frame, a more complex design may be desirable. For example, one strategy might be to use two separate sampling frames, one with no wealth information to sample respondents in general, and one with wealth information but incomplete coverage of poor people for oversampling relatively wealthy respondents. If the defects in the two frames are sufficiently well known, the two samples could be combined through weights, or the estimates from the survey could be made using a dual-frame estimation procedure.

Although oversampling of the wealthy, when done effectively, may lead to improvements in data quality for wealth measures, such improvements come at a cost. The general experience in countries where such oversampling is possible is that it requires a much more substantial effort to reach and persuade a very wealthy respondent to participate than is the case for a middle-class respondent. One explanation of such differences is the frequency with which very wealthy people employ other people to protect them from unwanted intrusions. Because many surveys must serve multiple purposes, it may make sense not to oversample the wealthy, but to select a larger general sample and tolerate higher variance and possible bias in wealth measures, if there is enough gain for the other purposes of the survey.

6.3.6. *Scope and coverage*

Household surveys on economic themes generally target the non-institutionalised resident population, i.e. people living in an independent accommodation and residing in the country at a certain time. The exclusion of people living in institutions follows both from practical considerations of reaching and interviewing people in such restricted environments as well as from the idea that people living in institutions do not have the same degree of decision-making authority as people living in independent households. If these populations grow over time, it may be necessary to re-examine the implications of their exclusion; also, in some cases such people may already be included in principle as absent members of survey households. The definition of a non-resident may also vary across countries, but the idea is that transient populations generally should not be included; in principle, non-residents might be included as an absent member of a household in the wealth survey of another country. Because there is some imprecision in the definition of both institutionalised people and non-residents, these populations are likely to contribute to non-sampling error.

Appropriate coverage of the target population generally depends on the quality of the *sampling frame*, i.e. the structure from which sampled units are drawn. A probabilistic sampling design requires that each unit has an *ex ante* known selection probability: ideally, this would be best attained by drawing names from a comprehensive list of households. In most countries, no such list exists; at best, the closest substitute would be a census-based list of individuals, which might include duplicates and might be at least somewhat outdated. If a list-based approach is adopted, particular care should be taken to anticipate and prevent incomplete coverage, that is, a situation in which certain groups of eligible households are omitted. Auxiliary frames covering certain segments of the population, or alternative sampling techniques, can be of help with under-coverage; they should, however, be used with a measure of care (see Box 6.4).

Where lists of households or individuals are not available, address lists may be a satisfactory substitute. Options that also include a name attached to the address, e.g. lists obtained from post offices, are to be preferred to options that refer only to a residential dwelling at a given address. The use of address lists, besides complicating the calculation of selection probabilities, may induce coverage problems for rural and other sparsely populated areas; GIS-based maps identifying dwellings, where available, could provide useful auxiliary information. In any case, an effort should be made to understand and document the degree to which the sampling frame over- or under-covers the target population.

6.3.7. *Collection methods*

Several data collection methods can be employed in household surveys. Some involve the participation of an interviewer, who either visits the respondents at their residence (face-to-face interview) or talks to them on the telephone, noting down answers on a paper questionnaire or storing them in a computer's memory. In other cases, questionnaires designed for self-administration are made available to respondents on the Internet, sent through surface mail or left by field representatives. Sometimes a self-administered questionnaire is used as a supplement to an interview mediated by an interviewer. In the case of wealth surveys, personal interviews conducted by experienced, well-trained interviewers are particularly recommended. Wealth is generally perceived by respondents to be sensitive information; direct interaction with an interviewer projecting an image of trustworthiness may increase response rates, both in terms of overall participation and in terms of co-operation in answering sensitive questions. Also, certain questions are

> ### Box 6.4. **Sampling highly mobile populations of new immigrants**
>
> In some countries, immigration is a long-established phenomenon. New immigrants join family members or groups of fellow nationals who already live in the country, finding employment through community networks and entering a geographically stable lifestyle from the moment of arrival. In other countries, where communities are not as established and/or where jobs are not as easily found, new immigrants are among the most mobile socio-demographic groups. They may change residence several times per year in the pursuit of job opportunities, e.g. in the agricultural sector or in home assistance to the elderly.
>
> All highly mobile subpopulations create difficulties in surveys, because administrative records of residence are often outdated, and once the sampled units are not found at their official address it may be impossible to track them down. In the case of new immigrants, these difficulties have a particularly large impact, because they may deprive policy makers of information on a key subpopulation for the planning of integration, labour and welfare policies. Auxiliary sampling frames can be of help: in some instances, some records that are produced only for immigrants (residence permits, work permits) are more updated than general residence records, and contain contact information. In other cases, alternative sampling techniques such as snowball sampling can be used, counteracting the lack of official information with data gleaned from informal networks. In both cases, however, care should be applied in estimating the effects of integrating this information on estimates. Assembling auxiliary sampling frames based on work permits, for example, will probably result in the overestimation of income, consumption and even wealth, because those who do not produce income are automatically excluded. Snowball sampling or variations thereof, being based on personal recommendations, result in over-representation of tightly knit groups, which could in turn be correlated positively to economic conditions. In general, every deviation from a clear plan of probabilistic sampling will induce alterations in the characteristics of estimates; while this may sometimes be necessary so as not to incur serious information gaps, it should be accounted for when releasing results.

technically complex, either because they concern items that are difficult to evaluate (e.g. a dwelling bought 70 years ago and received in inheritance by a 30-year-old respondent) or because they mention concepts that are not very familiar to the general public (e.g. hedge funds). Giving respondents a chance to ask the interviewer for clarification may reduce misunderstandings. Interviewers are also able to observe the progress of the interview and to record notes that may be helpful to the data creator during the later review of the data.

Whether or not an interview is directly mediated by an interviewer, in the case of wealth surveys it is recommended that at least the core measurements should be carried out using a computer-assisted interviewing (CAI) method. With a paper questionnaire, it is very difficult to implement a detailed and consistent review of the most essential aspects of the data at the time they are collected, which is the point when correction is most straightforward; in addition, such questionnaires generally require a separate data entry stage, which provides an opportunity for misinterpretation of handwriting and other transcription errors. In contrast, with CAI the underlying computer program can be set to conduct real-time checks on the plausibility of individual items and on the overall consistency of the responses obtained; such checks can also help in detecting errors made by interviewers in recording respondents' answers.

6.3.8. Collection unit

The most common collection units of observation for survey data on wealth are the individual, the household or a concept of economic family. The latter two types of approaches typically use a reference person to provide most of the data about the wealth of the unit, and sometimes allow for reporting on more personal forms of wealth by the relevant individuals. There are trade-offs in these approaches. Individuals may have the best understanding of their own pension rights, their credit card use and other such items, but it may be less clear to them how collectively owned items should be valued and how ownership shares should be distributed; indeed, in some cases ownership shares cannot be known for certain without breaking up a household. Thus, recording wealth solely at the person level introduces the possibility that some items held within a household may be omitted altogether, and that others may be counted multiple times.

Households are often straightforward: a single individual or one couple, with or without children. However sometimes, and particularly in some countries, households may be complex, containing multiple generations and multiple simple families within generations. In such situations, a large variety of ownership patterns may be present, and a similarly large array of potential measurement errors may be present as a consequence. In addition, where complex households are present in a relatively high degree, compositional differences across households cause variations in wealth within and across countries, obscuring the effects of underlying life-cycle factors and other behavioural patterns. Generally, some allowance for household composition is appropriate in the analysis of household-level wealth data. A close alternative to the household as a unit is a subdivision of the household, defined in terms of economic independence; if such economic units were sampled within households, comparison across households and across countries might be simplified. Nonetheless, because even such units vary in composition, allowance should be made for compositional differences. Most of the discussion in this chapter assumes that the desired approach is either the household or an economic family within the household.

6.3.9. Reference person

After gaining the co-operation of a household for a survey, a *screener questionnaire* must be administered to determine the person or persons who should provide information during the interview. A household reference person (HRP) is generally selected to collect core information about the composition of the household and the basic characteristics of its members. This person may also be in a position to provide accurate details on shared wealth. Data producers should set a rule for identifying the reference person, and define their role clearly. In some countries, the HRP is responsible only for providing demographic information and may be any adult; questions of an economic nature are instead asked of the financially knowledgeable person (FKP) identified in the screener questionnaire, who may or may not coincide with the HRP. To determine the FKP, the screener questionnaire should contain a question such as, "Among adult household members living here, who is the person most knowledgeable about financial matters?".

Given the logistical constraints of fieldwork activities, and the possible coexistence of different cultural norms in the same country, the rule used for identifying the HRP and the FKP should be context-dependent. If the sample is drawn from a list of names, using the named individual as the HRP might be counterproductive: if this person is not home at the time of contact, but their spouse is and is knowledgeable about the relevant matters, there

is no reason why he or she should not be the HRP. It is generally best to give some latitude to interviewers in deciding the person to serve as the HRP, consistent with rules set by data producers. In some cases, there may be no FKP in a household or the person who might be the FKP is too ill or disabled to do the interview; in such cases, a proxy for that person, who answers the questions on behalf of the household, may be allowed if that person is knowledgeable about the finances of the household. Such use of a proxy should be appropriately flagged in the data set. Interviewers should also be asked to provide *paradata* (see below) on the perceived level of competence of proxy respondents.

6.3.10. *Respondent co-operation*

Effective efforts to gain respondents' co-operation start with clear communication about the purpose of the survey and direct attention to their concerns about participation. Households should know the reasons why they are being asked sensitive questions. Whenever applicable, they should also be persuaded that their willingness to provide data is important for effective policy making; this may be particularly hard when trust in policy makers is low, e.g. during a crisis. Emphasising the importance of the data for longer-term research purposes may also be helpful. The confidential nature of the interviews should be stressed repeatedly when contacting potential respondents and the measures taken to protect confidentiality outlined. If possible, data producers should send a letter to households included in the sample some weeks before fieldwork starts, explaining to them that they have been selected for a statistical study of household economic conditions and that they should expect an interviewer to make contact with them within a given period; the letter should be signed by a public figure recognisable by households and generally deemed authoritative and trustworthy (e.g. the Director of the National Statistical Institute or the Governor of the Central Bank). In general, interviewers should be prepared to deal with a wide variety of questions about the content of the survey and its potential uses. Toll-free telephone numbers and e-mail addresses should be provided to households in case they want to acquire further information or to confirm the information they have been given. It must always be clear to the household that the entire survey depends on the representativeness of the sample and that the interviewer would allow a wide range of flexibility to accommodate the needs or reservations of the household. In some countries, surveys of people must be approved by special committees that are required to monitor the treatment of human subjects in research. Such committees typically insist on clarity in stating the purposes, risks and countervailing protections in a survey and on avoiding actions that might be interpreted as coercive.

Research suggests that offering households a tangible incentive to participate, in the form of a gift or a cash payment, can be very effective in obtaining co-operation. Such incentives must be evaluated in light of local cultural norms. Because adequate monetary compensation for respondents' time and effort is almost always impossible under realistic budget constraints, it is important to emphasise to respondents that the incentive is merely a token of respect. It may be effective in some instances to have the option of offering a donation to a designated neutral charity. In no case should there be a requirement to accept the incentive, but respondents should always understand that their contribution is valued and appreciated.

6.3.11. *Role of the interviewer*

In all household surveys there is an interaction between a person and a technical survey instrument, which usually must be the same for all respondents, aside from text that may be variable depending on previous answers. Introducing an interviewer into that process allows the possibility that quality standards can be maintained through that engagement as the particular situation of the respondent is encoded in the standardised framework. However, the introduction of an interviewer also raises the possibility that the respondent may react to the presence of an interviewer by providing inaccurate information, out of a concern not to appear unusual to the interviewer or some other concern about the interviewer. Interviewers must be trained in the technical skills necessary to gain co-operation and execute an interview, but also to cultivate an image of neutrality as they do their work.

In surveys where interviews are mediated by an interviewer, the selection and training of interviewers is of central importance. Respondents must be persuaded that the interviewer knocking at their door is a trustworthy person whom they can let into their house without fear, and who is not going to use the data provided for any purpose other than the one stated. Identification systems (badges, advance letters/telephone calls providing the interviewer's name to the household, etc.) are necessary, but they are not sufficient. Interviewers should display traits appropriate to the cultural context that enhance the likelihood of gaining the respondent's trust.

The interviewers most successful in gaining co-operation are generally those who are able to project personal empathy with the concerns of the respondent, but who are nonetheless able to remain neutral and non-threatening. Although in many countries interviewers are predominantly female, both male and female interviewers have been successful in surveys of wealth, with one gender sometimes being more effective than the other with certain sub-groups. The managers of interviewers should be attuned to the personal style of each of their interviewers and how they may be used most effectively with different types of respondents. Role-playing in training is often an effective way to teach both those directly involved and those who are observing how to engage with respondents, and it may also provide insights to managers about the relative strengths of different interviewers for interviewing different population subgroups.

Besides demonstrating strong "people skills", the ideal interviewer should also show appropriate technical competence to execute a high-quality interview. Because high levels of such skills do not always appear bundled in a single individual, sometimes it may be advisable to consider a degree of specialisation of tasks; for example, one interviewer might be devoted to gaining co-operation and another might be available on the telephone to conduct the interview. In any case, training should give all interviewers an understanding of the questionnaire, the general economic concepts underlying it and the broad technical aspects of the interview. Interviewers who are expected to complete actual interviews need to know more detail about the economic concepts referenced in the questionnaire, so that they can answer any questions respondents may have, or to have sufficient knowledge to explain information set out in a glossary made available for that purpose. When applicable, they should be trained in the detailed technical operations required to conduct an interview using a computer interface and to be able to deal with basic computer problems.[3]

Where possible, the one-time training should be supplemented with continuing education. At a minimum, interviewers should have regular contact with their supervisor to discuss problems. During the field period of the survey, trends may become evident and

it may be helpful to share information across all interviewers about coping with related situations. As the types of problems shift over the field period, it may be helpful to reinforce relevant lessons from the original training. When it is possible to identify particular problems in the work of individual interviewers, it is beneficial to intervene where possible to re-train the interviewer on the relevant material. In some surveys, a formal system of written feedback is used to provide comments to interviewers about individual cases and to solicit their input on problems detected. Materials developed from an overall view of issues confronted during the field period and posted on a project website accessible to the interviewers can be a cost-effective means of disseminating information.

6.3.12. Survey administration

Ideally, an interviewer might screen a household to identify the FKP and interview that person and any other relevant parties in the same session. Often it is wise to allow respondents time to consider the request for an interview and to check on the legitimacy of the project; prematurely pressing for a decision to participate may precipitate a refusal to continue. Many times the FKP will not be available at the time of the screening, and it is necessary to make an appointment with that person at a later time; if possible, the interviewer should obtain sufficient information for re-contacting the household to confirm the appointment without having to make an additional personal visit. Some respondents may be very pressed for time; making clear to busy respondents that it is possible to break the interview into short segments that can be completed either in person or by telephone may be helpful. Above all, the respondent should feel that the entire process operates at his or her own convenience. In the case of some wealthy respondents, there may be people whose job is to control access to the person; persuading a "gate keeper" may be difficult and time-consuming.

Steps should be taken to develop measures of the effort applied to individual survey cases and to monitor that level of effort to ensure that all the sample cases are given an opportunity to be informed about the survey and to participate. Such process data or *paradata* may be very important in assessing the potential for non-response bias in the final estimates. In some cases it may be possible and desirable to use *responsive design*, a technique that uses information about the sample management to direct effort efficiently to observations that would tend to reduce non-response biases.

6.3.13. Assessment and treatment of collected data

At the close of fieldwork activities, data producers are in possession of raw data, which constitute the basis for the estimation of population parameters. To reach this point with reasonable confidence, the data creators need to review the information for consistency, to deal appropriately with both unit and item non-response, and to apply appropriate procedures to create the estimates and related measures of statistical confidence. At the beginning of the post-survey processing of the data, it is recommended that the data creators generate a shadow variable for each main variable; the shadow variable should be used to keep track of the original status of the data contained in the main variable and any action taken on it.

6.3.14. Data checking, editing and imputation

An appropriate set of automated *data checks* in a CAI program can reduce the number of inconsistencies and errors in raw survey data, but this is not generally sufficient to eliminate all addressable errors. Data review and *editing* are usually required for the data to

be usable. Editing is an organised attempt to bring various sources of information together to assess the plausibility of reported data and to update the raw data (including setting new values or setting values to missing). Sometimes there are instances of incorrect CAI programming that go undetected even in a rigorous testing phase but that become evident in the raw data; inexplicable malfunctions of the CAI software may also generate errors. Timely evaluation of data during the field period can help to detect systematic problems in time to repair the program at that point. Automated checks on the raw data should include verification that the data match the logical flow of the questionnaire (e.g. households living in owner-occupied dwellings should not report a value for monthly rent payments).

More importantly, because some erroneous values can be detected only through complex analysis that cannot usually be performed using the CAI software, core variables should be analysed using both univariate and multivariate techniques to identify outliers. Ideally, suspicious values should always be verified by re-contacting households, but this is usually too costly to undertake for all except the most influential outliers. Data producers must thus make judgments about the validity of anomalous data values. At this stage, it may be helpful to have comments from interviewers, from the main data collection or from debriefing interviews or other *paradata*. Some errors (e.g. reporting values in a currency, adding extra zeroes to a reported number, incorrect use of negative numbers, etc.) are common or regular enough as to be reset mechanically either to a more plausible value or to a missing value. Editing should be random, with consistent application of the principles guiding the work. Conservatism in over-riding reported values should be emphasised. General guidelines on data editing are provided by several National Statistical Offices; international projects such as Euredit (EU) also exist, and should be taken as a reference.

The process of *imputation* consists of using a method to fill in missing values in the edited data. Items may be missing because they were not provided by respondents (*item non-response*), because the respondent gave only partial information (e.g. a range response to a question specified in terms of a continuous variable) or because the value provided was deemed to be incorrect during the data editing and set to missing. In wealth surveys, the share of missing values is sometimes relatively high: respondents might decline to answer on subjects that they find particularly sensitive, or they might not be able to value certain items. In general, the checking procedure should seek to address the process causing the problem rather than simply to find a way to rectify the results Thus, interviewers should help respondents by keeping them focused on the task, by reassuring them of their confidentiality, by convincing them to use records where possible, and by probing for partial information when nothing else is available.

It is usually argued that the data producers should provide imputations for missing values. Data producers often have information about survey observations relevant for imputations that is not included in the public version of the data. Because most statistical software performs only complete-case analysis, distributing data sets with a significant rate of item non-response imposes limits on users through the reduction in the number of observations available for analysis; cases with missing data may represent a systematic subset of the population, and omitting this group might introduce bias. Although some relatively sophisticated users may cope with estimation with missing data – by either using elaborate models or performing their own imputations – most users lack such skills. Even if users want to perform their own imputations, however, it would still be useful to have a standard set of imputations as a point of reference. Users who desire to do so may reset all imputed values to missing by using the information contained in the shadow variables.

Imputation of missing items also has costs, mostly in terms of the resources devoted to the development of the process.

In some countries, questionnaires with more than a few missing values on certain core variables are considered invalid and excluded from the data set. This approach is arguably the best strategy to maximise superficial data quality, since imputed values are not as informative as those collected directly from households. However, such a requirement may alter the incentive the interviewers or respondents face and lead to reporting unreliable values so as to avoid missing values. Moreover, the set of cases discarded may be systematically different from those retained in the data set, thus introducing the possibility of selectivity bias in the survey estimates. Discarding incomplete interviews is also very expensive, because more households have to be interviewed to attain the same final sample size.

While a number of technical options for imputation exist, all are based on the idea that there is enough commonality across households that relationships among cases without missing data on a given variable provide information about the likely value of a variable that is missing in another record. One of the simplest types of imputation model, *hot-deck imputation*, substitutes missing values in a case with values taken from a *donor household* with similar characteristics. A variety of regression-based approaches also exist, where missing values are substituted with fitted values, either based on an econometric model or on replication of the global correlation structure of observed values on missing ones. Imputed values should incorporate a stochastic element to preserve the variance of the variable in the observed population. In *single stochastic imputation*, a random term drawn from an appropriate distribution is added to a single hotdeck or regression estimate. In *multiple stochastic imputation*, the same process is repeated independently a given number of times. Multiple imputations allow a formal expression of the uncertainty of survey estimates; such uncertainty exists regardless of the method of imputation, but with simpler approaches estimation of the uncertainty becomes difficult or impossible.

Generally, even after checking, cleaning, editing and imputation, survey data sets are still not fit for immediate use. Weighting to account for selection and non-response is usually essential. The final sample on its own is not representative unless all units included have the same cumulative probability of selection and observation; this is unlikely to be the case in practice. For example, if the original sample oversampled areas near the sea, and blond people of all ages live disproportionately by the sea and older people have more time to participate in surveys, then the unweighted survey estimates would tend to be overly influenced by the condition of blond older people living by the sea. The sample can be realigned to the population through appropriate *estimation weights*, scale factors that indicate how many households are represented by each of the records included in the data set. Estimation weights result from a three-stage process. *First*, *design weights* are computed, corresponding to the inverse of each household's selection probability under the sampling design. *Second*, these weights are adjusted for unit non-response, based on models that predict participation probabilities of various socio-demographic groups (e.g. weights are inflated for respondents in low-participation groups, and deflated for those in high-participation ones). *Third*, estimation weights are derived by *calibrating* the sample to align with key dimensions of the population, such as the age distribution. In some surveys, the second and third steps are not distinguished.[4] Weights are usually made available to users in one or two forms: *not-normalised weights* that reflect the number of households represented by each case and thus sum to the total number of

households in the population, or *normalised weights* that reflect the fraction of households in the population represented by each case and that sum to one.

While weights are generally necessary in order to obtain unbiased estimates, they are sample-dependent and therefore stochastic; in most cases, they add to the variability of the final results. Techniques exist to limit this effect, such as the trimming of extreme weights. In the case of wealth surveys, the most difficult step in the construction of analysis weights usually lies in non-response adjustments. Given the substantial rate of non-response for wealthy households, the few who do participate risk being assigned very high weights, greatly inflating the variance of many key wealth-related measures. This may also lead to an under-representation of variance within the upper tail of the distribution, because so few cases would represent a very heterogeneous segment of households. Although this problem may manifest itself through weighting, it is more properly thought of as a problem of the inadequate representation of wealthy households. If only a few extremely wealthy households participate, it may be best to assign then a weight of one (reflecting their near uniqueness in the population), or to treat them as sufficiently different types of cases to be excluded from the final data set.

6.3.15. *Estimates and variability*

Estimates can be subdivided into positional and distributional. The former (e.g. simple or conditional means and medians) give an idea of the order of magnitude of a variable; the latter (e.g. simple or conditional standard deviations) give an idea of how heterogeneous the population is with respect to that variable. Most data producers concentrate on describing the distribution of wealth according to socio-demographic characteristics. Examples include: average/median wealth by household characteristics (e.g. income quintile, employment status, household size, education level of the highest earner), and inequality indices, such as the Gini coefficient, the 90/10 percentile ratio, etc. Whenever sample size allows, *domain-specific* estimates can be produced describing a particular sub-group: the debt-to-income ratio for the self-employed, the average interest rate on outstanding mortgages in a certain region, or the share of individuals over 65 owning life insurance. Published estimates generally include standard descriptive statistics and a few detailed indicators based on user needs. In light of the distributional/response behaviour issues described above, surveys may not be the best instrument for the estimation of total wealth.

Data creators and first-round analysts often publish a variety of descriptive statistics as a preliminary summary. Generally, such statistics include extensive tabulations of simple or conditional percentages (e.g. percentage ownership of at least one savings account, percentage ratio of debt payments relative to income, etc.), various simple or conditional indicators of size (e.g. mean or median savings account balances, quintile values of income, etc.) and a variety of other summary measures (e.g. Gini coefficients, standard deviations, etc.). The summary publication should include enough information about the survey methodology and context that technically knowledgeable readers can evaluate the plausibility of using the data for further research. Subsequent analysis may also employ more complex statistical models, such as regressions, Probit models, factor analysis, etc.

Wherever feasible, published estimates should include at least some indication of the degree of uncertainty about the results, typically in the form of a confidence interval or a standard error. The measure reported should be as comprehensive as possible, given the available data; most often, this will mean publication of a measure of sampling error. When data are multiply imputed, it should also be possible to report the combination of sampling

and imputation errors. There is usually no basis for estimating the magnitude of other sources of non-sampling error. Users should also be able to estimate measures of error from the data available to them. In many countries, however, complete design information cannot be released by the data creators owing to restrictions imposed by privacy laws; for example, for a geographically based design, the indications on the location of the sampled unit that would be necessary for calculating a simple design-based estimate of sampling error often cannot be made available. Where it is not possible to release the necessary design information for a simple design-based estimate of sampling error, the distributed data set should include *replicate weights* that can be used to simulate that estimate. Typically, replicate weights consist of weights computed using the same methodology as the main analysis weight, but computed over each element of a set of structured random subsamples of the final set of observations. The replicate weights are used to estimate a distribution of a given survey estimate over the sample replicates as a proxy for the distribution of sampling error.

6.3.16. *Preserving confidentiality*

Preservation of the confidentiality of respondents' data is an ethical and often a legal necessity. In the initial design of a survey, thought should be given to the question of what information can be made available for use at the end of the survey process – among the data creators and among researchers or policy makers beyond that group. This step is sometimes helpful in crafting survey measures that are useful but less sensitive than the most obvious ways of proceeding; it may also be helpful in establishing credibility with both respondents and interviewers.

Throughout the survey process, careful thought must be given to the control of not just information stored in computer systems but also paper documents, which are usually less easily controllable. Paper documents containing any type of identifying information should be kept to a minimum and procedures should be established for collecting and securing or destroying such documents. Electronic data held outside of the secure facilities of the data creator, particularly data held by interviewers, should be encrypted whenever feasible; such information should remain outside central control for the minimum time possible.

Standards for both what can be released from a survey and the types of users to whom the data may be released vary across countries. In general, it is not sufficient to remove obviously identifying information, such as names, pre-existing identification numbers, addresses, employer names, etc. It is also necessary to consider *indirect re-identification*, i.e. the possibility of using an attribute or collection of attributes to determine the identity of a given respondent with high probability. For example, unusual household structures taken together with detailed occupational data for household members might provide a basis for identifying the household. Increasingly, publicly available data about individuals makes it possible to identify people in unusual groups, thus raising the risk of re-identification. In general, it is not possible to remove all probability of re-identification without destroying the utility of the data for analytical purposes. However, it is possible to keep the probability sufficiently low. Data creators should consider all aspects of their data and consult the literature on the protection of confidentiality available at that time.

6.3.17. *Survey evaluation*

If a survey is a one-time project, data creators may limit the evaluation of the data to relatively simple summary measures, such as unit and item response rates, indications of

the comparability of survey estimates (aggregates, distributional estimates, change estimates) with estimates of the same quantities from other sources, etc. In some countries, there is an obligation to conduct a study to characterise potential non-response bias whenever the response rate to any official survey falls below a certain threshold; even where this is not a requirement, it is a good practice. For surveys that are expected to be repeated, evaluation should aim at identifying elements anywhere in the survey process that are capable of being improved. Data creators are encouraged to gather data and take notes over the course of the survey design and execution, which may help in the subsequent evaluation of the survey. Where possible, a debriefing of the full range of participants, other than respondents, should be conducted near the close of the survey work. Evaluation work should include research on basic aspects of survey measurement as well as comparisons of the outcomes of current and past practice.

6.4. Summary

This chapter has provided technical guidance on how to handle both general and specific challenges when designing and fielding surveys on household wealth. To that end, the chapter has taken potential data producers and users through the process of questionnaire design, sample selection, survey implementation and the production of estimates as well as the dissemination of the results.

The key highlights of this chapter can be summarised as follows:

● Sample surveys are a critical tool for the measurement of household wealth. Direct data collection at the micro level, especially in the context of multi-topic surveys also investigating income, employment and expenditure, allows users to take into account distributional facts and correlations between variables that are only latent in macro statistics. Appreciation of such surveys has grown along with an understanding of the potential importance of heterogeneity in explaining overall economic performance as well as in characterising responses to economic policies, (e.g. those directed at financial stability or tax design).

● Sample surveys focusing on household wealth confront data producers with certain challenges that are common across most types of surveys: samples have to be designed so as to maximise accuracy for a given cost, overcoming difficulties created by incomplete frames or lack of auxiliary information; complex computerised interviewing interfaces have to be implemented in order to ensure high data quality; field staff have to be selected and trained appropriately; whenever the survey is not compulsory, respondent co-operation must be encouraged, and the effects of non-response must be examined and appropriate adjustments made; the results have to be disseminated in a timely and clear fashion, offering users non-technical guidance on the variability of estimates. Some of the most serious problems occur during the interaction of the respondent with the questionnaire and, in the case of interviewer-mediated surveys, with the interviewer.

The distribution of wealth is more skewed than the distribution of other economic variables, and *ceteris paribus* both the unusually poor and the unusually wealthy are less likely to participate in sample surveys compared to the rest of the population. These facts combined suggest that in order to obtain accurate information on the overall distribution of household wealth, care should be taken to convince relatively poor people of their importance in a wealth survey, and special efforts should be focused on obtaining responses from the wealthy. Both efforts may have implications for sampling design and the allocation

of human resources; surveys where distributional estimates of wealth are not a priority may need to devote less attention to such issues, but the data creators should document clearly for subsequent users an appropriate level of caution for other uses of the wealth data.

For many people, wealth-related topics are more complicated than information collected in most other household surveys. Careful thought, research and evaluation needs to be devoted to the structuring and wording of a questionnaire, avoiding technical language as far as possible, and enabling respondents to report what they know, even if the information is only partial.

Experience suggests that wealth and income are among the most sensitive topics addressed in household surveys. Respondents will often need assurance that that their information will remain confidential. To this end, data creators must develop credible systems and procedures to maintain the confidentiality of respondents' information and develop effective means of informing respondents about the steps taken.

Notes

1. But even in those places it is necessary to turn to surveys to obtain important information not held for administrative purposes.

2. For example, information from a wealth registry or information on wealth or capital-income tax might be possible.

3. Training protocols should include the following: remarks from the data sponsor, if possible; explanations to interviewers of the purpose and the contents of the surveys; class-work on issues related to gaining co-operation, including role-playing exercises; class work on the contents and technical aspects of executing an interview; small workshops where interviews are simulated and interviewers are able to receive feedback from fieldwork managers and from each other; an exam covering procedural questions (e.g. "Which illustrative materials should I give to the household upon first face-to-face contact?"; "How do I answer doubts about the usefulness of the survey?"), technical questions (e.g. "What is the difference between a stock mutual fund and individual stock shares?" or "What is the evaluation rule for the main residence and how do I help households give the correct answer?"), and practical exercises (simulated full interview or sections).

4. Calibration can be a computationally intensive iterative process that requires abundant auxiliary information; it should be used carefully, limiting the number of calibration variables to the minimum necessary for the purposes of the survey.
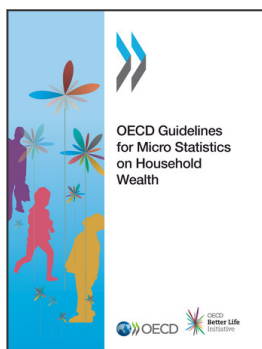
From:

OECD Guidelines for Micro Statistics on Household Wealth