

# International Household Survey Network Survey Quality Assessment Framework SQAF

---

DRAFT

Prepared by the Statistical Services Centre of the University of Reading

May 2009

# Table of contents

<b>Chapter 1: Introduction to the SQAF Checklist</b> .....	<b>1</b>
What is the SQAF checklist? .....	1
Completing the SQAF checklist? .....	1
Who should complete SQAF? .....	1
When to complete SQAF? .....	1
What can be expected from completing the SQAF checklist? .....	1
Why is it so long? .....	2
<b>Chapter 2: Survey Inception</b> .....	<b>3</b>
Objectives .....	3
Stakeholders and user groups .....	4
Main stakeholders for the survey .....	4
Other user groups .....	5
Survey sponsors .....	6
Who will sponsor? .....	6
Important issues to consider at the beginning of a survey .....	7
Key survey teams .....	7
Survey implementation team .....	7
Survey advisory team .....	8
Survey implementation team .....	8
Survey Manager .....	8
Field Manager .....	9
Data Manager .....	10
Data Analyst Manager .....	11
External consultants .....	11
Documentation .....	11
Survey inception documents .....	12
Documentation in survey archive .....	12
<b>Chapter 3: Budgeting and Financial Management</b> .....	<b>13</b>
Budget items .....	13
Financial incentives to help in delivery of quality results .....	14
Sensitivity analysis .....	15
Sample size .....	15
Cash flow .....	15
<b>Chapter 4: Sampling Design</b> .....	<b>16</b>
An introduction to sampling terminology for survey managers .....	16
Sampling design .....	18
Technical expertise .....	18
Documentation .....	19
Technical details of the design .....	20
Sampling scheme .....	21
Sample size .....	25
Drawing the Sample .....	26
Sampling during field operations .....	26
Sampling weights .....	27

Replacement of units.....	27
Suggested documents and other resources for the survey archive .....	27
<b>Chapter 5: Questionnaires .....</b>	<b>28</b>
Development of the questionnaire(s).....	28
Languages .....	29
Layout and format .....	29
Objectives .....	30
Supplementary information .....	30
<b>Chapter 6: Fieldwork Organisation and Implementation .....</b>	<b>31</b>
Introduction.....	31
Pre-fieldwork requirements .....	31
Materials for field work .....	32
Staff for field work .....	32
Timetable for field activities .....	34
Gantt chart for field work .....	35
Availability of staff .....	35
Field training .....	36
Contents of general training for field staff .....	36
Contents of training for enumerators .....	37
Contents of training for data entry operators .....	37
Contents of training for field supervisors .....	38
Allocation of workloads .....	38
Results of training activities.....	38
Survey Implementation .....	39
Fieldwork logistics .....	39
Supervision and quality control .....	40
De-briefing sessions.....	41
Field work documentation.....	41
Survey schedules .....	41
Quality assurance documents .....	41
Archiving of documents.....	41
<b>Chapter 7: Data Management.....</b>	<b>42</b>
Introduction.....	42
The Data Management Team .....	43
The Data Manager .....	43
Data Assistant(s) .....	45
Questionnaire Archivist .....	45
Programmers (database and data entry system designers) .....	46
Data Entry Staff.....	47
Fieldworkers: supervisors and enumerators .....	47
The Data Management timeline.....	48
The Questionnaire design.....	48
The Database design.....	49
Data Entry System design .....	49
Staff Training.....	50
Transfer and Logging of Questionnaires to the Central Office .....	51
Data Entry .....	51
Backups.....	54

Double data entry.....	55
Data Quality Assurance .....	55
Anonymisation.....	55
<b>Chapter 8: Tabulations and Data Analysis .....</b>	<b>56</b>
Introduction.....	56
Personnel and software.....	56
The Data Analysts .....	56
Staff and their experience .....	57
Software for analysis .....	57
Data Analysis Plan.....	58
Documentation, review and archiving.....	58
Contents of the Analysis Plan .....	59
Database issues .....	60
Completeness of information .....	60
Documentation.....	61
Table Results and Reporting.....	61
Sample size .....	61
Layout and formatting of tables .....	62
Analytical results.....	62
Technical Issues .....	63
Further Analytical Results and Reporting .....	63
Further Analysis.....	63
Technical Issues .....	64
User Satisfaction .....	64
General .....	64
Timeliness .....	64
<b>Chapter 9: Dissemination of Survey Products .....</b>	<b>66</b>
Changes to planned dissemination policies/procedures.....	66
Survey publications and media used for release .....	66
User groups.....	67
User satisfaction .....	68
<b>Chapter 10: Archiving - documentation and dissemination of microdata.....</b>	<b>69</b>
Introduction.....	69
Documentation in Support of Survey Outputs .....	70
Objectives .....	70
Fieldwork .....	70
Sampling .....	70
Data Management.....	70
Analysis.....	71
Survey Reports/Publications for Dissemination .....	71
Reports for dissemination .....	71
Contents of Dissemination Reports .....	72
Storage of Meta-Data .....	73
<b>Appendix - a data manager's checklist.....</b>	<b>75</b>
Introduction.....	75
Using Unique IDs to Identify Staff.....	75

Database and Data Entry Manuals .....	75
Data Entry System.....	76
Staff Training.....	76
Transfer and Logging of Questionnaires to the Central Office .....	77
Merging Databases.....	77
Backups.....	78
Double data entry.....	78
Exporting the Data.....	79
Anonymisation.....	79

# Chapter 1: Introduction to the SQAF Checklist

## What is the SQAF checklist?

The SQAF checklist is a series of questions written to help survey managers improve the quality of their surveys. It asks questions about multiple aspects of the survey process, encourages checking, documentation and implementation of systems to minimise the introduction of errors and ensure completeness of information.

## Completing the SQAF checklist?

### Who should complete SQAF?

The SQAF checklist has been designed as a tool for the survey manager. While experienced survey managers may be able to tackle all the questions on their own, it is more likely that managers will have to consult or delegate answering the question to specialised members of the survey team. In the latter case, it would be advisable for the survey manager to assess to what extent those questions have been addressed by the technical staff.

### When to complete SQAF?

The checklist is organised in chapters roughly matching stages in the development of a survey. We suggest that survey managers use the SQAF during the execution of the survey, rather than completing it at the beginning or end of the survey. It is hoped that survey managers will use it as a tool that is visited and revisited throughout the survey process. The SQAF checklist should help in developing plans and systems for quality control that will lead to an overall improvement in survey data quality.

While it is unlikely that all the questions in SQAF are relevant to each survey, it is expected that the process of thinking whether the questions apply will be useful in making decisions, gathering information or engaging in discussions with other members of the survey team. This process should help in establishing mechanisms that improve the overall quality of the survey.

### What can be expected from completing the SQAF checklist?

The SQAF checklist should give the survey manager

- A structured guide to the type of information that different members of the survey team should be able to provide if the survey is to be of good quality
- A list of questions that will help in managing and checking survey processes, staff and resources
- A tool to ensure that important aspects of quality control have been incorporated in the survey process

A completed SQAF checklist should also be a good basis to assess different aspects of the quality of the survey process and the data that it generates. Such document is useful to support the survey findings and to help in demonstrating the care taken by the survey team in setting up data quality control systems.

## Why is it so long?

The process of carrying out a survey involves many people and a significant number of stages. The quality of the survey results depends on the ability of the survey team to establish a system that prevents problems or traps them as soon as they occur. The challenge lies in being able to look at the process as a whole but also at the details of the survey work. This check list is long because it attempts to highlight aspects that need to be considered throughout the whole process of the survey. The team that prepared it does not expect a user to go through the whole check list in one go, that would be overwhelming. We envisage specific chapters of the checklist to be used at the appropriate time in the execution of the survey and for the information to be accumulated gradually. Another use of the check list can be as part of training programmes for survey managers.

There is a level of deliberate repetition in some chapters of the questionnaire. For example, anonymisation is discussed in several chapters because of the importance of considering this issue at different stages of the data management and archiving process. We hope that users of the checklist will find this repetition is useful rather than distracting.

The SQAF checklist development team will be grateful for suggestions for improvement and feedback about this tool. Please contact us at (XX).

## Chapter 2: Survey Inception

Bureaux of Statistics and Government Ministries regularly do surveys to gather information about agricultural production, health conditions of the population, national accounts, etc., to enable the government to make policy decisions on the basis of evidence generated by these surveys. Surveys are also conducted by NGOs, research organisations, universities, and other bodies for research and development purposes.

This chapter sets up checklists that will enable a survey manager to review survey start-up procedures. It covers survey objectives, who expressed the demand, how the survey work is being funded and the identification of stakeholders and user groups. It also highlights the key personnel needed during the actual conduct of survey operations.

### Objectives

No study should be undertaken until there is a clear specification of the study objectives and the questions which the study is aiming to address. A survey exercise is no exception! Objectives are often initially stated in broad terms, e.g. to assess the health conditions of children in rural Tanzania, or as a goal, e.g. to improve health conditions of children in rural Vietnam. However, they need to be specified in a more precise way through several statements, usually in the form of questions, which bring out the associated data element. Thus, the above example can lead to sub-objectives or more specific questions such as *What percentage of children have been vaccinated against measles?* *What is the neo-natal mortality rate?* etc. The checklist below addresses clarity in the definition of objectives and persons responsible for this component.

- 1 What is the goal or overall (broad) objective of the survey?

- 2 Who or what organisation or group was primarily responsible for conceiving a need for information defined by the above objective?



3 Was the survey conception based on demand for information that could only be generated through a survey? Yes  No

4 Has the survey objective been re-expressed as several questions which bring out the data element(s) needed to address that question? Yes  No

**If No**, when will this be done, and by whom?

--

5 **If Yes** to question 4, indicate the title of the document which lists questions that address the overall survey objective, and where it is located, or with whom.

Title of document

Where located and/or with whom

6 Who, or which groups, have been involved in specifying the overall objective as a series of data-related questions?

Name of person and/or Institution	Reason for interest in survey

## Stakeholders and user groups

It is common to find that demand for the information that will be generated from a survey arises from different people and institutions. These stakeholders play an important role at many stages of the survey cycle including demand for information, contributing to the funding, conceptualisation of objectives, technical aspects, implementation, use and dissemination of survey results. They can make important contributions to the capacity of the implementing agency, to the public perception of the survey and to the extent to which the results are used. They also provide useful feedback for improvements in future runs of the survey. It is therefore valuable to identify such stakeholders at the time of conceiving the survey need, and to keep these stakeholders involved and/or informed of survey activities.

It is usually the case that the stakeholders are also the main users of survey results. But there are other users, e.g. researchers, NGOs, universities, etc., who would use the survey results for research and development purposes. Identifying such groups is also a valuable activity at the survey inception phase and is covered below.

### Main stakeholders for the survey

7 Have stakeholders been identified? Yes  No

**If Yes**, please state their names, and other details as indicated below

<i>Name of stakeholder</i>	<i>Contact details</i>	Reason for interest	Participated in survey inception or planning
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>

### Other user groups

- 8 Have users of survey results, other than stakeholders listed above, been identified? Yes  No

**If Yes**, please state their names, and other details as indicated below

<i>Name of user</i>	<i>Contact details</i>	Reason for interest	Participated in survey planning
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>

## Survey sponsors

Once the need for information through a survey has been identified, a proposal needs to be set up and funding sought from one or more national or international sponsors. A record of this process is valuable to ensure that related documents enter into the survey archive. The government of a country is included in this group of 'sponsors'. The following checklist has been set up for this purpose.

### Who will sponsor?

- 9 Has funding for the survey been obtained? Yes  No   
**If Yes, who is paying for the survey?**

<i>Sponsor</i>	<i>Country of headquarters</i>	<i>National or International?</i>

- 10 **If No** to question 9, name the sponsors who will be/have been approached for funding the survey:

<i>Sponsor</i>	<i>Country of headquarters</i>	<i>National or International?</i>

- 11 Has a proposal for the survey been written, demonstrating the demand for information gathering, objectives being addressed and users of survey results? Yes  No   
 12 **If Yes**, has this proposal been submitted to one or more sponsors? Yes  No

- 13 **If Yes**, has the sponsor(s) accepted the proposal and approved funding? Yes  No
- 14 **If Yes**, state details relating to the funding below.

<i>Sponsor</i>	<i>Contact name for financial matters</i>	<i>Date approved</i>	<i>Date funding will start</i>

### Important issues to consider at the beginning of a survey

- 15 Have you shared the detailed objectives of the survey with all the stakeholders and sponsors? Yes  No
- 16 Have you asked them for feedback and suggestions? Yes  No
- 17 Have you prepared a policy for dissemination of results? Yes  No
- 18 Have you proposed a schedule for publication of results? Yes  No
- 19 Do you have a policy to allow access to the survey's microdata? Yes  No
- 20 Do you need authorisation to carry out field work? Yes  No
- 21 Have you considered the need for public awareness campaign prior to the field work? Yes  No
- 22 Have you established mechanisms for feedback from the users of the survey results? Yes  No

### Key survey teams

To ensure a well-planned and executed survey, it is valuable to set up two different high level teams:

- *Survey implementation team*. This team is responsible to the sponsor or primary stakeholder for delivery of survey outputs. Sometimes called the Survey 'core' team.
- *Survey advisory team*. This team is formed who have an interest in the survey results for use in policy matters or for research and development purposes, and who could act in an advisory capacity to the implementation team. The function of the *survey advisory team* is to provide advice to the survey implementation team, monitor survey progress, and provide technical inputs where needed.

### Survey implementation team

- 23 Will there be a survey manager, responsible for all aspects of survey design, implementation and production of survey outputs? Yes  No
- 24 Will there be a survey field manager, responsible for all aspects of field implementation? Yes  No
- 25 Will there be a data manager, responsible for data entry, management of all survey data, and the survey data archive? Yes  No
- 26 Will there be a survey statistician who will be available for advisory inputs from the planning stage to the stage of data analysis? Yes  No

- 27 Will there be a data analyst manager (this could be one of the persons above) who will be responsible for delivery of data analytical results? Yes  No

**NOTE: If the answer is No to any of the above, the survey work should not proceed. More details concerning this team appear in checklists below.**

### Survey advisory team

- 28 Who are the stakeholders who will serve on the *survey advisory team*? Give names and affiliations below, and indicate the reason for their interest in participating as an advisory team member.

<i>Name and institution</i>	<i>Contact details</i>	<i>Reason for interest</i>

### Survey implementation team

The composition of the survey implementation team varies depending on factors such as the size of the survey or the institutional setup of the organisation that runs the survey. This section collects basic information about the survey implementation team and gives an overview of the responsibilities of key members of this team as well as the type of personnel that answers to them. In some cases one person may take the role of more than one of the positions described below, and in others the responsibilities may be allocated to more than one person. As the survey manager has responsibility for organisation and recruitment of the survey team, this section may be of help to survey managers in deciding the structure of their team and documenting details concerning the management team.

#### Survey Manager

The survey manager is the project manager for the survey. He/she has direct responsibility for:

- Budgeting and financial management
- Overall supervision of activities and responsibility for timely delivery
- Organisation and when necessary recruitment of staff for the survey team
- In coordination with field manager plan field activities
- Communications with stakeholders
- Coordination of development of survey instruments
- Overall responsibility for quality assurance
- Delivery of survey information products
- Ensure archiving of survey data and metadata

Main personnel that often answer directly to the survey manager are:

- Field Manager

- Data Manager
- Data analyst manager
- External consultants

29 Who is the survey manager for this particular survey?

Name

Title

Contact details

30 Does the survey manager have autonomy over the financial management of the survey project? Yes  No

30.1 **If No**, who makes the final decisions with respect to financial matters?

30.2 Does the survey manager have direct access to this person? Yes  No

30.3 **If No**, what procedures exist when financial issues need discussion?

### Field Manager

The field manager deals with operational matters in the field. More specifically, he/she is responsible for:

- Recruitment and management of field personnel
- Design of field operational procedures and the corresponding manuals
- Ensuring that field staff are trained for field work
- Overall responsibility to ensure the timely and complete execution of field activities
- Overall responsibility for design and implementation of quality assurance procedures in the field.

Main personnel that often answer directly to the field manager:

- Field supervisors
- Enumerators
- Other field staff such as drivers

31 Who is the field manager for this particular survey?

*Name*

*Title*

*Contact  
details*

### Data Manager

The data manager is responsible for all aspects of data management and archiving. In particular, he/she is responsible for:

- Compilation, computerisation and delivery of a clean database (or a collection of datasets) to the survey analysts
- Development of computer programmes and databases for the survey data
- Overall responsibility for data quality assurance systems
- In coordination with field manager establish the data quality assurance system for data in the field.
- In coordination with programmers establish a data quality assurance system for digital storage and management of data

Main personnel that often answer directly to the data manager

- Programmers / Data Assistants
- Data entry clerks and their supervisors
- Data archivist

32 Who is the data manager for this particular survey?

*Name*

*Title*

*Contact  
details*

## Data Analyst Manager

The data analyst manager is responsible for:

- Production of statistical summaries using the data delivered by the data management team
- Ensuring correct data processing and production of results as agreed with the survey implementation team

Main personnel that often answer directly to the data analyst manager

- Data Analysts
- Research assistants

33 Who is the data analyst manager for this particular survey?

*Name*

*Title*

*Contact details*


## External consultants

External consultants, either national or international, provide specialised inputs that require skills not available from within the survey implementation team. While their input may be managed by different members of the survey implementation team, the survey manager needs to ensure that clear terms of reference are prepared for their contributions and that a requirement for full documentation of inputs is always in place.

34 Who were the external consultants (if any) for this particular survey?

<i>Name and title</i>	<i>Contact details</i>	Brief description of contribution

35 Were there written and specific terms of reference for the work of these consultants? Yes  No

## Documentation

Several documents arise at the inception phase of the survey. It is highly desirable that all the key documents are prepared (if not already available) and that they are lodged with the Data Manager for archiving. The checklist below serves this purpose.



**Survey inception documents**

- 36 Is there a document that provides the background and justification for the survey, along with survey objectives and questions to be addressed from survey results? Yes  No
- 37 Is there a document that describes definitions used during the survey, e.g. poverty measurements, household, etc. Yes  No
- 37.1 If standard definitions have been used, provide appropriate references for the origin of these definitions ?

- 37.2 If survey specific definitions have been used, have they been clearly stated without ambiguity and documented? Yes  No
- 38 Is there a document that describes the Terms of Reference for all external consultants? Yes  No

**Documentation in survey archive**

- 39 Have the following documents been lodged with the data manager for archiving?
  - 39.1 The survey proposal Yes  No
  - 39.2 Description of background, justification, and survey objectives Yes  No
  - 39.3 A definitions document (standard and survey specific) Yes  No
  - 39.4 Terms of reference for external consultants Yes  No
- 40 If the Data Manager is still to be appointed, name the person(s) responsible for ensuring the above documents are forwarded to the Data Manager for archiving

Name of person	Role in survey team

## Chapter 3: Budgeting and Financial Management

- 1 Is there a draft budget for the survey? Yes  No   
**NOTE: If the answer is No, the survey work should not proceed.**
- 2 Has the survey secured funding for all its implementation? Yes  No   
**NOTE: If the answer is No, the survey work should not proceed.**

### Budget items

- 3 Have you considered including any of the following in the budget?
- Basic salaries
- |                        |                              |                             |
|------------------------|------------------------------|-----------------------------|
| Survey manager         | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Data manager           | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Field manager          | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Principal data analyst | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Accountant             | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Secretarial support    | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Supervisors            | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Interviewers           | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Data entry operators   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Drivers                | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
- Travel costs
- |                         |                              |                             |
|-------------------------|------------------------------|-----------------------------|
| Survey manager          | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Data manager            | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Field manager           | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Supervisors             | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Interviewers            | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Data entry operators    | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Drivers                 | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Specialised consultants |                              |                             |
- Transport
- |                       |                              |                             |
|-----------------------|------------------------------|-----------------------------|
| Vehicles              | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Fuel                  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Vehicle maintenance   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Other transport costs | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
- IT equipment
- |   |                              |                             |
|---|------------------------------|-----------------------------|
| Computers for data entry                | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Computers for data processing           | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| UPS, stabilisers                        | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Palmtops/other field data entry devices | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Scanner (other optical input devices)   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Printers                                | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Backing up devices                      | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

Software	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Specific measuring tools	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Reproduction of questionnaires	Yes <input type="checkbox"/>	No <input type="checkbox"/>
IT consumables	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stationery	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Pre-survey public awareness campaign	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Translation	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Pilot survey	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Office space	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Training	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Hiring of venues	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Subsistence costs	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Training materials	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Coordination meetings	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Dissemination of survey results	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Web site development	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Specialised consultants		
Fees	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Travel	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Accommodation	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Subsistence		
Contingency	Yes <input type="checkbox"/>	No <input type="checkbox"/>

*This is a generic list of budget items. It is likely that these will be present in most survey budgets. However this list cannot be guaranteed to be exhaustive, neither can it be said that every item listed should appear in a budget.*

- |     |  |                              |                             |
|-----|--|------------------------------|-----------------------------|
| 4   | Have you compared your budget with budgets for previous surveys carried out under similar settings to check for omissions? | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 5   | Can all the skills and technical expertise be found within the institution running the survey?                             | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 5.1 | <b>If No</b> , have you identified external consultants?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 6   | Have you prepared specific terms of reference for each external consultant?  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

### Financial incentives to help in delivery of quality results

In many surveys the payment is associated with the performance of activities such as 'going to the field', 'interviewing', 'data processing', etc. This is payment for 'activities'. In some surveys the payment is conditional on the delivery of specific products, for example 'delivery of a final survey report', 'delivery of a field report according to agreed specifications', 'delivery of a specific number of completed, quality assured questionnaires'. While it is likely that some payment for activities will be needed, the use of payment for products gives the survey manager extra tools to help improve the quality of the survey results. This is done through the use of incentives associated to the timely delivery of good products and the establishment of penalties for failure to deliver the expected products.

- 7 Is the survey using any of the following reward approaches

- |     |  |     |                          |    |                          |
|-----|--|-----|--------------------------|----|--------------------------|
| 7.1 | Payment for activities   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 7.2 | Payment for products   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 8   | Have you considered establishing a system of performance based incentives? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

### Sensitivity analysis

A survey budget is executed over a period of time and therefore may be affected by factors that are outside the control of the survey implementation team. Before a budget is finalised it is convenient to assess how certain hypothetical situations would affect the capacity to deliver the expected survey results. This could be called a "sensitivity analysis".

- |     |  |     |                          |    |                          |
|-----|--|-----|--------------------------|----|--------------------------|
| 9   | Have you considered the following factors as part of the sensitivity analysis for your budget? |     |                          |    |                          |
| 9.1 | Variations in the price of fuel  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 9.2 | Increase in inflation rate   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 9.3 | Devaluation of the currency in which the budget is prepared                                    | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 9.4 | Delays in the data collection process  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

### Sample size

- |    |   |     |                          |    |                          |
|----|---|-----|--------------------------|----|--------------------------|
| 10 | Have you considered how robust your draft budget is to changes in the foreseen sample size? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 11 | Was the budget updated once the design of the survey was completed and the sample drawn?    | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

### Cash flow

- |    |  |     |                          |    |                          |
|----|--|-----|--------------------------|----|--------------------------|
| 12 | Have you assessed the ability of your institution to provide disbursements at the time when they are required?                   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13 | Have you put in place a work plan that ensures that the flow of resources is ensured to be at the time and where it is required? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

## Chapter 4: Sampling Design

### An introduction to sampling terminology for survey managers

The following paragraphs provide brief explanations of technical terms used in sampling that a survey manager should be aware of. They can be read in advance of completing the form or/and referred to when completing the information for this section. They are included here using simple language and without mathematical formulae.

Sampling <sup>1</sup>	Sampling is the process of selecting a number of cases from all the cases in a particular group or universe.
Probability sampling <sup>1</sup>	A probability sample is a sample selected by a method based on the theory of probability (random process), that is, by a method involving knowledge of the likelihood of any unit being selected.
Non-probability sampling <sup>1</sup>	<p>A sample of units where the selected units in the sample have an unknown probability of being selected and where some units of the target population may even have no chance at all of being in the sample.</p> <p>Forms of non-probability sampling are numerous, such as voluntary samples (only responses of volunteers are used), quota samples, expert samples.</p> <p>Data collected using non-probability sampling cannot provide valid conclusions about the whole population – their results are only valid about the members of the sample.</p>
Sampling error <sup>1</sup>	<p>That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a sample of values is observed; as distinct from errors due to imperfect selection, bias in response or estimation, errors of observation and recording, etc.</p> <p>The totality of sampling errors in all possible samples of the same size generates the sampling distribution of the statistic which is being used to estimate the parent value.</p>
Standard error	Measures the variability of the estimate, or precision. The larger the standard error of an estimate the less precise it is.
Non-sampling error <sup>1</sup>	<p>An error in sample estimates which cannot be attributed to sampling fluctuations.</p> <p>Non-sampling errors may arise from many different sources such as defects in the frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations or misunderstanding or bias or negligence or dishonesty on the part of the investigator or of the interviewer, mistakes at the stage of the processing of the data, etc.</p>
Sample size and errors in estimation	Standard errors are inversely proportional to the square root of the sample size. This means that the gain in precision for every extra unit in the sample size is bigger when the sample size is small. As the sample size increases, the benefits of every extra unit in the sample become smaller quite quickly. The practical

<sup>1</sup> Source: OECD Glossary of Statistical Terms. Online version <http://stats.oecd.org/glossary/index.htm> accessed 18 April 2008

	<p>consequences of this are that reasonable precision may be affordable, but extreme precision can be very, very expensive!</p> <p>While large sample sizes tend to provide smaller more precise estimates, survey managers should be aware that large sample sizes increase the occurrence of non-sampling errors. Therefore the gains in precision must be carefully balanced against the risk of introducing errors that cannot be measured.</p>
Sampling unit	A sampling unit is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made.
Stratification	<p>Stratification refers to the division of a population into strata. Strata are non-overlapping subsets of the whole population (often, but not always, geographically defined) within each of which a separate sample is selected.</p> <p>Stratification is usually done with one of these two objectives:</p> <ul style="list-style-type: none"> <li>• To potentially improve the overall precision of the estimates by gaining control over the composition of the sample. For instance, we may want to ensure that the sample contains certain predefined proportions of households headed by men and women, or in urban and rural areas, or in different regions of the country.</li> <li>• To produce estimates for subgroups of the population that otherwise could be poorly represented in the sample. For instance, a non-stratified sample of Argentina will contain a lot of households from Buenos Aires but very few from a less populated province such as Tierra del Fuego. If we want estimates of sufficient precision for all provinces, we need to ensure that our sample contains enough households from each of them.</li> </ul> <p>These objectives are not complementary:</p> <ul style="list-style-type: none"> <li>• If the objective is to obtain precise estimates for the population as a whole, the sample should be allocated among strata more or less in proportion to their population;</li> <li>• If the objective is to obtain estimators of comparable precision for all strata, the sample should be of about the same size in each of them.</li> </ul>
Sampling frame	A list of all members of a population used as a basis for sampling. In multi-staged sampling, sampling frames may be constructed for different stages in the sampling process.
Multi-stage sampling	Multi-stage sampling is a sampling method by which a sample is selected in stages. The sampling units at each stage are sub-sampled from the units chosen at the previous stage. The sampling units belonging to the first stage are called primary or first stage units; and similarly for second stage units, etc. The sampling units at the last stage of the process are called the final or ultimate sampling units.
Sampling weights	Because a sample is used to estimate characteristics of the population, each value in the sample makes a contribution to the estimation of the population parameter. This contribution is its weight. Because of the complex sampling designs that are used, in most cases sampling units carry the different weights, and these weights need to be derived. The derivation of weights is based on the probability of selection of a sampling unit. Weights can be derived as soon as a sampling scheme has been designed, but these weights will need some adjustment after data have been collected to take account of non-responses.

## Sampling design

Sampling design often requires the intervention of experts (either from within the organisation or external consultants) who provide input that ranges from advice on specific points to full development of the sampling methodology.

This section includes detailed information about the people involved in the sampling design and the contributions that they have made. Compiling the information for this section should help in contacting the people responsible for the sampling design when and if necessary.

## Technical expertise

- 1 Who was responsible for the design of the sampling scheme within the implementation team for this survey?

*Name*

*Title*

*Contact details*

- 2 Did the survey team have inputs from a sampling expert from within your institution? Yes  No

2.1 **If yes,**

*Name*

*Contact details*

- 3 Did the survey team have inputs from any external consultants for designing the sampling scheme? Yes  No

3.1 **If yes,**

*Name*

*Contact details*

*Name*

Contact details

--

## Documentation

A sampling scheme is a complex process that needs to be documented in detail. It is important that the survey manager has access to complete documentation about the technical aspects of the sampling scheme. It must be ensured that those technical aspects are described with the level of detail necessary for the implementation of the sampling scheme and for their use during the analysis of the data. All the technical documents need to be kept in an archive that allows easy access to the information when required.

Full documentation is important to enable statistical analysis. Incomplete documentation make it impossible to calculate appropriate estimates of the standard errors for the survey estimates and the survey manager must ensure that details about sampling stages, stratification, clustering and sampling weights are well documented.

Complete documentation of the sampling scheme has to be finalised before the survey takes place, and before any external consultants complete their contracts. The submission of full technical details should be part of the terms of reference of the contract of any consultant engaged by the survey organisers to produce a sampling design. The review of this documentation by a competent statistician is recommended so as to ensure completeness and clarity.

The rest of this chapter of the SQAF should help to ensure that the documentation is complete.

- |   |   |     |                          |    |                          |
|---|---|-----|--------------------------|----|--------------------------|
| 4 | Is there a document that describes the sampling methodology?  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 5 | Has this document been archived as part of the survey metadata?   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 6 | Was the document reviewed by anyone not involved in the design of the sampling scheme to ensure that it describes the sampling scheme in enough detail to allow a competent statistician to implement it? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

6.1 **If Yes**, by whom

Name

--

Contact details

--



### Technical details of the design

The definition of target population<sup>2</sup>, and more importantly of the characteristics of the study population<sup>3</sup>, determine to whom the results of the survey apply. The aim of the design team should be to have the same definition for both populations. However this is not always possible due to practical considerations such as budget or time, or due to problems that arise during the field work. The fact that the definitions do not always match exactly is well known and does not affect the quality of a survey provided that:

- The two definitions are not drastically different and the survey population can be considered close enough to the target population.
- Any differences are fully documented so that the relevant caveats can be written at the time of presentation of the results.

### Population definition

- 7 Is the target population clearly defined in the sampling methodology? Yes  No
- 8 Is there a difference between the target population and the study population? Yes  No
- If yes**, please give details.
- 8.1 Is this difference made explicit in the description of the sampling methodology? Yes  No

### Use of a Master Sample<sup>4</sup>

- 9 Was a master sample used for this study? Yes  No
- 9.1 **If yes**, what master sample was used?.

- 10 Did you take a subsample of the master sample? Yes  No

### Study units

The main study unit in a survey refers to the unit about which information is being collected. For example in a household survey, the main study unit is the household. The study unit frequently is also the ultimate sampling

<sup>2</sup> The target population is the set of elements about which information is wanted and estimates are required.

<sup>3</sup> The study population is the set of units from which the sample is drawn. In sampling there are practical constraints that force the survey to narrow down the target population to a set of sampling units that can be reached.

<sup>4</sup> A sample drawn from a population for use on a number of future occasions, so as to avoid ad hoc sampling on each occasion. Sometimes the master sample is large and subsequent inquiries are based on sub-samples from it.

unit.

A sampling unit is one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made. The ultimate sampling unit may be defined as the smallest unit which is the subject of sample selection.

11 What is the definition of the main study unit in the survey?

12 Is the definition used above a standard definition? Yes  No

12.1 **If yes**, where does the definition come from?

### Sampling scheme

13 Did the sampling scheme use multi-stage sampling to reach the study units? Yes  No

**If No**, go to question 15

13.1 **If Yes**, complete the following information about each stage in the sampling scheme

Sampling stage	Description of the Sampling unit	Sample selected using probability sampling	
		Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stage 1		Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stage 2		Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stage 3		Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stage 4		Yes <input type="checkbox"/>	No <input type="checkbox"/>
Stage 5		Yes <input type="checkbox"/>	No <input type="checkbox"/>

**Note:** In some surveys the sampling stages required to reach study units vary within the same study due to the way units are organised. For example, households in urban areas may be reached by sampling provinces, towns, neighbourhoods, and enumeration areas, while households in rural areas are reached by sampling districts first, and then enumeration areas. Use this section to describe exactly the stages that are proposed to reach the study units, describe all the variations that occur in your survey. Use one table like the one above for each variation in the sampling stages.

*Details of sampling stages*

14 For each stage in your sampling design complete the following information

*First sampling stage*

a Sampling unit: name and definition

b Number of units selected at this stage

out of

c Did you have to construct a sampling frame starting from scratch for this stage?

Yes

No

c.1 **If yes,** How was it constructed?

c.2 **If No,** what sampling frame did you use to select units at this stage?

d When was the sampling frame last updated? (date)

d.1 If this sampling frame was updated for this survey, how was it done?

e Did the survey team carry out a verification of the sampling frame for units at this level?

Yes

No

e.1 **If No,** explain why?

e.2 **If Yes**, how was the verification carried out?

- f Is the final sampling frame stored in electronic format? Yes  No
- g Has the sampling frame for this stage been included in the archive of the survey? Yes  No

**Stratification**

- h Did you use stratification at this stage? Yes  No

If stratification was not used skip to the “second sampling stage”

h.1 **If Yes**, list the strata and provide the information about each stratum

Stratum	Number of units allocated

- i How did you decide the number of units that were allocated to each stratum?
- Equal sample sizes
- Proportional allocation
- Other (please give a reference for the description of the method used)

- j For the selection of units within each stratum did you use...
- Equal probability of selection
- Probability proportional to the size of the unit (PPS)
- Other (please give a reference to the description of the method used)

k If you used PPS, what measurement of size was used?

***Subsequent sampling stages***

Please copy these pages, starting from question 14, for each stage in the multi-stage sampling scheme.

**Sample size**

15 What is the total number of study units that this survey should have according to the sampling scheme?

16 To what extent did the following criteria influence the final sample size for the survey?

	No influence	Moderate influence	High Influence	Very high Influence
Budget	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Precision for key estimates	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

17 If the sampling scheme was designed with a view to achieving a specific level of precision for one or more particular population characteristics, please state

Population characteristic being estimated	Level of precision aimed for <i>(indicate value of allowed standard error, width of confidence interval or coefficient of variation)</i>
---	---


### Drawing the Sample

- 18 Is the sampling frame stored in electronic format? Yes  No
- 19 Does the sampling frame contain a full listing of all the sampling units for every stage in the sampling scheme? Yes  No

**If No,** to what stage in the sampling scheme can this sampling frame be used?

**If the sampling frame is not available in electronic form,** go to “Sampling during field operations”

**If an electronic sampling frame (or partial electronic sampling frame) was used,**

- 20 Was the sample drawn using a computer? Yes  No

**If yes**

- 20.1 What software was used to draw the sample?

- 20.2 Was there a computer program, syntax or script, written to draw the sample? Yes  No

- 20.3 Who wrote this program?

*Name*

*Contact details*


- 20.4 Has this program been archived? Yes  No

### Sampling during field operations

- 21 Did the field teams carry out sampling in the field? Yes  No

**If yes**

- 21.1 Who was responsible for drawing the sample?

21.2 Where is the methodology for such sampling described?

21.3 Were any field listings produced to generate sampling frames? Yes  No

21.4 Were the field listings computerised? Yes  No

21.5 What checks were put in place to ensure that the sampling was done following the planned methodology?

### Sampling weights

22 Who is in charge of producing the sampling weights for this survey?

Name

Contact  
details

23 Is there documentation that describes the derivation of these weights? Yes  No

23.1 If Yes, does this include instructions on how to adjust for non-response? Yes  No

### Replacement of units

24 Is there a defined procedure for replacing non-respondents at the ultimate sampling stage? Yes  No

24.1 If yes, where is it documented? Yes  No

### Suggested documents and other resources for the survey archive

25 Are the following documents ready for archiving?

25.1 Description of sampling methodology Yes  No

25.2 Technical details for derivation of weights Yes  No

25.3 Sampling frame including any listings carried out in the field specifically for this survey Yes  No



## Chapter 5: Questionnaires

### Development of the questionnaire(s)

1. Who has the responsibility for managing the process of questionnaire development?

Name

Contact  
details


2. Was the questionnaire developed in consultation with the stakeholders? Yes  No

**If Yes,** which of the stakeholders contributed to the development of the questionnaire?

3. Was the questionnaire pre-tested<sup>5</sup>? Yes  No

4. Was there a pilot test of the questionnaire?<sup>6</sup> Yes  No

**If Yes,**

- 4.1. In how many sites was the pilot test carried out?

- 4.2. How many interviews were carried out in total?

- 4.3. Who was responsible for making changes to the questionnaire after the pilot test?

5. Is there a clear system to identify different versions of the questionnaire?

Yes  No

6. What is the final version of the questionnaire?

<sup>5</sup> Pre-testing a questionnaire often refers to a test carried out with a reduced number of interviewees to have an initial assessment of how the questionnaire works. See pilot test below for comparison.

<sup>6</sup> Pilot testing refers to a test of the questionnaire and field procedures carried out in a number of sites chosen to reflect the diversity of conditions under which the survey will run, with interviews carried out on subjects that would qualify for the survey, by the type of interviewers and supervisors that would be part of the field teams.

## Languages

- 7 Is the questionnaire to be administered in the language of the respondents? Yes  No

- 7.1 **If Yes**, state the languages and tick the boxes to confirm that for each language the questionnaire has been translated, back-translated and approved.

Language	Translated	Back-translated <sup>7</sup>	Checked and approved
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Layout and format

- 8 Has the data manager checked the questionnaire design from the point of view of layout for data entry (scanning or CAPI)? Yes  No

- 9 Has the questionnaire been checked for
- a. Appropriate flow of questions Yes  No
  - b. Skip instructions Yes  No
  - c. Appropriate use of language Yes  No
  - d. Clear instructions for the interviewer Yes  No
  - e. Appropriate use of units of measurement Yes  No
  - f. Sensible recall periods Yes  No
  - g. Consistent and sufficient pre-coding of questions Yes  No
  - h. Codes for different types of missing values and non-responses Yes  No
  - i. Sensible duration of the interview Yes  No

- 10 Does the questionnaire have a prominent place where a unique identifier will be recorded? Yes  No

- 11 Is there a control panel<sup>8</sup> that formalises the process of passing on Yes  No

<sup>7</sup> Back translation refers to the process of translating the questionnaire into the original language to check for loss of meaning in the questions as a result of translation. This needs to be done by a different translator and an assessment of the level of mismatch between the original questionnaire and the back translated questionnaire is made to decide whether any unacceptable deviations from the original have been introduced

- responsibility for the questionnaire from interviewer to supervisor to the data management team?
- 12 Does the questionnaire include a section to be completed by the supervisor after he/she has checked it for completeness, clarity and consistency in the data? Yes  No

### Objectives

- 13 Has the questionnaire been checked to ensure that the information collected is sufficient to address the objectives of the survey? Yes  No
- 13.1 **If Yes**, who carried out this check?

### Supplementary information

- 14 Are there any other interview tools apart from the questionnaire such as response cards or codebooks? Yes  No
- 14.1 **If Yes**, list the interview tools that have been prepared for the survey

- 15 Is there an enumerator (or interviewer) field manual containing explanations about how to administer the questionnaire? Yes  No
- 16 Is there a supervisor field manual containing explicit instructions about how to conduct supervision? Yes  No
- 17 Is there a format for reporting field activities for  
Enumerators Yes  No   
Supervisors Yes  No
- 18 Are the GPS coordinates of location of respondents to be collected? Yes  No

---

<sup>8</sup> By Control Panel we mean a place for the interviewer, supervisor and data entry clerks to sign and date the questionnaire when they have completed their tasks for that questionnaire

# Chapter 6: Fieldwork Organisation and Implementation

## Introduction

Quality assurance during field organisation and implementation is well recognised as an important component in survey operations since it has a direct impact on generating high quality and reliable data. An outline of procedures that contribute to this component, and staff needed for this purpose, would have been discussed and documented at the survey planning stage to ensure an adequate budget and time allocation for all field operations. In this chapter, we re-visit and expand on these procedures (e.g. field staff training, supervisor checks on survey returns, etc), by providing checklists that aim at assuring survey managers that quality control is in place to give reliable survey data, and hence good quality survey outputs.

Except for the first section below – which lays down the necessary pre-requisites for field work, the remaining sections in this chapter appear in the chronological order in which activities will take place in the field. Checklists provided in each section are written in the past tense so that once the survey fieldwork has been completed, the level of quality can be assessed. However, it is important to note that every checklist should be read in advance of the field work so that appropriate steps can be taken to ensure the quality criteria implied by the checklists are met. It would also be highly desirable to make reference to these checklists, during, and as the field work progresses, since this will help maintain the necessary level of quality that the survey requires.

It is assumed that the final responsibility for field quality assurance lies with the Field Manager – whose responsibilities have been outlined in Chapter 2 of this manual. However, in most surveys there would be further team members who are responsible for the survey objectives and design, e.g. survey manager, data manager, field work trainers, stakeholders, as defined in Chapter 2. Field work supervisors also feature prominently in this chapter, and in most cases, the field manager would also serve as one of the supervisors. One or more of the implementation team members could also serve as supervisors.

Note that data entry is generally the responsibility of the data manager and is thus covered in more detail in Chapter 7. However, it is also included here for those cases where the data entry is carried out in the field.

## Pre-fieldwork requirements

To start field operations, it is obvious that the necessary staff must be in place, they must be trained, and materials needed for the training components and survey implementation must be ready at hand. The number of staff needed and the time needed for field activities will of course depend on the nature of the survey, the length of the questionnaire, the time-scale allowed or available for field work, and the number of sampling units to be covered during data collection. With respect to enumerator numbers, it is desirable to train a larger number than is required in order to allow for possible drop-outs and scope for removing others who fail to reach the necessary standards following their training for survey work. The checklists below are aimed at ensuring that the necessary staff are available and that all other requirements are met, prior to commencement of field organisation and implementation.

**Materials for field work**

- 1 Has the field instruction manual been completed with sufficient copies for distribution to field staff? Yes  No
- 2 Is the final version of a pilot-tested survey instrument available, with sufficient copies for conduct of fieldwork and for use during fieldwork training? Yes  No
- 3 Has the sampling plan been finalised and documented? Yes  No
- 4 Has the sampling coverage (e.g. geographical areas, households or other units) been clearly defined and documented for allocation to field teams at the time of training? Yes  No
- 5 Has the data entry program been designed and tested against the finalised questionnaire or other data recording instruments? Yes  No
- 6 Has the team needed for leading the field team, training of field enumerators, supervising the field work, collating and checking survey returns and passing them to the data entry team, been established? Yes  No

IF “NO” IS THE ANSWER TO ANY OF THE ABOVE, PROCEED NO FURTHER

**Staff for field work**

7 Who was the field manager for this survey?

Name	
Title	
Contact details	

8 Who conducted training for the master trainers? (If relevant only, e.g. when the survey is large):

<i>Name</i>	<i>Job Title or role in team</i>	<i>Contact details (email, phone)</i>

9 Who conducted training for the enumerators? (Include master trainers if any)

<i>Name</i>	<i>Job Title or role in team</i>	<i>Contact details (email, phone)</i>

10 Who conducted training for the data entry operators? (Include master trainers if any)

<i>Name</i>	<i>Job Title</i>	<i>Contact details</i>

10.1 Were all the above trainers closely associated with the development of the data entry system and/or screens set up for data entry? Yes  No

10.2 **If No**, explain how the those named above acquired this familiarity

11 Who conducted training for the supervisors?

<i>Name</i>	<i>Role in survey team</i>	<i>Contact details</i>

11.1 Were all the above trainers fully aware of the survey objectives and expectations? Yes  No

11.2 Were all supervisors present during the enumerator training? Yes  No

- 12 Has there been a fair and transparent process, e.g. by interview, to select enumerators for training Yes  No

If Yes, please give details of the selection process:

- 13 Was there a procedure in place to select enumerators that were “best” in terms of reaching the highest level of competency after training? Yes  No

If Yes, please give details of the selection process after training:

- 14 Give the number of enumerators who were finally selected to participate in survey work?

Type of enumerator	Number
1. Professional/Permanent	
2. Recruited for survey	
<b>Total =</b>	

- 15 Were additional enumerators appointed or reserved as backups for possible enumerator drop-outs during survey work? Yes  No

If yes

- 15.1 How many enumerators have been kept in reserve?

- 15.2 Is the number in reserve included under “Total” in question 14 above?

Yes  No

- 16 How many supervisors were appointed to quality check the work of the enumerators?

- 17 Was the number of supervisors adequate in relation to the number of enumerators?

Yes  No

### Timetable for field activities

It is important to have a timetabled list of activities for organisation of the field work and its completion, prior to field work training and implementation. A draft time schedule for field work may have been set up at the time of survey planning, but it is desirable to re-visit the plan since modifications may be required in view of findings during the piloting of the survey instrument (e.g. length of time to complete the recording schedule) and the coverage agreed when developing the sampling plan. It is also likely that the initial draft schedule was not sufficiently detailed in providing for activities within the time period allocated for field work.

Staff availability is also relevant when finalising the field activity schedule. The checklist below will assist the field manager in confirming that all relevant activities have been scheduled in a realistic way.

### Gantt chart for field work

- 18 Was a Gantt Chart, giving the time frame (in weeks say) for each field activity, available prior to commencement of field work? Yes  No

**If No**, when was the Gantt Chart completed? (Give date)

- 19 Were the following activities timetabled into the Gantt chart?

19.1 Training of Master Trainers (where needed) Yes  No

19.2 Training of field supervisors Yes  No

19.3 Recruitment and training of field enumerators Yes  No

19.4 Recruitment and training of data entry personnel Yes  No

19.5 Implementation of survey work by each group of enumerators Yes  No

19.6 Supervision of field work by field supervisors and/or field manager Yes  No

19.7 Planning the time allocation to field personnel according to their roles and requirements in the field Yes  No

- 20 Was there more than 1 week's gap between field training and implementation of survey work? Yes  No

**If Yes**, explain reasons for delay and how it affected trainees recall of issues covered during training

- 21 Was sufficient time allocated for field work to allow supervisors to check questionnaires and for enumerators to re-visit households and secure correct answers to questions where only proxy<sup>9</sup> answers were available from the interviewed household member? Yes  No

### Availability of staff

- 22 Which persons were involved (state their title or role in survey team) with the activities below, and what was their total count?

Activity	Job Title or role in team	Total number
(a) Training of master trainers (if any)		
(b) Training of field supervisors		
(c) Training of field enumerators		
(d) Training of data entry operators		

<sup>9</sup> A proxy answer is an answer provided by a member of the household on the basis of information that he/she has about other member of the household or an aspect of the household affairs that he/she is not fully involved with.



23 How many persons in the following groups were trained:	Number
(e) Master trainers (if any)	
(f) Field supervisors	
(g) Field enumerators/interviewers	
(h) Data entry operators	

- 24 Was the allocation of time in the Gantt chart sufficient, given the numbers stated in questions 22 and 23 above? Yes  No
- If No**, state what action will be/has been taken to overcome this difficulty

## Field training

The training of the enumerators (interviewers), and the data entry personnel, as well as those responsible for supervising their field work, must precede survey implementation. Every effort should be made to conduct the training at a central location so that all trainees receive the same training and have a common understanding of the survey objectives and data quality requirements. Where it is not possible to conduct the training at a single location, training of a relevant number of “master trainers” must take place, ensuring their understanding of survey objectives and ability to follow and communicate field manual instructions to their trainees is on par with that of the team of trainers.

Allocating sufficient time for the training is crucial. The amount of time needed will depend on the nature of the survey. For national level surveys, allocating four weeks for the training component is not unusual. For large surveys of this type, considerable time will also need to be allocated for the training of master trainers.

In the checklist below, reference has been made to enumerators and data entry personnel as though they are different groups, but there may be surveys where some field staff fulfil both functions. Where the groups are different, each group should have a reasonable awareness of the other group’s work. This strengthens the quality of results from the whole field operation.

### Contents of general training for field staff

- |  |                              |                             |
|--|------------------------------|-----------------------------|
| 25 Did the general training for field staff cover the following areas?                                       |                              |                             |
| 25.1 Survey objectives   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 25.2 Expected use and users of survey results  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 25.3 Questionnaire discussion and ensuring each question and its purpose was well understood by the trainees | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 25.4 How the questionnaire data relate to survey objectives  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 25.5 Who to consult if unexpected happenings occurred  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 25.6 Importance of documenting unforeseen events   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

25.7	Importance of generating reliable survey data	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
25.8	Demonstration and/or discussion about the software being used for data entry	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

### Contents of training for enumerators

26	Did the specific training for enumerators cover the following areas?				
26.1	An overview of the planned sampling strategy	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.2	How to construct the sampling frame (where unavailable) at a given site	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.3	The process of selecting the sampling unit (e.g. household) from the sampling frame	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.4	Process of introducing the survey to the respondent	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.5	Action to take if selected households (or other enumeration units) refuse to participate	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.6	Discussion of the Field Instructions Manual	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.7	The need to check all questions have been asked and answers recorded for every single question	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.8	How to minimise non-responses to particular questions	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.9	Whether proxy answers can be accepted for particular questions	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.10	The need to include notes of reasons for missing responses (even if the note say “cannot tell”)	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.11	Highlighting questions which would allow internal consistency of the questionnaire to be checked while in the field	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.12	Practical work in small groups doing “mock” interviews with each other	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.13	Practical work in the field situation to test trainees ability to conduct successful interviews	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.14	Practical work concerning equipment (where relevant), e.g. use of equipment for anthropometric measurements; weighing scales	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.15	An assessment of enumerator capability before they are accepted to the survey team	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.16	Clarity regarding who receives the completed questionnaire returns and timetable for this activity during the period of field work	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.17	Incentives (not necessarily in monetary terms) to increase motivation amongst enumerators to collect reliable data during fieldwork	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
26.18	Clarity regarding potential penalties for neglect or untimeliness in their work	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

### Contents of training for data entry operators

27	Did the specific training for data entry operators cover the following areas?				
27.1	Software being used for data entry	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
27.2	Familiarity with computer equipment being used				
27.2	Practice in entering “mock” data, e.g. those generated during piloting of the questionnaire or practice interviews done by enumerators	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

27.3	Understanding and dealing with skip questions	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
27.4	What to do when unexpected/incorrect data are found in data sheets	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
27.5	Dealing with dates and agreeing on a common format for use	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
27.6	How to deal with missing answers	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

### Contents of training for field supervisors

28	Did the training for field supervisors cover the following areas?				
28.1	Familiarity with the finalised questionnaire and survey objectives	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
28.2	Clear understanding of the survey areas where teams will operate?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
28.3	Frequency of undertaking specific (documented) supervision tasks with each team being supervised	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
28.4	Checklist(s) to be completed to assess quality assurance in each field site	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
28.5	Attendance (as observers) of enumerator training activities	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
28.6	A system for providing feedback to enumerator trainees on their interview practice sessions				

### Allocation of workloads

29	Was the allocation of enumeration teams to selected sampling units (e.g. households), and a timetable for visits, discussed and agreed?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
30	Was the system for supervisor checking of completed questionnaires and onward transfer of the questionnaires to the data entry teams, discussed and agreed?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
31	Was the allocation of supervisors to enumeration/data entry teams, and time for these activities, discussed and agreed?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
32	Have the above allocations, including dates for visits, etc., been documented?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>

### Results of training activities

33	Did problems or difficulties arise during training on the use of the Field Instruction Manual?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
	<b>If Yes</b> , was the Field Instruction Manual updated to overcome such problems and/or difficulties	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
34	Did problems or difficulties arise during training data entry personnel on use of the data entry system?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
	<b>If Yes</b> , was the Data Entry System updated to overcome such problems and/or difficulties	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
35	Have any additional problems, encountered during the training, been documented and a copy lodged with the survey manager?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
36	Was the finalised schedule of plans for all field activities shared with all field workers?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
37	Was a training evaluation completed by enumerators?	Yes	<input type="checkbox"/>	No	<input type="checkbox"/>
	<b>If Yes</b> , what is your perception from this evaluation about the extent to				

which the enumerator training has been successful?

Very successful       Successful       Reasonably successful       Not successful

- 38 Was a training evaluation completed by supervisors? Yes  No   
**If Yes**, what is your perception from this evaluation about the extent to which the supervisor training has been successful?

Very successful       Successful       Reasonably successful       Not successful

## Survey Implementation

Quality assurance during survey implementation is largely concerned with the smooth operation of logistics underlying the field work and quality checking the activities of the field supervisors, enumerators and data entry operators. The logistics must be considered in thorough detail in advance of field work to match fieldwork activities and ensure the efficient use of time by all field personnel and other support staff like vehicle drivers. Having a day-to-day timetable of activities to be done by each field team member is a MUST. Sufficient time must also be available to allow at least 1 day per week of off-duty time for every staff member so as not to endanger the quality of the information collected through fatigue and/or illness. Arrangements should also be in place to provide adequate back-up should any member of the team fall sick during the survey period.

### Fieldwork logistics

- 39 Was there a documented plan which shows the processes and time schedule for smooth transfer of field data returns to supervisors for quality checks, then to the data entry teams, and finally to the field manager? Yes  No
- 39.1 Did the above plan include a system for recording safe receipt of questionnaires from the field to the supervisor and then into the data entry operation? Yes  No
- 39.2 Did the plan allow tracing the status of each questionnaire from field operations to data entry personnel and finally to field manager? Yes  No
- 40 Was adequate transport arranged and drivers allocated to teams according to the documented field work plan? Yes  No
- 41 Were procedures (budget, timeliness, etc) in place to ensure fuel, oil and vehicle maintenance issues were dealt with adequately during field work? Yes  No
- 42 **For large surveys:** Was there a break within a week or two of field work commencement to discuss and resolve any problems encountered? Yes  No
- 42.1 Where problems were found above, were field instruments or work schedules updated if necessary and lodged with the survey manager? Yes  No
- 43 Was there a documented plan to show how computerised files were transferred to the data manager and the data analysts? Yes  No
- 44 Was there an adequate number of computers available for data entry to proceed immediately after the first batch of completed survey returns were received? Yes  No
- 45 Was there a system in place to take regular back-ups of data already entered? Yes  No
- 46 Was there a system to provide additional staff or time in case a survey team member falls sick? Yes  No
- 47 Was there a planned and documented timetable for payment of salaries and expenses of field workers? Yes  No

- 47.1 **If Yes**, was this schedule made clear to the field workers? Yes  No
- 47.2 **If Yes**, was the plan for payments kept to within 1-2 days of the stated timetable? Yes  No

### Supervision and quality control

- 48 Did each supervisor accompany the enumerator team(s) under his/her responsibility? Yes  No   
**If No**, did the supervisor regularly undertake the following tasks
- 48.1 Check that the team(s) visit the correct households according to their visit schedule Yes  No
- 48.2 Check that team(s) follow instructions in field manual correctly Yes  No
- 48.3 Make unannounced visits to team(s) to observe their interview process Yes  No
- 48.4 Meet up with team(s) to discuss progress and any difficulties Yes  No
- 49 Were the following tasks undertaken by the supervisor to check on the quality of field returns?
- 49.1 Follow a standard checklist to confirm quality and accuracy of answers recorded by enumerators Yes  No
- 49.2 Check that all questionnaire returns due from the team are received in time Yes  No
- 49.3 Re-visit a few households, randomly selected from his/her team's allocation, to confirm a few of the key results in the questionnaire Yes  No
- 49.4 Keep a record of results of all his/her the quality checking procedures Yes  No
- 50 Did the supervisor follow procedures outlined below?
- 50.1 Check that teams complete all activities according to their work schedule Yes  No
- 50.2 Meet deadlines for sending checked survey returns at regular intervals to the relevant data entry team Yes  No
- 50.3 Keep in regular contact with the field manager to report progress and any field related problems Yes  No
- 50.4 Keep a record of any problems that arise and steps taken to resolve these Yes  No
- 51 Did the field manager, or other survey implementation team member, follow procedures outlined below?
- 51.1 Make unannounced visits to the field to check progress and to check the work of the supervisor(s) Yes  No
- 51.2 Make a few random checks on the questionnaire returns of each supervisor Yes  No
- 51.3 View some results from entered data and check the data for possible inconsistencies or other errors Yes  No
- 51.4 Prepare a draft field report of difficulties encountered during field staff recruitment, field training and/or field implementation, and steps taken to resolve these and to record a summary of activities undertaken Yes  No

### De-briefing sessions

- 52 Was there a de-briefing meeting with survey implementation team members and field staff to discuss the draft field report and to make suggestions for improvement in future surveys Yes  No
- 53 **If Yes**, did the de-briefing session record the overall response rate, i.e. the percent of field returns completed relative to the coverage planned in the sampling document? Yes  No
- 54 **If Yes**, state this response rate as a percentage Response rate =    %
- 55 Were reasons for non-responses and recommendations from the de-briefing documented and included in the field report? Yes  No

### Field work documentation

It is inevitable that many of the documents prepared in advance of field work, e.g. field manuals, survey instruments, etc., will need revision during training activities and during survey implementation. Any changes need to be captured immediately, and the final documents transmitted to the survey manager as well as the data manager for archiving purposes. The checklist below ensures that all necessary documentation has been captured.

### Survey schedules

- 56 Have the following documents been viewed since completion of field activities and updated?
- 56.1 Questionnaire and/or other survey instrument Yes  No
- 56.2 Field instruction manual for enumerators Yes  No
- 56.3 Field instruction manual for supervisors Yes  No
- 56.4 A report relating the original sampling frame to sampling actually achieved in the field, and reasons for any differences Yes  No

### Quality assurance documents

- 57 Have the following documents been viewed since completion of field activities and updated?
- 57.1 Checklist(s) for supervisors to confirm quality assurance in field Yes  No
- 57.2 A report on errors found with data entry and steps taken to resolve them Yes  No
- 57.3 A report on difficulties found during field implementation and steps taken to resolve them Yes  No

### Archiving of documents

- 58 Have all survey schedules listed above been lodged with the data manager for inclusion in the survey archive? Yes  No
- 59 Have all quality assurance documents listed above been lodged with the data manager for inclusion in the survey archive? Yes  No

# Chapter 7: Data Management

## Introduction

Data management is the function that provides access to data, monitors the storage of data and controls input/output operations. The aim of data management is to produce reliable, consistent and fully-documented datasets that can be analysed throughout the survey and archived at the end in such a way that they can be used by other data users well into the future<sup>10</sup>. Data management continues throughout the lifespan of the survey and beyond. It should not be confused with data analysis which is the process of transforming raw data into usable information.

The questions included in this chapter of the Survey Quality Assessment Framework are intended as a guide to the desirable characteristics of the data management process. Not all of them are essential for good data management, but the questions are asked to help survey managers identify some aspects that should be considered in establishing a good data management system. An appendix at the end of the SQAF adds more detail to some of the questions given here. These more in-depth questions are to help the data manager in his/her work.

Although we have used the past tense in phrasing the questions in this section, the document is designed as a checklist to be used at all stages of the survey (before, during and after).

There are two main sections in this chapter:

1. In the first section we consider the Data Management Team and the roles of staff that make up this team. Depending on the size and complexity of the survey the data management team changes in size and in practice team members often take responsibility for several functions, or a single function might be shared by more than one person.
2. Data management can also be looked at from a process point of view and in the second section, under the heading Data Management Timeline, a number of questions have been included that highlight critical data management issues at different stages in the survey.

---

<sup>10</sup> Data archiving and metadata documentation is greatly helped by tools such as the Microdata Management Toolkit prepared by the IHSN (<http://www.surveynetwork.org>)

## The Data Management Team

In a typical survey there are many people who handle the data prior to analysis. By data in this context we are also referring to the information contained in the completed questionnaires and not just the electronic database. These “data handlers” all have areas of responsibility for data management and are all part of the data management team. This section looks at issues of data quality by asking questions and providing checklists for specific roles of staff that form the data management team. These roles include:

- Data manager
- Data assistants
- Questionnaire archivist
- Database designer
- Data entry staff
- Fieldworkers (enumerators and supervisors)

### The Data Manager

The data manager is a senior member of the survey team on a par with the fieldwork manager and the principal investigators. He/she is one of the core members of the survey team. In this document we use this title for the person who is in charge and has overall responsibility for the development, implementation and quality assurance of the data management system.

The data manager must have experience of team management and should have a good working knowledge of database management systems. Ideally he/she will have previous experience of handling data from large-scale household surveys.

**A word of warning:** Traditional training of statisticians rarely gives them the skills needed to be good data managers. In general, the tasks of a data manager are better performed by someone with good management and computer skills.

1 Who was the data manager for this survey?

*Name*

*Title*

*Contact  
details*




- 2 Were specific <sup>11</sup>terms of reference (TORs) written for the data manager? Yes  No
- 2.1 **If yes**, are the TORs documented and stored with the survey documentation? Yes  No
- 3 Was the data manager engaged with the survey...
- 3.1 from the design stage? Yes  No
- 3.2 to the archiving stage? Yes  No
- 4 Did you expect to need a data manager for the entire survey? Yes  No
- If no,**
- 4.1 At which stages of the survey did you expect the data manager inputs?
- 
- 4.2 What arrangements were made for assurance and coordination of the data management process?
- 
- 5 Was the data manager employed full-time or part-time for the survey? Full-time   
Part-time
- 5.1 **If part-time**, what percentage of his/her time did he/she work on the survey?
6. Did the data manager have previous experience managing survey data? Yes  No
- 6.1 **If yes**, please give details.
- 

---

<sup>11</sup> The terms of reference (TORs) is a document listing specific tasks and responsibilities and the characteristics of expected outputs.

6.2 **If no**, how was the data manager deemed to be qualified for this role?

### Data Assistant(s)

In larger surveys one or more data assistants are employed to help the data manager with organising the data management system, running data comparisons and consistency checks, backing up of the data, etc. Often these assistants are only employed for a limited time directly after data collection.

7 Were there any data assistants employed for the survey?

Yes

No

**If yes,**

7.1 How many data assistants were employed?

7.2 Who was the line manager of the data assistants?

### Questionnaire Archivist

The questionnaire archivist would be responsible for:

- Logging receipt of completed questionnaires as they are delivered to the central office
- Matching the incoming questionnaires to the number expected according to the sampling design
- Filing and storage of completed questionnaires
- Organising a signing-out system for the questionnaires so that he/she knows the whereabouts of all questionnaires at any time
- Keeping track of which questionnaires have been entered and/or scanned.

The questionnaire archivist reports directly to the data manager.

8 Who was the questionnaire archivist for this survey?

<i>Name</i>	
<i>Title</i>	
<i>Contact details</i>	

9 Did this person also have another role within the survey? Yes  No

9.1 If yes, please give details

--

**Programmers (database and data entry system designers)**

The survey database is a complicated instrument often including tables, forms, queries, reports and automated modules. These modules add specific functionality such as automatic skips in the data entry forms and range checks for the variables. Designing the database is a highly skilled task and often external consultants are brought in either to assist in the design or to produce the entire database.

10 Who was responsible for the development of the **survey database**?

<i>Name</i>	
<i>Title</i>	
<i>Contact details</i>	

11 Were there clear and specific TORs for the programmer developing the database? Yes  No

11.1 If yes, are these TORs documented and stored with the survey documentation? Yes  No

12 Who was responsible for the development of the **survey data entry system**?

<i>Name</i>	
<i>Title</i>	
<i>Contact details</i>	

13 Were there clear and specific TORs for the programmer developing the data entry system? Yes  No

- 13.1 **If yes,** are these TORs documented and stored with the survey documentation? Yes  No
- 14 Were there plans to scan the questionnaires for optical character recognition? Yes  No
- If yes,**
- 14.1 Who was in charge of developing or adapting the software for data capture?
- 
- 14.2 Were there clear and specific TORs for the programming developing the scanning software? Yes  No
- If yes,** are these TORs documented and stored with the survey documentation? Yes  No

### Data Entry Staff

There are generally two levels of data entry staff. First there are the data entry supervisors. As their title suggests they supervise the data entry. They must be familiar with the data entry system and able to answer questions from the data entry operators. They collect the questionnaires from the archivist and allocate batches to the data entry operators. They must keep track of any problems encountered with data entry and they report directly to the data manager.

The data entry operators carry out the bulk of the data entry. These are often unskilled staff employed solely during the data entry phase of the survey. They must be trained in the use of the data entry system. They report to their allocated supervisor.

- 15 How many data entry supervisors did you employ for the survey?
- 16 How many data entry operators did you employ for the survey?
- 17 Before being employed did the data entry staff have to pass any tests to ascertain their skill level? Yes  No
- 17.1 **If yes,** please give details of the testing process
- 
- 18 How were the data entry operators remunerated? Fixed salary   
Per questionnaire   
Per hour

### Fieldworkers: supervisors and enumerators

The enumerators and their supervisors play an important role in ensuring the quality of the data collected by the survey. Enumerators collect the data and must complete the questionnaires accurately and clearly. They should sign and date the questionnaires once they have completed the interview. They report directly to their supervisors. Their training must include sessions on completing the questionnaires.

Supervisors collect the completed questionnaires from the enumerators and must check each questionnaire

for completeness, accuracy, clarity and consistency. Ideally, those designing the questionnaire will produce a checklist detailing all the visual consistency checks to be done on the completed questionnaires. The supervisors should work through the checklist for every questionnaire and should sign and date the questionnaire once they are satisfied that the data recorded are consistent. By signing the questionnaire the supervisor takes responsibility for the quality of the data in that questionnaire.

Depending on the technology and method of administration of the questionnaire chosen by the survey, specific tasks of fieldworkers may change slightly, however their main responsibilities remain pretty much the same. Chapter 6 includes detailed questions concerning the work of the enumerators and supervisors.

- 19 How many fieldwork supervisors were employed for the survey?
- 20 Were the supervisors given a checklist to be completed to ensure each questionnaire is complete, accurate, clear and consistent? Yes  No
- 21 How many enumerators were employed for the survey?
- 22 Were the lists of supervisors and enumerators included in the database before the start of data entry? Yes  No

## The Data Management timeline

Data management can also be looked at from a process point of view. In this section a number of questions have been included that highlight critical data management issues at key stages in the survey. These are:

- Questionnaire design
- Database design
- Data entry system design
- Staff training
- Fieldwork
- Transfer and logging of questionnaires to the central office
- Data entry
- Data Checking
- Archiving

## The Questionnaire design

The data manager should consider the questionnaire from the data management point of view. He/she should advise on layout bearing in mind ease of data collection and ease of data entry. While the data manager and/or the database designer do not decide on the questions to be asked they contribute to the design of the questionnaire to suggest desirable characteristics of the layout and structure. For example, they will highlight the need to provide space at the front of the questionnaire for the enumerators, supervisors and data entry staff to sign and date, ensure that appropriate space is provided for writing answers to questions such as “other please specify”, and in general link the database and questionnaire structures.

- 23 Was the data manager involved in the questionnaire design? Yes  No
- 24 Was there a data handlers section at the front of the questionnaire with space for:
- 24.1 The enumerator to write their name, sign and date after completion Yes  No

- of the interview?
- 24.2 The fieldwork supervisor to write their name, sign and date after he/she has completed the data quality checks and is satisfied with the quality of the data? Yes  No
- 24.3 Data entry operator to write their name, sign and date after he/she has entered the data into the database? Yes  No
- 25 Did the printed questionnaire include the variable names for each question? Yes  No
- 26 Was the key identifier for the household repeated at the top of each page of the questionnaire? Yes  No

### The Database design

- 27 Was the database delivered to the required specification and on time? Yes  No

27.1 **If no**, please give details.

- 28 What software was used for the database design?

- 29 What were the reasons for choosing this particular software?

- 30 If the database was developed by an external consultant or with the help of an external consultant, was there capacity in the data management team to make changes to the database as needed? Yes  No

30.1 **If no**, how were changes managed?

### Data Entry System design

As far as possible the data entry screens should match the questionnaire visually. Where possible the skips in the questionnaire (e.g. If No, go to Q23) should be programmed into the data entry screens. Checks for completeness should be included to ensure data entry staff do not miss questions (or pages). Where possible range checks should be included. A data entry manual should be prepared to accompany the data entry system.

31 Which software package was used to create the data entry system?

32 What were the reasons for choosing this particular software?

33 Were there any missing value codes used in the survey data? Yes  No

33.1 **If yes**, where are these codes documented?

### Staff Training

Fieldwork supervisors and enumerators will generally be trained in the use of survey procedures for conducting interviews. Their training should also include sessions on completing the questionnaires and checking the accuracy of the data.

Data entry staff will also need to be trained in the use of the data entry system.

34 Did the field supervisors' training include sessions on completing checklists to ensure the questionnaires are complete, consistent and accurate? Yes  No

35 Were the data entry operators trained specifically for the work in this survey? Yes  No

35.1 **If No**, why not?

**If yes**

35.2 How long was this training?

35.3 Who ran the training sessions?

- 35.4 Did the data entry supervisors receive additional training over and above that which was given to the data entry operators? Yes  No   
**If yes**, how long was this extra training?

### Transfer and Logging of Questionnaires to the Central Office

After fieldwork the completed questionnaires – whether in paper or electronic format – will need to be transferred to a central repository. This is the stage where the questionnaire archivist takes over responsibility for the completed questionnaires.

- 36 Was there a well defined process in the central office for receiving the completed questionnaires? Yes  No   
 36.2 **If yes**, is this process documented? Yes  No

### Data Entry

Data entry is often considered an unskilled task, but, as with data collection, lack of care in data entry can have serious consequences for the survey.

Traditionally data entry was done after fieldwork in centralised offices. Some survey practitioners now advocate entering the data while the teams are still in the field. There are advantages and disadvantages to each approach. In this check list we consider three possibilities for data entry:

1. In a centralised survey office by a team of data entry staff
2. In the field by one or two data entry staff working as part of the field team
3. Directly onto the computer by the fieldworker during the interview

- 37 How was data entry carried out for your survey? (more than one answer is possible)
- In a central location after fieldwork
- During fieldwork by data entry staff working as part of the field team
- Directly onto the computer by the fieldworker during the interview
- Other, specify

*(Please go to the appropriate section(s) according to the answer to the previous question)*



*In a central location after fieldwork*

The advantages of centralised data entry are:

- Larger pool of trained data entry staff available
- Problems with the data entry system can easily be addressed
- Better security for data, questionnaires and equipment
- Data entry is more easily supervised

The disadvantages are:

- There might be a substantial time gap between data collection and data entry
- It is difficult to return to the field to check inconsistencies

38 How was the data entry organised

- Central server with a single copy of the database
- Individual computers with multiple copies of the database

*(now skip to the section on Backups)*

**1. During fieldwork by data entry staff working as part of the field team**

The advantages of entering the data while still in the field are:

- It is easier to revisit the household if anything is not clear or consistent on the questionnaire
- The interview is still fresh in the mind of the fieldworkers who are in the local team

The disadvantages of this system are:

- Supervision of data entry is difficult
- Problems with the data entry system cannot be resolved easily
- A higher level of skill is required of the data entry staff (at least supervisor level)
- Storage of backups and databases could be a problem
- There are issues of safety of field personnel
- Systems for ensuring an adequate supply of power are needed
- Systems are needed for safeguarding the data in case of system, software or equipment failure

39 How many data entry staff were within each field team?

40 Within each field team who was responsible for data quality assurance?

41 What mechanisms were used to cascade improvements or changes to the data entry system once the fieldwork had started?

42 How did you provide technical backup to the field teams on IT related problems?

*(now skip to the section on Backups)*

## 2. *Directly onto the computer by the fieldworker during the interview*

The advantage of this method of data entry is that there is no delay between data collection and entry – data collection is synonymous with data entry. However, there are many disadvantages:

- Fieldworkers need to be highly skilled both in interviewing and in the use of the data entry system – dependency on the skills of the field staff is increased
- Fieldworkers need to carry equipment for data entry with them in addition to their usual set of materials
- There are issues of safety of field personnel
- Systems for ensuring an adequate supply of power are needed
- Systems are needed for safeguarding the data in case of system, software or equipment failure
- There is no opportunity for double data entry – total reliance on automatic and post-interview consistency checks.

43 How many days training did the fieldworkers have in the use of the data entry system?

44 Were the fieldworkers supervised for some of the interviews?

Yes  No

44.1 **If yes**, provide details of the supervision method used

45 What equipment (laptop, palm pilot, etc.) was used for the data entry?

46 What contingency plans did you have in place in case of hardware or software failure?

47 How did you ensure a constant supply of power?

48 What plans were put in place to ensure the safety of the field personnel?

49 Describe the procedure used for transmitting the databases to the central survey office.

50 Did the report from each field team include a section describing the data entry and how well this worked including any problems encountered?

51 What mechanisms were used to cascade improvement or changes to the data entry system once the fieldwork had started?

52 How did you provide technical backup to the field teams on IT related problems?

### Backups

Backups are an essential part of data management and regardless of the method of data entry you must have a system for taking regular backups of the data and key documentation.

53 Describe your backup strategy

### Double data entry

Double data entry is a process in which the same data are independently entered into two separate copies of the database. The two databases are then compared and any discrepancies checked against the original questionnaires.

54 Did you use double-data entry? Yes  No

### Data Quality Assurance

55 Has an assessment of the overall data quality been prepared for this survey? Yes  No

55.1 If yes, is this document stored with the survey archive? Yes  No

### Anonymisation

In household surveys the respondents often provide the researcher with confidential information. It is important to ensure that individuals cannot be identified from any of the data that is put into the public domain. This obviously includes removing names and addresses, but you must also consider other data by which individuals could be uniquely identified, e.g. the only Marathi speaking family within a community of Hindi speakers.

56 Who is in charge of ensuring anonymisation of datasets?

# Chapter 8: Tabulations and Data Analysis

## Introduction

Ensuring high quality in the production of analysis outputs is a crucial component of survey operations, given that such outputs are aimed at addressing the objectives of the survey. The aim of this chapter of the SQAF is therefore to provide data managers with a checklist which draws attention to different facets that contribute to good quality analytical results. It is hoped that this will enable survey managers to ensure quality assurance of results produced by the data analysts before such results are formally published.

In the checklist below, the main emphasis is on tabulations since they form a key component in the analysis of all surveys. It is well recognised that the quality of survey results can be much enhanced by having well-structured and well-presented tables that summarise the data effectively, and give the user an assessment of the magnitude of population characteristics disaggregated by (a) geographical units such as a country's regions or other administrative divisions; (b) urban and rural populations; and/or (c) gender. Summarising survey results as tables generally forms a first stage analysis for surveys with a research orientation, but could also form the full analysis output for surveys primarily aimed at learning about the distribution of key population characteristics or special indicators such as those measuring poverty.

Where the survey objectives are much wider, further analytical outputs, involving for example modelling techniques or multivariate procedures, may be applicable. Some additional checklist questions are included towards the latter part of this chapter to provide managers with some quality assurance guidelines about these more advanced analytical results. In the main however, the objective of this chapter is to provide a quality assurance framework for survey results, largely in terms of the quality and relevance of tabulations, but also covering other general issues such as personnel and software, production of a data analysis plan, checking quality of data received from the data processing team together with associated documentation, and user satisfaction.

## Personnel and software

### The Data Analysts

Survey data analysis is generally handled by one or more persons who have overall responsibility for the delivery of survey results or have been assigned to undertake data analysis tasks for the person who has this responsibility. In either case, they must have a good knowledge of the survey objectives, familiarity with the software being used for the analysis, and experience in survey data analysis techniques. The checklist below is aimed at ensuring these quality requirements are met.

### Staff and their experience

- 1 Will the data analysis (fully or partly) be done or has it been done by any persons in the organisation responsible for the survey?

**If Yes**, what are their names, title, level of involvement in the survey design, and experience in undertaking survey analytical work?

<i>Name</i>	<i>Job Title</i>	<i>Involvement in the survey design (none, low, medium, high)</i>	<i>Experience (low, medium, high)</i>

If external data analysts are involved, give their name, institution and contact details

<i>Name</i>	<i>Institution</i>	<i>Contact details</i>

### Software for analysis

- 2 What software will be used (or has been used) by the data analysts?

<i>Name of software:</i>	<i>Version Number:</i>

- 3 What level of competence did the analysts have in using above named software prior to undertaking the data analysis?

<i>Name</i>	<i>Software</i>	<i>Competence (low, medium, high)</i>

## Data Analysis Plan

It is inadvisable for data analysis to commence without the production of a Data Analysis Plan because such a plan is part of good statistical practice. The lack of a plan for analysis can often produce results that are not relevant to the survey objectives, or more importantly, omit tables that are needed.

The primary advantage of preparing a Data Analysis Plan at the survey planning stage is that it will

- (a) serve to highlight which data variables are really needed to achieve the survey objectives, e.g. the relevance of all information in a questionnaire survey can be ascertained by checking whether they enter somewhere in the analysis plan;
- (b) set a common standard for all data analysts to follow in terms of table layout, headings, font sizes, spacing, etc., and
- (c) outline the list of tables to be produced and analysis techniques to be followed so that the analysts have a clear protocol to follow, thus enabling the analysis to be completely quickly and efficiently.

The checklist below aims to highlight a number of features related to the Data Analysis Plan so that survey managers can be confident that quality assurance criteria in analytical results are met.

### Documentation, review and archiving

- 4 Was a Data Analysis Plan discussed and documented by the survey team at the time of designing the survey? Yes  No

4.1 **If No**, indicate reasons or constraints why this did not happen

--

4.2 **If Yes**, name the person(s) responsible for the final written version of the Data Analysis Plan and his/her role in the survey team

Name:	Role in team:

- 5 Were all data analysts consulted or given the opportunity to contribute to the Analysis Plan? Yes  No

6 Name those who contributed to discussions and/or documentation of the Analysis Plan and whether they are/were internal or external to the institution responsible for the survey

Name	Internal or External

--	--

- 7 Has the Data Analysis Plan been reviewed by someone who has not been involved in the discussion of its production or its write-up? Yes  No
- If yes,** name the person and his/her role in the survey team or his/her designation/title

Name	Role in team/Designation/Title

**If no to question 7,** how confident are you that the Analysis Plan is appropriate for achieving the survey objectives? Skip to question 9.

Fully   
Partially   
Not at all

- 8 Following the reviewer’s feedback on the Data Analysis Plan, was the Plan revised? Yes  No
- If yes,** name the person who did this, and his/her role in the survey team

Name:	Role in team:

**If no to question 8,** give reasons why a revision was not done

- 9 Was the final version of the Data Analysis Plan submitted to the Data Manager for inclusion in the Survey Archive? Yes  No
- If yes,** name the person who made the submission and the name of the person receiving the document

Name of person submitting:	Name of person receiving:

**If no to question 9,** give reasons why not, or the time at which this will be done

**Contents of the Analysis Plan**

- 10 Which of the following components were included in the Analysis Plan
- 10.1 A list of the analysis objectives in accordance with survey objectives Yes  No
- 10.2 An outline of procedures for an exploratory data analysis Yes  No



10.3	Possible actions to be taken if outliers <sup>12</sup> are identified	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.4	A listing or a dummy version of tables that are needed for addressing the survey objectives	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.5	For all tables, a clear statement of which are column variables (and row variables for 2-way tables), together with the summary statistic (count, row/column percentage, mean, etc) for key variable(s) included in table cells	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.6	An example showing the layout structure and style for tables (both 1-way and 2-way tables)	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.7	A description of the format to be used for tables, e.g. font sizes, spacing, etc.	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.8	Name of the software to be used for the data analysis, and its version number	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.9	An outline of the purpose and statistical tests to be applied to table results, e.g. chi-squared analysis, where appropriate	Yes <input type="checkbox"/>	No <input type="checkbox"/>
10.10	An outline of other statistical analysis procedures that will be undertaken and their purpose, together with a list of relevant references	Yes <input type="checkbox"/>	No <input type="checkbox"/>

## Database issues

The Data Processing Team (*sometimes referred to as the Data Management Team*) would generally be expected to supply a cleaned up/validated database to the Data Analyst, as soon as it is ready, with all the information collected during the survey. However, it would be appropriate to confirm the quality and completeness of this information prior to commencing the data analysis. If the survey manager is not involved with data analysis aspects, then this section is best completed by the person responsible for delivery of analytical results. The survey manager can be assured of good progress if all questions are answered with a “Yes”. If this is not the case, reasons for “No” answers need to be determined through discussions with the Data Manager.

## Completeness of information

11	Are the following variables available in the database?		
11.1	Name or ID code for fieldwork supervisors	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.2	Name or ID code for Enumerators/Interviewers/Data Collectors	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.3	Name or ID code for Data Entry Operators	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.4	Dates when data collection took place	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.5	Administrative/Geographic units relevant to sampling design	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.6	Unique ID for ultimate sampling units, e.g. households	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.7	Sampling weights <sup>13</sup> for each sampling unit	Yes <input type="checkbox"/>	No <input type="checkbox"/>
11.8	Variables corresponding to each item in the survey questionnaire or	Yes <input type="checkbox"/>	No <input type="checkbox"/>

<sup>12</sup> An observation is regarded as an outlier if it does not conform to the pattern of distribution exhibited by similar observations

<sup>13</sup> Weighting the data by the sampling weights is necessary to generalise the results from the survey sample to the target population. The principle underlying the weighting process is that of taking each record and expanding it to represent the number of units of the population from which it has been drawn.

other recording instrument

- 12 Does the total number of records in the database correspond to the number of questionnaires/sampling units covered in the survey? Yes  No

12.1 **If No**, indicate reasons for the difference

- 13 Does the total number of records in the database correspond to the number of planned sample size as stated in Chapter 4, question 15? Yes  No

13.1 **If No**, indicate reasons for the difference

- 14 If there was a **No** response to any of questions 11, 12 or 13, have reasons for this been ascertained satisfactorily with the Data Manager? Yes  No

### Documentation

- 15 Has the Data Processing Team provided the Data Analysts with a Data Dictionary that describes each variable in the database, along with an explanation of codes used for categorical variables? Yes  No

15.1 **If No**, would action be taken to ensure that the Data Dictionary is prepared and submitted to the Data Analysts? Yes  No

- 16 Has the Data Processing Team provided the Data Analysts with a Data Validation Report? Yes  No

16.1 **If No**, does the Data Processing team expect to prepare a Data Validation Report? *(This is highly recommended since it provides information regarding types of quality checks done on the data).* Yes  No

## Table Results and Reporting

This section is a crucial component of quality assurance since tabulations produced during data analysis must be in accordance with planned procedures outlined in the Data Analysis Plan. Table results and their reporting must provide the user of the survey results with clear messages concerning the key findings. Of major importance is to ensure that sample sizes are clearly stated, deviations from expected sample sizes are explained, that editorial standards are met – with tabulations well presented using a standard format for titles, legends, etc., and that emerging results are reported, where appropriate, with reference to their statistical significance. The checklist below covers these issues and a few others.

### Sample size

- 17 Can the total number of records used for each table be determined from the table itself (through its heading, as a footnote or as part of the table)? Yes  No

**If No**,

17.1 Will the data analysis reporting team be asked to provide this information? Yes  No

- 17.2 Are there valid reasons for omitting reference to the sample size? Yes  No
- 18 In tables which are based on the full data set rather than a subset, does the number of records used for the analysis coincide with the number of case records computerised? Yes  No
- If No,**
- 18.1 Does the reporting show the numbers missing? Yes  No
- 18.2 Are there explanations in the tabulated results, or in the reporting of the results, for the different types of missing values? Yes  No
- 18.3 Will you be seeking such explanations and requesting that they be incorporated in the reporting of results? Yes  No
- 19 Has the non-response rate<sup>14</sup> for the survey been highlighted in the results report? Yes  No
- 20 How large is the non-response rate (as a %) over the whole survey?

### Layout and formatting of tables

- 21 Are the table layouts consistent with respect to variables used commonly for disaggregating the data (e.g. variables such as gender, geographical regions, urban/rural, etc.)? Yes  No
- 22 Are the table layouts consistent with respect to their style and format? Yes  No
- 23 Are the number of significant digits consistent across all tables (e.g. 3 significant digits would mean having figures such as 149, 33.2, or 4.27)? Yes  No

### Analytical results

- 24 Have all tables suggested in the Data Analysis Plan been produced? Yes  No
- If No,**
- 24.1 Has the Data Analyst leader provided reasons for this reduction? Yes  No
- 24.2 Will this have a serious impact on delivery of outputs to meet the survey objectives? Yes  No
- 25 Is there consistency in results across different tables (e.g. in terms of sample sizes, numbers in disaggregated subsets, etc)? Yes  No
- 26 Can the results of each table be interpreted without reference to the reporting text? Yes  No
- 27 Are there syntax program files which would allow all tabulations to be reproduced in the event that errors in the data are later discovered and corrected, hence requiring a re-run? Yes  No
- 28 Is there a system for someone to check the accuracy and validity of computational processes leading to data tabulations? Yes  No
- 28.1 **If yes,** name the statistician/programmer who was involved with this checking procedure and his/her role in the survey team?

Name	Role in team

<sup>14</sup> Non-response rate = (number of responding units)/(total number targeted by the sampling design)\*100

## Technical Issues

- 29 Is there reference in the reporting to indicate that sampling weights have been used in data tabulations? Yes  No
- If No,**
- 29.1 Does the reporting make it clear that the results only apply to the surveyed sample rather than the target population? Yes  No
- 29.2 Are there valid reasons why sampling weights have not been used in the analysis? Yes  No
- 30 Where statistical tests of significance have been carried out, have the exact p-values for significance been stated (either in the reporting, or in the table) rather than depicting these with 1-3 stars(\*)? Yes  No
- 31 If outliers have been reported, is it clear how they were handled during the analysis? Yes  No
- 32 Have data imputations been done prior to data analysis? Yes  No
- If Yes**
- 32.1 Has the method used been documented in the reporting? Yes  No
- 32.2 Briefly outline the method of imputation used

## Further Analytical Results and Reporting

Although it is often the case that the bulk of the analysis of survey data involves data tabulations, surveys with more special requirements can lead to standard or more advanced statistical analysis procedures. In such cases, it would be important to ensure that the methodological procedures have been documented in the Data Analysis Plan and implemented during the actual analysis of the data. Other issues can also arise such as the way in which outliers are dealt with, and actions to take in dealing with missing values.

The checklist below covers some of the more common issues that arise during analyses that proceed beyond mere tabulations or include more advanced forms of statistical analyses. It must be recognised however that this checklist does not give assurance that the correct methodology has been applied, or that the interpretation of results is correct. For these issues, it would be necessary to get an independent review of the statistical results from a suitably qualified statistician.

### Further Analysis

- 33 Have further data analysis procedures been done in accordance with the planned statistical analysis approaches outlined in the Data Analysis Plan? Yes  No
- 33.1 **If No,** has the Data Analyst leader provided reasons for the changes? Yes  No
- 34 Has the methodology used for the analysis been clearly documented in a way that allows the analysis to be repeated? Yes  No
- 35 Has the purpose of methodologies undertaken in further analyses been written in a way that can be understood and appreciated by a non-statistician? Yes  No
- 36 Are syntax program files available that will enable the analysis to be re-produced and checked, where necessary, by an independent reviewer? Yes  No

### Technical Issues

- 37 Is there reference in the reporting to indicate that sampling weights have been used in further analyses? Yes  No
- 38 Are measures of precision (standard errors) or confidence intervals available for all key estimates that have been reported? Yes  No
- 39 Can the statistical software used for the analysis take account of sampling weights and provide correct standard errors for survey estimates? Yes  No
- 40 Is the coefficient of variation (CV) for key survey estimates of acceptable magnitude? Yes  No
- 41 If the survey report includes CVs (rather than standard errors or confidence intervals), state the CVs, expressed as a percentage, for those estimates regarded as most important for achieving the survey objectives

Estimate	Coefficient of variation

### User Satisfaction

Quality of analysis results is of course best judged by the users themselves, so it would be appropriate to give some consideration to users' needs and the evaluation of their level of satisfaction with tabulated results. This final section includes just a few checklist questions to ensure that users' needs are met.

#### General

- 42 Do you have a procedure in place to evaluate users' satisfaction with the production and quality of survey results? Yes  No
- 43 Do you have flexibility within your data analysis team to respond to user demands for additional analyses? Yes  No
- 44 Within how many days of such requests would you expect to agree to a time scale for delivery of additional analyses results, recognising the need for time to discuss the exact requirements?
- 45 Did the main users of the survey have an input towards the development of the analysis plan? Yes  No

#### Timeliness

- 46 What is the time lag between the last date of data collection and completion of data analysis results?
- | Month | Days |
|-------|------|
|       |      |

46.1 What is your perception of user satisfaction with the above time lag?

Very satisfied: <input type="checkbox"/>	Satisfied: <input type="checkbox"/>	Unsatisfied: <input type="checkbox"/>	Very unsatisfied: <input type="checkbox"/>
--	-------------------------------------	---------------------------------------	--

- 47 Do you have a procedure to monitor progress during the data analysis phase so as to ensure that analytical results are completed in the required time?      Yes       No

## Chapter 9: Dissemination of Survey Products

Chapter 2 highlighted the need to identify stakeholders and user groups to whom the survey products would be disseminated, and also highlighted the need to set up and agree on policies and procedures for dissemination. However, on completion of the production of survey outputs, it is well to re-visit these decisions and consider whether changes are required. Hence this chapter allows the updated list of user groups and methods for dissemination to be stated and considers the timeliness of dissemination activities.

### Changes to planned dissemination policies/procedures

- 1 Were there changes to previously planned dissemination policies? Yes  No   
**If No**, have the agreed policies and procedures been documented and submitted to the survey archive? Yes  No
- 2 **If Yes**, to question 1, did changes occur with respect to any of the following?
- 2.1 Which survey products to disseminate? Yes  No
- 2.2 User groups who would receive survey publications Yes  No
- 2.3 Procedures for dissemination, e.g. media to use Yes  No

### Survey publications and media used for release

- 3 Provide the full list of survey publications, together with dates (intended and actual) of dissemination. Append extra pages if necessary.

Publication name	Intended date	Actual date	Time lag acceptable?
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>

- 4 What are the media (hard copies, emails, web, etc) through which the above publications will be disseminated?

Publication name	Media for dissemination	Date of release

### User groups

- 5 Were users of survey products identified at the survey inception stage? Yes  No

**If No**, please explain when this was done and how the users were identified:

- 6 Name the stakeholders and user groups, e.g. Government ministries, research institutions, etc., and the contact person for receiving survey publications. Append extra pages if necessary.

<i>Ministry/Institution</i>	<i>Name of contact</i>	<i>Job title of contact</i>



## User satisfaction

- |   |   |     |                          |    |                          |
|---|---|-----|--------------------------|----|--------------------------|
| 7 | Was there a system in place to evaluate users' satisfaction with the survey reports they received?                | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 8 | Was there a system in place to provide additional information, e.g. data, field manual, etc) on request to users? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 9 | Were users clear about whom they should contact if they had questions relating to survey reports?                 | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

# Chapter 10: Archiving - documentation and dissemination of microdata

## Introduction

Archiving is generally considered to be the domain of the data manager, but it is important that the survey manager and other members of the team understand that they too have responsibilities in this area.

Many consider “Archiving” to be synonymous with “Data Archiving”. However, data by themselves are of little use – a whole range of documentation and meta-data must be included in the archive (thus we have omitted the term “Data” from the chapter heading). The data manager may be in charge of the archive but others in the team will need to provide documentation and reports to do with the survey design, sampling, data analysis, etc. These have been mentioned in earlier chapters but brought together here for clarity and consistency.

Two levels of archiving may be considered. The first is the internal survey archive which is only available to the survey team. This should include all the (raw) data, all documentation (procedures, reports, minutes of meetings, etc.) generated during the course of the study, copies of syntax files for data checking and analysis, analysis results, etc.

The other level of archiving is to submit the data to a public or national data archive. This would include anonymised datasets, questionnaires, sampling documents, etc. The archive must include enough documentation to enable secondary users of the data to fully understand the study and the data.

This chapter is intended to make survey managers and data managers aware of the range of documentation they should expect to include in the survey archive. Remember all information related to the survey may be useful and should be archived (even if not all will be made available in the public domain).

- 1 Who was responsible for collating documentation and data files for the survey archive?

*Name*

*Title*

*Contact details*

- 2 Has an internal survey archive been produced? Yes  No

- 2.1 **If yes**, who has copies of the archive CD/DVD?

- 3 Has the archive been published on a website where you have control? Yes  No

- 3.1 **If yes**, what is the URL of the website?

- 4 Has the data been submitted to a public archive, national or regional data holder? Yes  No

4.1 If yes, how can secondary users access the data and documentation?

## Documentation in Support of Survey Outputs

Previous chapters have highlighted the types of documentation that should be available to provide the necessary level of quality assurance with respect to processes that take place from the planning of a survey through to data tabulation and production of analytical results. Here we bring together the key documents that relate directly to the survey products being disseminated so that the evidence base for the quality of these products is clear.

### Objectives

- 5 Have the following documents been submitted to the data manager for archiving purposes?
- 5.1 Document describing survey background, justification and objectives Yes  No
- 5.2 Definitions used in the survey (standard and survey specific) Yes  No

### Fieldwork

- 6 Have the following documents relating to field activities been submitted to the data manager for archiving purposes?
- 6.1 Field data collection instruments (e.g. questionnaires)? Yes  No
- 6.2 Field instruction manual(s) Yes  No
- 6.3 Checklist(s) used during fieldwork to confirm quality assurance Yes  No
- 6.4 Field implementation report(s) including difficulties encountered, resolutions, and recommendations Yes  No

### Sampling

- 7 Have the following documents relating to the sampling process been submitted to the data manager for archiving purposes?
- 7.1 Sampling frame(s) used during field operations Yes  No
- 7.2 Document describing the sampling methodology in detail Yes  No

### Data Management

- 8 Have the following items relating to data collected and its management been included in the survey archive?
- 8.1 All survey data sets accompanied by their data dictionaries<sup>15</sup> Yes  No

<sup>15</sup> By data dictionary we mean a document containing descriptions of all variables in the datasets including name, label, codes and missing values.

- 8.2 A Data Quality Assessment Report Yes  No
- 8.3 A Data Management Report highlighting key features of all computer related tasks ranging from data entry to the production of the archive Yes  No

**Analysis**

- 9 Have the following documents and files relating to the data analysis been submitted to the data manager for archiving purposes?
  - 9.1 The Data Analysis Plan Yes  No
  - 9.2 All analysis program files, appropriately annotated with comments to clarify the purpose of each analytical component undertaken Yes  No
  - 9.3 The Data Analysis Methodology report Yes  No

**Survey Reports/Publications for Dissemination**

The main survey publication will be the report that describes the methodological approaches and the survey results. However, it is likely that there will be other reports that would be disseminated either in response to a request from stakeholders, or because they would be helpful to researchers intending to make further use of the data. Which documents would be disseminated in this way is a decision that will be made by the core survey team with due regard to confidentiality issues. We are assuming that these documents would also be included in the survey archive but here we are assuming dissemination of individual documents prior to the final submission of the survey archive at the end of the study.

The checklist below covers a range of documents that would be expected to form the core of survey products available for dissemination. A further checklist includes the minimum requirements concerning the contents of each of these reports to ensure that the corresponding survey product will be valued by the user.

**Reports for dissemination**

- 10 Were the following survey products available for dissemination prior to the end of the study?
  - 10.1 Sampling Methodology report Yes  No
  - 10.2 What was the time lag between its publication and the date of completion of the fieldwork?
  - 10.3 Data Management report Yes  No
  - 10.4 What was the time lag between its publication and the date of completion of the fieldwork?
  - 10.5 Data Analysis Methodology report Yes  No
  - 10.6 What was the time lag between its publication and the date of completion of the fieldwork?
  
- 10.7 The main survey report, including an overview of methodological aspects, the key survey results and discussion of the findings Yes  No
- 10.8 What was the time lag between its publication and the date of completion of the fieldwork?
- 11 Were the above time lags acceptable to the funding body or organisation sponsoring the survey work? Yes  No

## Contents of Dissemination Reports

- 12 Did the Sampling Methodology report include the following key components?
- |      |  |     |                          |    |                          |
|------|--|-----|--------------------------|----|--------------------------|
| 12.1 | Source and/or details of the sampling frame  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 12.2 | The sampling process, with listings or descriptions of strata, clusters and other units  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 12.3 | Sample sizes used in each stratum and at all hierarchical stages of the sampling process                                       | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 12.4 | The final questionnaire used during field implementation   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 12.5 | Details concerning sampling weights used during data analysis work or reference to a document where these details may be found | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
- 13 Did the Data Management report include the following key components, or a reference to where these components could be found?
- |      |  |     |                          |    |                          |
|------|--|-----|--------------------------|----|--------------------------|
| 13.1 | Software used for data entry and for data management, and reasons for choosing these particular software | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13.2 | A list of survey related documents that would be available in the data archive                           | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13.3 | A list of data sets available in the data archive, with a brief description of the contents of each      | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13.4 | An assessment of the quality of data in the archive  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13.5 | An audit log of problems encountered during data management, and steps taken to resolve them             | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 13.6 | A list of all syntax/program files used during data preparation and analysis                             | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
- 14 Did the Data Analysis Methodology report include the following key components?
- |      |   |     |                          |    |                          |
|------|---|-----|--------------------------|----|--------------------------|
| 14.1 | Personnel contributing significantly to the production of analysis results  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.2 | Description of statistical methodologies used, including a list of references   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.3 | Software used for data analysis and reasons for using these software  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.4 | Key survey results and a discussion of these  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.5 | The use of sampling weights   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.6 | Reasons for any limitations in survey results, e.g. due to the sampling process, data collection methodologies, statistical methods used, etc | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 14.7 | An assessment of the quality of the analytical results  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
- 15 Did the main survey report include the following key components?
- |      |  |     |                          |    |                          |
|------|--|-----|--------------------------|----|--------------------------|
| 15.1 | A justification for why the survey was conceived and carried out | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 15.2 | An overview of the sampling methodology                          | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 15.3 | An overview of the data collection process                       | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

- |      |  |                              |                             |
|------|--|------------------------------|-----------------------------|
| 15.4 | Procedures followed to ensure confidentiality of survey respondents                  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 15.5 | A discussion of non-sampling errors  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 15.6 | Presentation of sampling errors for key survey results, e.g. using confidence limits | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 15.7 | An assessment of the quality of the survey as a whole                                | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

## Storage of Meta-Data

Meta-data comprises the information without which the survey data are rendered next to useless. It is recommended that you make use of the Microdata Management Toolkit provided by the International Household Survey Network (IHSN) to document your study in accordance with the Data Documentation Initiative (DDI) and the Dublin Core (DCMI) metadata standards<sup>16</sup>. The toolkit generates a metadata file in XML (Extensible Markup Language) which consists of the following sections:

<b>Documentation Description</b>	A description of the XML file including study title and production date.
<b>Study Description</b>	A description of the study itself including investigators, dates, methods, etc.
<b>File Description</b>	A detailed description of each data file including contents, processing checks, etc.
<b>Variable Description</b>	Detailed information on each variable including names, labels, category codes, etc.
<b>External Resources Description</b>	Materials related to the study other than data files e.g. questionnaires, fieldworker manuals, data entry system, etc.

The *Quick Reference Guide for Data Archivists* published by the IHSN gives details for each section.

- |      |   |                              |                             |
|------|---|------------------------------|-----------------------------|
| 16   | Did you use the guidelines provided by the Data Documentation Initiative (DDI)?                                     | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 17   | <b>If no, why not?</b>  |                              |                             |
| 17.1 | Did you use the Microdata Management Toolkit produced by the IHSN to help with the management of data and metadata? | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 17.2 | Were all electronic documents converted to PDF format?  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 17.3 | Have key variables (unique identifiers) been identified and documented for each data file?                          | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 18   | Have relationships between data files been validated?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 19   | Were frequency tables produced for all categorical variables?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 20   | Were values outside the expected range for categorical variables  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

<sup>16</sup> DDI (Data Documentation Initiative) and DCMI (Dublin Core Metadata Initiative) are international XML metadata specifications. For more information on these standards and on the IHSN Toolkit please visit [www.surveynetwork.org](http://www.surveynetwork.org)

- investigated?
- 21 Were descriptive statistics (min, max, mean, etc.) produced for all continuous variables? Yes  No
- 22 Were values outside the expected range for continuous variables investigated? Yes  No
- 23 Were missing value codes assigned and documented for all numeric and categorical variables? Yes  No
- 24 Have all files (data files and document files) been scanned with up to date virus detection software? Yes  No
- 25 Have you ensured that all the micro data has been thoroughly anonymised? Yes  No

# Appendix - a data manager’s checklist

## Introduction

Chapter 7: Data Management was aimed at helping the survey manager identify some aspects of a good data management system. There are more in-depth questions that the data manager needs to consider and this chapter is an appendix has been put together as an additional checklist for data managers.

## Using Unique IDs to Identify Staff

If problems occur in the data is it often useful to be able to identify the chain of “data handlers” – that is the individuals who have collected the data, checked it, and computerised it. A **Data Handlers** table at the front of the questionnaire provides space for these individuals to sign and date the questionnaire. In order to computerise this information, staff should be assigned unique IDs. These IDs are easier to enter making it easier to determine who has handled the data.

- |   |  |     |                          |    |                          |
|---|--|-----|--------------------------|----|--------------------------|
| 1 | Were the fieldworkers each given a unique ID that was used on the questionnaires and in the database?          | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 2 | Were the fieldwork supervisors each given a unique ID that was used on the questionnaires and in the database? | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 3 | Were the data entry staff each given a unique ID that was used on the questionnaires and in the database?      | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| 4 | Were all the IDs and names of the data handlers included in the database before the start of data entry?       | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

## Database and Data Entry Manuals

The Database Manual should describe the database in detail. This is particularly important if the database was developed by an external consultant

The Data Entry Manual describes the data entry process and should be used for training and as a reference for the data entry staff.

- |     |   |     |                          |    |                          |
|-----|---|-----|--------------------------|----|--------------------------|
| 5   | Was a database manual produced?   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
|     | <b>If yes</b>   |     |                          |    |                          |
| 5.1 | Are the following included in the database manual:  |     |                          |    |                          |
| a.  | A list of tables and their corresponding primary keys <sup>17</sup> ?   | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| b.  | A detailed explanation about how each primary key was defined?  | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |
| c.  | A description of any other objects in the database such as forms, queries, reports, or modules, in particular any added functionality | Yes | <input type="checkbox"/> | No | <input type="checkbox"/> |

<sup>17</sup> A primary key is a unique identifier comprising one or more fields within the table. The primary key cannot be duplicated and cannot be blank.



- such as appending data, editing data, etc.?
- d. A diagram of the relationships built into the database? Yes  No
- 6 Was a separate data entry manual produced? Yes  No
- If yes**
- 6.1 Was the data entry manual used in the training of the data entry staff? Yes  No
- 6.2 Was each data entry operator given his/her own copy of the manual? Yes  No

## Data Entry System

- 7 To what extent did the data entry screens match the questionnaire visually? Completely   
Partially   
Not at all
- 8 Did the data entry system include automatic skips matching those in the questionnaire? Yes  No
- 9 Were there checks in the data entry system to ensure completeness – i.e. checks to make sure a value was entered for each question<sup>18</sup>? Yes  No
- 10 Were range checks (valid values) included where appropriate? Yes  No
- 10.1 **If yes**, where are these range checks documented?
- 
- 11 Did anyone test the data entry system by entering several of the completed questionnaires? Yes  No
- If yes**,
- 11.1 Who carried out this test?
- 
- 11.2 How many questionnaires did he/she enter?
- 11.3 Was the testing process and its results documented and acted upon? Yes  No

## Staff Training

One of the roles of the data manager is to ensure that data entry is completed as efficiently and accurately as possible. This usually involves overseeing the training of the data entry staff although the actual training sessions may be run by another member of the data management team.

- 12 Were the data entry operators trained specifically for the work in this survey? Yes  No

<sup>18</sup> Non-responses are a common occurrence in surveys, therefore codes for non-responses need to be included in the data entry program.

12.1 **If No**, why not?

**If yes**

12.2 How long was this training?

12.3 Who ran the training sessions?

12.4 Did the data entry supervisors receive additional training over and above that which was given to the data entry operators?

Yes  No

**If yes**, how long was this extra training?

## Transfer and Logging of Questionnaires to the Central Office

As well as having overall responsibility for the electronic data, the data manager has overall responsibility for the management of the completed questionnaires. After fieldwork the completed questionnaires will need to be transferred to a central repository. This is the stage where the questionnaire archivist takes over responsibility for the completed questionnaires.

- |      |  |                              |                             |
|------|--|------------------------------|-----------------------------|
| 13   | Was there a well defined process in the central office for receiving the completed questionnaires?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
|      | <b>If yes</b> , did this process include:  |                              |                             |
| 13.1 | Logging receipt of completed questionnaires?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 13.2 | A way to match received questionnaires to the expected number determined by the sampling design?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 13.3 | Filing of completed questionnaires?  | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 13.4 | A system for tracking data entry for each questionnaire?   | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| 13.5 | A system keeping track of current locations of questionnaires – e.g. a library type system in which users have to sign questionnaires in and out of the central store? | Yes <input type="checkbox"/> | No <input type="checkbox"/> |

## Merging Databases

In most surveys the data entry is split between several data entry operators. Each operator may be linked to a central server with a single copy of the database, or there may be multiple copies of the database on separate computers. In the latter case there will need to be a system for merging the data from each of the databases.

It is generally easier to carry out any data checks and subsequent corrections prior to merging the databases as there is then a smaller pile of completed questionnaires to look through when data inconsistencies are found.

- 14 How was the data entry organised
- Central server with a single copy of the database
- Individual computers with multiple copies of the database

14.1 If **individual computers**, how were the separate databases merged?

- 14.2 Were the databases merged before or after checks were carried out on the data?
- Before
- After

## Backups

Backups are an essential part of data management and you must have a system for taking regular backups of the data and key documentation.

- 15 How often were backups of the database(s) made?

- 16 Where were the backup media stored?
- 

- 17 Does your backup strategy include backing up electronic copies of survey documentation?      Yes       No

## Double data entry

Double data entry is a process in which the same data are independently entered into two separate copies of the database. The two databases are then compared and any discrepancies checked against the original questionnaires.

- 18 Did you use double-data entry?      Yes       No

**if yes**

18.1 Describe the process for the double data entry comparisons?

- 18.2 Did you compare records (matching by key fields)?      Yes       No

- 18.3 Did you compare individual values in the fields? Yes  No
- 18.4 Were corrections made to both copies of the database? Yes  No
- 18.5 Was a log kept of all discrepancies found by the double data entry comparisons? Yes  No
- 18.6 Is this log stored with the survey documentation? Yes  No

**If no to question 18,**

Describe any procedures you used in lieu of double data entry

## Exporting the Data

The data will most likely need to be exported from the data entry system to a format ready for analysis. Some systems export just the raw data values while others also export variable and value labels and missing value codes.

- 19 What format was used for data exported from the data entry system?

- 20 Did the export process include variable and value labels? Yes  No

- 20.1 **If no**, how were variable and value labels added to the exported data file(s)?

- 21 Did the export process include missing value codes? Yes  No

- 21.1 **If no**, how were missing values codes added to the exported data file(s)?

## Anonymisation

In household surveys the respondents often provide the researcher with confidential information. It is important to ensure that individuals cannot be identified from any of the data that is put into the public domain. This obviously includes removing names and addresses, but you must also consider other data by which individuals could be uniquely identified, e.g. the only Marathi speaking family within a community of Hindi speakers.

22 Have the exported data files been made anonymous? Yes  No

22.1 **If yes**, how was the anonymisation done?

22.2 **If no**, when do you plan to anonymise the data?

23 Have all text values been thoroughly checked to ensure they do not identify individuals? (E.g. in response to the question “Who is the head of household?” “Fred Bloggs” should be removed from the data file but “My father” is acceptable.) Yes  No