Chapter 9

Quality assurance for household wealth statistics

This chapter focuses on the quality assurance of statistics on household wealth to ensure their fitness for purpose. It provides guidelines on best practice methods of assessing quality. These guidelines complement those on measurement, analysis and dissemination in previous chapters. They are intended for use by both data producers and data users.

The importance of quality assurance frameworks as a tool for defining data quality and assessing the quality of a set of statistics is discussed. The roles of existing frameworks developed by international bodies – including the UNECE, Eurostat and the OECD – are noted and the key features of these frameworks are described. Their relevance to micro statistics on household wealth is considered. The recommendations that follow are intended to cover both micro (individuals or families) and macro data (e.g. National Accounts). Although the quality of aggregate statistics is largely determined by the quality of the components, which are usually constructed using micro data, the estimation procedures error evaluation procedures may differ between the two categories and thus may require different considerations.

9.1. Quality assurance frameworks

The concept of "quality" has many different meanings, depending on the context in which it is defined and in its intrinsic subjective nature. The International Organisation for Standardisation defines quality as the "degree to which a set of inherent characteristics fulfils requirements" (ISO 9100). In producing wealth data there are often multiple requirements, and accommodating them may require a trade-off in terms of the quality of information to support individual objectives.

Using OECD definitions, the quality of statistics can be defined by reference to the following seven criteria: *institutional environment*, *relevance*, *accuracy*, *comparability, coherence, timeliness* and *accessibility*. Drawing on existing quality frameworks, this section briefly describes the different dimensions of data quality that should be considered in assessing the quality of household wealth data.

Each dimension is illustrated by quality issues that often arise in existing household-level wealth data. These include: variations in the understanding or attention of survey respondents; the accuracy of asset valuations (e.g. market value may not be approximated, or the price volatility of some assets may lead to inconsistent valuation); the impact of sampling errors (e.g. distortions due to outliers, or bias in mean/median measures and in the degree to which the wealth distribution is skewed); gaps or other deficiencies in coverage (e.g. differences from the standard coverage for certain types of households, such as those containing immigrants or those regarded as institutional households); non-response and under-reporting of data; the extent of comparability with macro and other micro sources; and the availability of confidentialised unit record data for use by the general public.

9.1.1. Institutional environment

The first dimension of quality is the institutional environment. This dimension refers to the institutional and organisational factors that may affect the image of the data producer. The institutional environment can be evaluated by considering different attributes:

- Impartiality and objectivity are related with the data production and dissemination using standardised statistical procedures in such a way that these practices are objectively transparent.

- Independence refers to the extent to which the institution producing the statistics is independent from political pressure and other regulatory or administrative bodies, as well as from private sector operators and potential conflicts of interest. The mandate for data collection is defined as the extent of the legal act by which administrative organisations, firms and households may be compelled to provide data to the institution.
- The adequacy of resources is the extent to which the resources available to the agency are sufficient to meet its needs for the data production.
- Finally, the quality commitment is the extent to which processes, staff and facilities are in place for reaching the target quality levels.

Caution may be required when using wealth data from private operators who are not compelled by law to collect such data, as they may have an interest in providing a biased representation of reality; for instance, the evaluation of property prices by real estate agents may suffer from a conflict of interest.

9.1.2. Relevance

The *relevance* of data is the degree to which statistics meet the needs of actual and potential users. Relevance thus depends upon both the coverage of the required topics and the use of appropriate definitions or concepts.

As noted in Chapter 3, there are broader and narrower concepts of household wealth. In general, the definition of household wealth refers to the sum of real assets and financial assets less financial liabilities. However, more comprehensive concepts of wealth may include pension entitlements, various state-contingent assets, human capital and public resources. For some relatively extended wealth definitions, quality may suffer, particularly when there is a need to incorporate estimates of future situations in a present value, as is generally the case in estimating the present wealth value of future pension rights.

9.1.3. Accuracy

The concept of *accuracy* is related to the degree to which the data allow estimation of the population characteristics they are designed to describe. Accuracy has many attributes, and in practical terms there is no single aggregate measure to summarise it. Typically, this characteristic is more easily described in terms of sources of errors. In a survey, errors cause survey responses or distributions of survey responses to deviate from their true values.

The total survey error (TSE) refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. Data producers should optimally allocate the available resources to minimise TSE for investigating a limited number of relevant population characteristics. Insofar as possible, major sources of error should be analysed as part of the initial development of a survey, so that resources can be assigned efficiently to reduce errors to the extent possible, while still satisfying specified costs and timeliness objectives.

The sources of error in sample surveys can be divided broadly into two categories: sampling and non-sampling error. The former includes errors in estimating the interested population parameters that derive solely from the sampling or estimation process. Non-sampling errors mainly relate to measurement, data collection and processing; this class comprises quite diverse specific types of error that are usually harder to control than sampling ones. Administrative data collected for non-statistical purposes usually cover the

whole reference population; thus they are generally affected by non-sampling errors only. Similar considerations apply for census data.

For sample survey estimates, an evaluation of sampling errors can be carried out with the computation of classical standard errors of estimators. In simple cases, these can be obtained by means of algebraic formulas; more frequently, the adoption of complex sample designs suggests approximate solutions (e.g. the Jackknife Repeated Replication method or other replication methods).

Non-sampling errors can be classified in specification error, coverage error, non-response error, measurement error and processing error.

- Specification error occurs when the collected data do not include relevant economic variables for the objectives of the survey, where the relevant variables can only be approximated, or where the elaboration of questions and instructions are ambiguous.
- Coverage error exists when some statistical units belonging to the reference population are not included in the sampling frame or when the density of the sampled population differs in some other way from the reference population.
- Non-response error occurs because some households do not participate in the survey at all or they decline to answer or cannot answer individual questions in the survey.
- Measurement error arises during the data collection process and includes errors made by the interviewer or by the respondent, and errors in the survey instrument or other measurement protocols.
- Processing error includes errors emerging from data entry, data editing, or other computer programs or processes that affect the data after they are collected.

In budgeting a survey, there is a clear trade-off between sampling and non-sampling errors. Resources can be devoted to procuring a large sample and thus minimising random sampling error, or else concentrated on a smaller sample but with better interviewer controls, a higher response rate, more accurate data collection procedures, and other measures intended to improve the survey process.

Most often, in household sample surveys not all the units selected for the survey will actually be interviewed. The difference between the target and the actual sample reflects unwillingness to participate or other factors, with the most common one being difficulty in contacting the selected household. When non-response occurs, the estimators of the population parameters will generally be biased unless the pattern of non-response is completely random. In the absence of information to the contrary, it seems prudent to assume that there are reasons that some households are more or less likely to participate in a survey and that those reasons might be confounded with the variables of interest in the survey. In some surveys, substitutes for individual non-respondent cases will have been introduced into the sample. While use of replacements can allow a given realised sample size to be maintained, it does not address the possibility of bias. Where it is possible to closely match non-respondents and substitutes along key dimensions of a survey, there may be some ground for treating estimates using the sample with substitutes as approximately unbiased. Nonetheless, every effort should be made *ex post* to evaluate the reliability of any system of substitution.

Response rates may vary over different groups. For example, surveys in many countries often show lower response rates in urban areas, particularly the largest cities. Because older people usually are more likely to be at home than others, they are also more likely to be

reached and persuaded to do an interview than other types of households. Similarly, relatively wealthy people may be difficult to contact and less willing to participate.

Where auxiliary data are available on respondents and non-respondents, it is recommended that the sample survey designers estimate response rates at the level of the available classification variables and investigate the implications for bias in survey estimates. It is also important to consider differential co-operation across interviewers. *Paradata* (process data generated in the execution of a survey, such as the time and date of all attempts to obtain an interview with a given respondent, together with the characteristics observed for all sample members) may also be useful in understanding patterns of non-response and their potential implications for bias. In fact, some characteristics of both respondents and non-respondents can be detected. In conducting personal interviews, for example, the characteristics of the neighbourhood and of the building are observable. Comparing respondents and non-respondents as regards these characteristics can help to understand the possible bias arising from the response process. Information on the characteristics of non-responding households can also be inferred by analysing the effort required to get an interview from responding households. A comparison of the households that were interviewed at first visit with those that agreed to be interviewed only after their first refusal provides information on non-response. When the non-response rate is high and the analyses show a possible presence of bias, one should also produce adjusted estimates by re-weighting the interviewed households by the inverse of the estimated propensity to participate, to the extent that this is possible.

Several statistical techniques, based on various assumptions, can be employed to address non-response issues. Knowledge of the distribution of some relevant characteristics for the entire population can be used to adjust the corresponding sample characteristics with the census or administrative compositions. Moreover, a significant deviation of the sample distribution from that of the population gives indirect information about random missingness in the response process. The sample composition can then be aligned with population distributions by means of post-stratification techniques. When auxiliary information is available in the form of knowledge of marginal distributions, the Iterative Proportional Fitting method can be employed. More generally, calibration techniques, based on a linear regression model, offer a wide variety of solutions to adjust the sample weights so as to reproduce external known information.

Longitudinal household surveys present other problems. A household may not be fixed over time, and only some part of the original household (perhaps living in a household with other people not in the original survey) might be available to be interviewed later. Even for a given household unit, non-responses may differ from one wave to the next, because non-response in later waves may be affected by the experience of earlier waves. As in the case of cross-sectional surveys, every effort should be made to understand the patterns of non-response and the implications for bias in key survey estimates.

In sample surveys, bias due to non-sampling errors may sometimes be reduced by adopting a few simple practices in the initial contact with the household. It is usually recommended that respondents be sent a letter explaining the purpose of the survey and encouraging participation. Additional material, such as a clearly designed booklet describing the main uses of the information and providing explanations and assurances of confidentiality, may also be helpful. In a wealth survey, it may be particularly important to offer the respondent a means of verifying the validity of the survey. The availability of a toll-

free telephone number for respondents to call to obtain more detailed information can be a valuable asset for participation. Symbolic incentives or gifts have also proved useful in encouraging participation. For in-person or telephone interviews, the role of the interviewer as the mediator of the survey and the representative of the sponsor is very important. Effective interviewers need to have both a high level of persuasive skill and sufficient logical skill to navigate a complex technical interview. The training, experience and compensation of interviewers are important factors in non-response. Interviewers should adapt their schedule to the respondents' availability. In this regard, it is also recommended that surveys develop and enforce a protocol to ensure that interviewers make a minimum number of attempts to obtain an interview and that those attempts take place at different times of the day and week.

A best practice for reducing non-response to questions about money values is to record range information. For each money amount for which the respondent cannot or will not provide an answer, an alternative answer consisting of a range containing the answer may be solicited in a variety of ways. There is a long history of surveys that use a "range card", i.e. a list of a sequence of ranges with a means of identifying each range without have to read the entire range. Evidence also exists that allowing respondents to offer their own ranges may provide a tighter range than alternative approaches. There is also experience of using a logical decision tree to specify a sequence of "unfolding brackets", using questions in the form, "Is it EUR 10 000 or more?" Some surveys have used a combination of all three approaches. Several studies have shown that relatively large proportions of respondents who initially refuse to answer or don't know the exact answer to an income question will provide range information. Although range information is only a partial answer to the intended question, it does allow for the possibility of more efficient estimates. Such information may also still help to reduce biases if respondents who provide a range of information are systematically different from other respondents.

The sampling frame is a list or a mechanism from which a sample is drawn. For most household sample surveys, the target population is the civilian non-institutionalised population. Sometimes the sampling frame is a list of target population members, such as a population census or fiscal registry. At other times, a method, such as area-probability sampling, is used to select an unbiased sample without the need to enumerate or know the entire population. In principle, the frame should allow a non-zero probability of being placed on the selection of every member of the target population, and no element should be duplicated or have the uncontrolled possibility of being selected in multiple ways under a given mechanism. Unfortunately, sampling frames sometimes fail to satisfy these requirements. The accuracy of data obtained from household surveys may depend to an important extent on the quality of the sampling frame from which the sample was selected. In sample surveys, the most common and critical frame omissions involve population "non-coverage" errors. A non-coverage error refers to the incompleteness of the sampling frame in assigning *ex ante* a positive probability to the selection of each unit of the target population. For instance, for a list-based sample, this problem may arise for particular subgroups of the population, such as illegal immigrants or households that have a higher geographical mobility, for which the lists rapidly become inaccurate. Whenever it is possible, a measure of sample under-coverage should be computed and the implications for the key survey estimates should be considered.

The specification error arises from the discrepancy between the concept implied by the survey question and the concept that should be measured in the survey. This error is often rooted in the planning stage of a survey, where the specification of the desired information is

inadequate and/or inconsistent. Specification error may arise from poorly worded questions, inadequate instructions, or confusing framing or sequencing of topics in the survey.

Processing errors comprise editing, data entry, coding, assignment of survey weight errors, and any other incorrect manipulation of the data before it reaches its final state. Such errors arise during the data collection and processing stages. In this class, one of the most critical errors is the miscalculation or misspecification of the survey weights; such errors may produce severe bias in the estimates. Data entry errors may be reduced through appropriate implementation of Computer-Assisted Personal Interviewing (CAPI) to administer the questionnaire. CAPI allows the automatic routing of questions, contingent on answers to other questions, and including a variety of consistency checks.

Measurement error may arise from the action or inaction of respondents or interviewers. Respondents may misunderstand questions, they may have an inadequate understanding of their own situation, they may be unwilling or unable to check records during the interview, or unable to answer a question at all, or they may deliberately provide incorrect information in response to questions. Interviewers may cause errors directly by failing to follow instructions or other survey protocols, by incorrectly entering information in the questionnaire, or by falsifying data. They may also cause errors indirectly if their way of speaking or acting influences respondents to provide incorrect information. If measurement error differs across groups in ways that cannot be controlled by in estimations, then some differences across groups seen in the data may be illusory.

The components of wealth are usually evaluated at a market price, i.e. the price at which a particular asset may be sold at a given time on the market. However, households may not know the precise market value of their assets. For example, this situation may occur for dwellings bought a long time before the interview or for highly volatile financial assets. The analysis of wealth values over time should take into account a certain weakness in the information provided.

Even involuntary errors in reporting values of some phenomena (e.g. the size of the respondent's dwelling), due to rounding or to lack of precise knowledge, may cause serious problems to estimators. In particular, the "classic" measurement errors (independent of the true latent value) inflate the standard errors of estimates.

The evaluation of measurement errors is useful both for producers, as it can give an insight into improving the questionnaire or collection procedures, and for users, who must be conscious of the limitations of the data they use. Often, measurement errors can be evaluated only indirectly, through examination of inconsistent or implausible values, or through comparison of survey responses at the household level or for groups of households to estimates obtained from other sources.

In household wealth surveys, the most critical type of measurement error is the under-reporting of wealth assets, which may arise from recall difficulties, or from a reticence to report what is perceived to be sensitive information. In particular, the propensity to report wealth may differ from country to country, depending on cultural norms and more practical issues related to tax evasion. This type of error can produce severe bias in estimates, and special techniques are required to overcome it. To evaluate the under-reporting problem, a useful approach is to compare estimates derived from different data sources (sample surveys, administrative registers, fiscal data and National Accounts). A typical example is the discrepancy between the number of dwellings declared by households in the sample survey and the number owned by households according to the census or the administrative register.

In the presence of under-reporting behaviour, estimates of real and financial wealth will be underestimated by comparison with the macro amounts. If the under-reporting is not uniform across the different wealth components, averages and other statistics will be biased. For example, if the under-reporting is higher for financial assets like equity and investment fund units, which are more frequently held by rich people, it is presumable that the concentration index will also be downward-biased.

Different approaches to measuring the under-reporting can be constructed using statistical matching procedures performed between answers of household surveys and data from other sources, such as the statistics held by commercial banks on their customers. In particular, micro data may allow measuring both the non-reporting and under-reporting of wealth assets for different groups of households.

It is recommended that, in the course of the interviews, interviewers provide additional information, e.g. comparing household's answers and the objective evidence they can see for themselves: the type of neighbourhood and type of dwelling, the standard of living implied by the quality of furnishings, and so on. In fact, the interviewers' opinions can be a good instrument to assess the credibility of the sample survey responses.

Compared to sample surveys, administrative data usually allow an analysis of specific geographical domains (e.g. house registers or fiscal data) and high-frequency statistics (e.g. stock market indices). The monetary costs specifically attributable to the production of these data and the statistical burden of respondents are usually limited. However, using administrative data requires a deep knowledge of the regulatory environment for which these data are collected. In particular, typical drawbacks of these data are lack of coverage of specific sub-populations (e.g. unlisted companies), incoherence between legal and statistical definitions (as in the case of official registers of the values of dwellings) and a lack of data freshness. Reporting errors may also occur in fiscal data.

The evaluation of the accuracy of aggregate statistics does not rely on the tools applied for micro data. In fact, these estimates are obtained using complex procedures in which measurement and processing errors are rarely monitored, preventing an analytical computation of the estimation error. An indirect quality indicator of such statistics can be obtained by analysing the number and size of past revisions.

9.1.4. Comparability

Comparability refers to the degree to which data can be compared over domains, across countries, and over time. Comparability aims to eliminate (or at least reduce to the maximum extent) the effects of differences in definitions and measurement procedures when statistics are compared. Therefore, consistent procedures, particularly ones based on the use of international agreed definitions and standards, are important. Known deviations from standards should be fully documented for data users.

The fluctuations in the market prices of certain assets may produce large differences over time, even in the absence of stock variations, in both the amount and the inequality of wealth. Typically, a rise in the stock market is associated with an increase in inequality, as the shares are held mostly by wealthier households. The contrary tends to happen when there is rise in the housing market.

In comparisons over time, any change in the survey, such as the mode of the interview, the interview questions or question ordering, the sampling frame, the strata definitions or the oversampling of a specific domain, may produce significant effects in the comparability

of estimates. The extent of these effects should be accounted for through specific tests. A typical example of a change in the survey is the effect of a reformulation of a question, which can be evaluated by randomising the old question on one half of the sample and the new formulation on the remaining half. Measures may also decay in terms of the inter-temporal comparability if the institutional structures dealt with by households change. For instance, the introduction of tax amnesties, which regularises undeclared or untaxed assets, may produce an increase in the reporting of wealth.

Comparisons of wealth across countries may be strongly affected by institutional differences in entitlements. Pension systems in particular often have distinctive features at the national level, which may have important effects on the accumulation of other forms of wealth that are more easily measured. Such institutional features may also change over time in ways that are both hard to predict and induce further changes in other types of wealth. Even in cases where the relevant institutional features are reasonably fixed, individuals may be unable to report the details needed to estimate the present value of a given type of entitlement. Because some entitlements are contingent in nature, it may be difficult or impossible to estimate the relevant probabilities at the level of individuals, even when the overall distribution of probabilities needed to understand aggregate outcomes is known. Great care should be taken in international comparisons involving countries with different pension systems. Other institutional systems, such as tax-deferred savings accounts or real estate financing arrangements, may also differ substantially across countries, and the implications of such factors for comparisons of wealth data should always be considered carefully.

There should be similar concern about comparisons over time or between countries that have very different ratios of private to total wealth. The amount of household wealth may be influenced by that of public wealth (or debt). At the same time, the imputation of public assets and liabilities to single households is, at best, a complex operation.

9.1.5. *Coherence*

The *coherence* of survey data concerns their adequacy to be reliably combined in different ways and for various uses. Coherence may be divided into internal and external coherence. The former refers to the coherence between different economic variables collected in the same cross-section or inferable from the longitudinal component of the survey. The latter is related to the coherence with external sources of information, such as the national accounts or population census.

The comparison of information on income, wealth and expenditure offers a first and valuable possibility for checking the internal coherence of the collected micro data. Anomalous relationships between consumption, income and wealth can in fact immediately reveal data problems. Moreover, the information collected over time on the same units allows constructing a household balance sheet verifying the accounting identities between these economic variables. Panel data also allow measuring the time consistency of the time-invariant variables. In case of time-varying variables, the evaluation of the data reliability requires the adoption of models for disentangling the true dynamics from the measurement error.

The editing and imputation procedures are standard practices to check and restore the internal consistency of the collected or produced data. Even if there is a common agreement that invalid or self-contradictory entries should be automatically removed, the

excessive use of data processing can itself affect the quality. In fact, researchers may edit data more than necessary, because of a low capability of identifying "true" errors. High percentages of acceptable data erroneously classified as unacceptable affects the effectiveness of the editing process, by introducing a slippery non-sampling error. Therefore, it is recommended that edit and imputation procedures lead to the amount of data processing strictly necessary.

In some cases, useful information on the internal coherence can be obtained by the comparison of estimates of the same phenomenon constructed in two different ways. For example, in a wealth household survey the estimate of the total number of houses owned by households and rented to others can be compared with the corresponding estimate drawn from the number of households living in dwellings rented from other households.

As noted in Chapter 2, detailed comparison of macro and micro statistics of household wealth can improve the understanding of the quality of both data sets, including their strengths and weaknesses. For example, it can help to identify items that are under-reported in the micro statistics as well as items that are under-estimated by the sources and methods used in the macro statistics. This can lead, in turn, to improvements in the accuracy and coherence of both sets of statistics.

In many countries, common sources for comparing sample estimates on the number of dwellings are the census and administrative registers. Financial Accounts, constructed following international standards, provide more general and harmonised sources for the micro-macro comparisons of financial assets and liabilities. However, some definitions or conventions used in the Financial Accounts do not always favour a precise comparison. For example, non-profit institutions serving households (NPISH), such as charities and trade unions, are often grouped together with households; their economic weight is limited but not negligible. The regular confrontation between micro and macro statistics can be helpful in explaining the differences between them to users, thereby improving interpretability.

The classification of the assets held by households may sometimes be ambiguous. For instance, if a household owns a company that in turn holds a dwelling, it is not obvious *a priori* whether the asset should be classified as a real estate holding, a personal business or a financial asset. Although the SNA conventions clearly define the accounting rules for such situations, differences in legal and/or accounting conventions across countries may lead to different answers. Careful consideration of such differences is particularly important for comparisons across countries. Moreover in principle and in practice, constraints on data collection may lead to further qualifications. In sample surveys the availability of additional information can be used to give a statistical representation that better fits the actual situation.

There are some asset categories that include varying types of assets, whose treatment may have substantial effects on the external coherence of survey data with other measurement frameworks. For example, in some countries managed accounts are not uncommon, and such accounts may be invested in a variety of more specific asset types. If a survey respondent is not heavily engaged in monitoring such an account, that person may not know the more detailed portfolio composition, or have only a general idea about it. Some types of trust accounts, annuities or insurance contracts, and tax-preferred accounts may have similar characteristics. To the extent that it is feasible to learn more from respondents about portfolio composition, it may be possible to increase the measured coherence of survey and external data.

9.1.6. Timeliness

Timeliness refers to the interval of time between publication and the period to which the data refer. It is important that data and the corresponding estimates be made available as soon as possible, so that policy decisions can be made on reasonably up-to-date data.[1]

Aggregate statistics can usually be provided with higher (quarterly or yearly) frequency and can be used for business outlook analyses. In contrast, more time is usually required to collect and process micro statistics, which are therefore employed to study structural economic changes. For instance, the Household Finance and Consumption Network (ECB) recommends three years as the minimum frequency for gathering household financial budgets using sample surveys. A lower frequency would save costs but at the expense of significantly diminishing the utility of the survey data for policy purposes. However, in some circumstances, the lack of frequent data can be overcome by combining this data with more updated external sources of information (population distributions, aggregate variables, etc.), for instance through the use of microsimulation.

There is a clear trade-off between accuracy, cost and timeliness. A larger sample can reduce the random sampling error at the cost of increasing the time required to collect, clean and edit the data for analysis. In the same way, using a smaller and better-trained group of interviewers or increasing the intensity of the field activities in order to increase the response rate may have a direct effect on the length of the field period.

9.1.7. Accessibility

Accessibility refers to the degree to which users are able to use the data. The concept of accessibility spans the physical requirements for access, the structure of data files, the tools available for access, the restrictions placed on accessing the data, and the adequacy of supporting documentation.

Survey data on wealth and other such information collected for scientific purposes should be made available as broadly as possible for the intended purposes, to the extent that the confidentiality of the respondents can be protected adequately. There are trade-offs between the breadth of the data released and the ease of access, and between the disclosure risks and public benefits of the research. Although secure central data repositories can be used to allow access to sensitive information that cannot be released more generally without risking identification of the respondents, the requirement to be physically present in such facilities inevitably limits data access and the range of analysis. Data enclaves, in which users are given access to sensitive data securely via the Internet, can expand the range of uses of data that can be shared in this way, but the most sensitive information generally should not be included. The broadest audience is reached with data sets that are made more freely available to researchers, but such data sets must be scrupulously anonymised to fulfil the ethical and, generally, legal necessity to protecting respondent confidentiality. The anonymisation process involves editing the content of records to eliminate information that can be used to identify the respondents directly or indirectly. Direct keys to households' identity (e.g. name, codes) must be suppressed; in addition, actions must be taken to reduce the likelihood that other variables or combinations of variables might identify respondents. Suppression of variables such as geographic information is very common, as are collapsing of categorical values into broader

1. Another quality dimension is *punctuality*. This attribute refers to the time lag between the actual and the planned dates of publication.

categories, truncating some values at an upper and/or lower value, rounding off monetary amounts, and the constrained simulation of some values. In some extreme instances, certain household records may be eliminated from the public data set altogether, if the necessary amount of data suppression to allow release would render the survey observation useless for analytical purposes.

Accessibility is increased by offering data in a variety of formats (e.g. SAS, Stata, R, etc.). Because existing data formats may not persist forever, however, consideration should also be given to archival management of the data and supporting material. The construction of ASCII files, which embed a minimal amount of structure and are generically readable, is one common solution. Restricted versions of the data and supporting information should also be archived with care. In some cases, privacy constraints may be a function of the age of the data, and it is important to preserve access for future researchers. Consideration should also be given to providing users with software tools to address calculations that are specific to the data set.

For researchers to understand the data more fully, they should have as much access as feasible to the tools and other structures that led to the creation of the data they observe. At a minimum, such information should include a representation of the survey questionnaire; a set of auxiliary tools, such as illustration cards and other accompanying devices; and a full listing of codes used, including those entered directly during the interview and those coded subsequently. Ideally, users should also be provided with flag variables describing the status of each variable (e.g. originally missing and imputed in the final data) and methodological research into the reliability of the data, particularly when response rates are low.

9.2. Summary

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

- The quality of statistics can be considered in terms of seven criteria: institutional environment, relevance, accuracy, comparability, coherence, timeliness and accessibility. There are inevitably trade-offs between costs and quality, and between various aspects of quality.
- In a high-quality institutional environment, the statistical agency producing the data is impartial, objective, independent from political and other institutional pressures, and free of potential conflicts of interest. It is adequately resourced to produce the statistics of interest, and has a mandate to collect the relevant data.
- The relevance of data is the degree to which statistics meet actual and potential users' needs. Thus, relevance depends upon both the coverage of the required topics and the use of appropriate definitions or concepts. There may be a trade-off between relevance and other aspects of quality; for example, more comprehensive and more relevant data items may be less accurately measured than more narrowly defined data items that are easier to collect.
- Accuracy is related to the degree to which the data correctly allow estimation of the population characteristics they are designed to describe. Sampling error refers to an inaccuracy that arises because data is collected only from a sample that may not be fully representative of the total population of interest. There are several distinct categories of non-sampling error.
  - Specification error occurs when the collected data do not include relevant economic variables for the objectives of the survey, where the relevant variables can only be approximated, or where the elaboration of questions and instructions is ambiguous.

- ❖ Frame error, or coverage error, exists when some statistical units belonging to the reference population are not included in the sampling frame or where the density of the sampled population differs in some other way from the reference population.
- ❖ Non-response error occurs because some households do not participate in the survey at all or they decline to answer or cannot answer individual questions in the survey.
- ❖ Measurement error arises during the data collection process and includes errors made by the interviewer or by the respondent, and errors in the survey instrument or other measurement protocols.
- ❖ Processing error includes errors emerging from data entry, data editing, or other computer programs or processes that affect the data after they are collected.

- Comparability refers to the degree to which data can be compared over domains, across countries, and over time. Comparability aims to eliminate the effects on statistical comparisons flowing from differences in definitions, survey instruments and measurement procedures, sampling frames, and institutional structures. Where these differences cannot be avoided, attempts should be made to measure the impact.
- The coherence of data refers to their adequacy to be reliably combined in different ways and for various uses. Internal coherence refers to the coherence between different economic variables collected in the same cross-section or inferable from the longitudinal component of the survey. External coherence is related to the coherence with external sources of information, such as the national accounts or population census.
- Timeliness refers to the interval of time between publication and the period to which the data refer. It is important that data and the corresponding estimates be made available as soon as possible, so that policy decisions can be based on reasonably up-to-date data.
- Accessibility refers to the degree to which users are able to use the data. The concept of accessibility spans the physical requirements for access, the structure of data files, the tools available for access, any restrictions placed on accessing the data, and the adequacy of supporting documentation.