

Copyright © 2015 Banco Interamericano de Desarrollo. Esta obra está bajo una licencia Creative Commons IGO 3.0 Reconocimiento-No Comercial-Sin Obra Derivada (CC-IGO BY-NC-ND 3.0 IGO) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando crédito al BID.  No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI. El uso del nombre del BID para cualquier fin que no sea para la atribución y el uso del logotipo del BID, estará sujeta a un acuerdo de licencia por separado y no está autorizado como parte de esta licencia CC-IGO.

Notar que el enlace URL incluye términos y condicionales adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.

**TIP**

Documenting the choice of a survey method.

All information related to the survey may be useful and should be archived (even if not all will be disseminated to the public). This includes not only technical documents such as the questionnaires or list of codes (obviously needed by data users), but also administrative reports (potentially useful for implementation of future surveys), and other documents such as a compilation of the comments provided by stakeholders at the time the questionnaire was designed, etc.

**WARNING**

Comparability does not mean repeating the same errors over and over again.

**WARNING**

It is a mistake to think that users enter the process only at the analysis stage.

**WARNING**

Increasing the sample size does not only cause the budget to increase. There are concomitant difficulties of scale – for example, the core team might no longer be big enough to be able to monitor the activities effectively – and other sampling issues (see Chapter 5, “Sampling Error and Sample Size”) which might mean that the increase creates more problems than it was meant to solve.

**WARNING**

 It seems very obvious, but it is sometimes difficult to achieve the goal of a full-time project manager, where the statistical agency is highly structured into departments, divisions and so on.

**WARNING**

 The corollary of granting a high degree of financial autonomy is that rigorous auditing must be conducted after money has been disbursed.

**WARNING**

 The contingency line is not meant to cover up for incompetent drafting of the budget, and it should not be expected to fill gaps in it.

**WARNING**

 Underestimating the value of some items, or forgetting that the price of some materials might change between drafting the budget and purchasing them, can seriously endanger the success of the project

**WARNING**

Non-random sampling methods (such as purposive sampling, convenience sampling or quota sampling) cannot provide valid conclusions about the whole population – their results are only valid about the households in the sample.

**TIP**

The inverse square root formula means that a law of diminishing returns underlies the relationship between standard errors and sample size. Additional increments in sample size bring about progressively more disappointing reductions of sampling error

**WARNING**

 Since there are no exact formulas for non-sampling errors, their incidence tends to be overlooked at the sampling design stage, but they are a very real risk. There comes a point at which efforts to improve the accuracy of a survey by enlarging its sample size become counterproductive

**WARNING**

 Beware of people who report the size of a sample as a percentage of the size of the population – this usually reveals ignorance. What matters is not the sampling fraction but the sample size.

**WARNING**

 Most household surveys need to be analyzed with weights because their samples are disproportionally stratified

**TIP**

 The sample size is the most important single indicator of overall survey precision. The second most important is the number of clusters.

**WARNING**

 Some researchers report the square root of the design effect (called the design factor, deft) instead of the design effect (deff). When interpreting a reported design effect, always check if it is a deff or a deft.

**TIP**

Census data managers should not see the production of the summarized list of EAs as a by-product of the census data processing activities, but rather as a precondition of efficient census data processing.

**WARNING**

Avoid segmentation if you can, but if you cannot, remember that segmentation is a de facto additional sampling stage. All the details of the process (sketches, quick counts of dwellings, etc.) will be needed later to compute weights.

**TIP**

 In some countries it is possible that the local authorities have records of construction permits which can be used to update the frame or define alternative sampling strategies.

**TIP**

If you want to allocate a sample proportionally, you don’t need to define explicit strata. Implicit stratification is simpler and costless. It even makes estimation easier!

**WARNING**

Never work through the country one region at a time. This “sweeping” technique will impede the accurate measurement of seasonality.

**WARNING**

Household listing is a field operation that requires training and supervision. It may account for between 10 to 20 percent of the fieldwork costs.

**WARNING**

Statistical software will compute the standard errors of a complex sample only if instructed to do so. Otherwise it will simply assume that the sample is a simple random sample and grossly underestimate the standard errors.

**TIP**

Replacing non-respondents is not a solution for non-response bias, but it can be a preventive measure, because it eliminates a perverse incentive. Interviewers should not see non-response as way of making their work easier.

**WARNING**

Don’t assume that the staff of the national statistical agency will be good interviewers. Their very experience means that the reverse may well be the case!

**WARNING**

It is sometimes suggested to deal with non-response by replacing households; increasing the size of the sample to compensate for non-respondents; using formulae or correction models; imputing missing values with methods such as cold-deck, hot-deck, etc. Neither of these actions is a solution to the problem. The best way to deal with non-response is to prevent it.

**WARNING**

Integrating data entry into fieldwork implies that all members of the field team fully understand the program messages. The data-entry program can become the ideal tool for supervising the quality of fieldwork, but only if it has been written from the perspective of genuine integration into that fieldwork.

**WARNING**

Good data entry workshops will standardize the editing task as much as possible. However, most of the corrections which are applied are based on informed guesswork which cannot be checked with the source of the information. Only in exceptional cases would there be a return to the field.

**TIP**

Don’t underestimate the time the field team needs to spend in each sample point. Measuring the yield of the field teams in “households per week” rather than in “hours per household” will help you avoid this mistake.

**TIP**

A CAPI application should include supervision techniques which can be used in addition to direct observation of interviews. For example, interviews can be digitally recorded; data can be time-stamped at the time of entry to ensure the integrity of the interview process; and information from multiple sources (reference tables, other respondents, other questionnaires) can be cross-checked automatically against the information as it is keyed in during the interview.

**WARNING**

Programmers naturally look at the questionnaire as if it were a computer program, and love to see it skip automatically, but this is not a good idea.

**TIP**

To implement the checks on unit prices and energy consumption, the program will need to consult a table of price ranges and calories per gram.

**WARNING**

Beware of blindly copying questionnaire forms from other surveys, or from sources such as Volume III of Grosh & Glewwe (2000). These are precious resources which can and should be consulted, but questionnaires must always be fully adapted to the country in which the survey will take place.

**TIP**

Make sure that the different versions of the questionnaire can be clearly distinguished from each other while they are being circulated for comment.

**WARNING**

If integrated data-entry is not being used, the centralized data-entry support must still be included in the field test

**WARNING**

Avoid including unnecessary instructions in the manual for actions which should be perfectly clear from the questionnaire, such as: “For sex, write 1 if male and 2 if female”

**WARNING**

The use of a CAPI application eliminates the need for a paper questionnaire, but this doesn’t make the questionnaire development process any easier.

**TIP**

Record multiple answers only if multiplicity is an essential analytical need. Otherwise, you can always avoid multiplicity by adequately wording the question. For instance, rather than asking “What are your favorite sports?” ask “Which sport do you like more?” or “Which sport did you last watched?”

**WARNING**

Where differences are found between the original version and the back-translation, don’t assume that it is the translated version which is wrong! Check the questionnaire in the main language too, in case there is an ambiguity there which needs to be corrected.

**WARNING**

Some people say that the diary method cannot be used in illiterate countries, but this is not true. It’s a little harder but it can be done

**TIP**

The interviewer should go through the whole list first, asking the respondent if each item was consumed or not. Only when all the items have been mentioned should s/he ask for the quantities consumed. Otherwise, respondents realize that if they say they have consumed an item, they will then be asked tiresome questions about it. As the interviewer proceeds down the list, the respondent often starts to deny that the household has consumed subsequent items, in the hope that this will shorten the interview.

**WARNING**

The deliberate recording of certain expenses in more than one section must be dealt with carefully by analysts, to avoid the risk of potential double-counting.

**WARNING**

Be careful with housing expenditures. Housing expenditures are second only to food in a household’s budget.

**WARNING**

Blank and zeroes are very different things.

**WARNING**

If control totals are not removed from the datasets, inexperienced analysts may end up counting things twice; for instance adding up the total expenditure on fruits to the detailed expenditures on apples, oranges, etc.

**WARNING**

Decentralized data-entry workshops are almost always the worst possible option. It is very difficult to ensure uniform management and training if data-entry operators are based in several different locations. Decentralization tends to amplify and multiply the defects and weaknesses of centralized data entry.

**WARNING**

As far as possible, avoid extrapolating totals from your estimates. If possible, limit yourself to percentages, averages, per capitas, and so on. There is sure to be at least one other survey which was previously done in that country with different totals, and you will have to explain over and over again why yours are not the same.

**TIP**

When organizing data records for dissemination, consolidate variables which refer to the same statistical object into a distinct - for example, individual data about health or education

**WARNING**

On no account should the data be consolidated into a single record at household level!

**WARNING**

As far as possible, avoid extrapolating totals from your estimates. If possible, limit yourself to percentages, averages, per capitas, and so on. There is sure to be at least one other survey which was previously done in that country with different totals, and you will have to explain over and over again why yours are not the same.

**TIP**

When organizing data records for dissemination, consolidate variables which refer to the same statistical object into a distinct - for example, individual data about health or education