

## **Innovative techniques that can be used to reduce response errors in household based surveys**

Pietersen, Dion Dhemandra

*Statistics South Africa, Survey Operations*

*170 Andries Street*

*Pretoria (0037), South Africa*

*E-mail: dionp@statssa.gov.za*

### Introduction

Statistics South Africa (Stats SA) is South Africa's national statistical agency. According to Minister in the Presidency: National Planning, Mr Trevor Manuel, "It provides official statistics through which government can identify issues, plan programmes and address them, and monitor their implementation as well as their desired impact on society." The department conducts both economic and social surveys within the country, some of which are conducted monthly, quarterly and annually. Within the Survey Operations / Population and Social Statistics Surveys cluster the following surveys are conducted: the General Household Survey, Income and Expenditure Survey, Quarterly Labour Force Survey, Domestic Tourism Survey, as well as ad hoc surveys and projects.

The divisions' core area of functionality is in the field; so reduction of non sampling errors becomes an intrinsic element of the division. Due to the numerous surveys and projects that are undertaken within the division, senior management identified strategic projects and surveys for improvement. The paper presented will focus on one of these priority projects – the Labour Force Survey and associated modules.

At Stats SA, the October Household Survey (over the period 1994 – 1999) was South Africa's principal vehicle for collecting South African labour market information. The survey covered an wide range of subject matter from births to deaths, migration, employment and housing to crime and education. The demand for labour market information as well as conformity to the International Labour Organization resulted in the establishment of a Labour Force Survey (LFS) programme within the organization. The first LFS was conducted in 2000. The survey was conducted in March and September every year. Data for the LFS was released 6 months after the collection cycle.

Even though the LFS provided information to stakeholders, there were still some criticisms related to coverage, scope, timeliness and frequency of the survey. The decision to re-design the LFS was well documented in the report written by the International Monetary Fund consultants in June 2005. The re-engineered Quarterly Labour Force Survey was introduced in 2008 addressing all issues outlined in the criticisms.

The paper to be presented will focus on innovative approaches that the division begun implementing to reduce response error, some of which will be related to:

1. Field Operations
2. Questionnaire design and analysis
3. Data processing

## 1. Field Operations

- Field staff

As per the IMF report, a retrospective analysis of previous LFS showed that the usage of contract field staff resulted in unresolved and unaccounted cases of response error. Previously contract field staff were appointed for 3 weeks, after the 3 weeks contracts were terminated and accountable for field activities during the analysis stage of the survey was lacking. A decision in the re-engineering process was thus to appoint a permanent survey staff, these included:

- Provincial Survey Coordinators (x9)
- District Survey Coordinators (x54)
- Provincial Quality Monitors (x27)
- Survey Officers (x233)
- Provincial Admin Officers (x9)

For the field staff a series of systems were developed to enhance the collection process and introduce a more define process of accountability in the field. The systems that were developed do not only serve the field operations team, it is also an integral link to the Data Processing Center. The primary system that has been developed is the Instrument Tracking System. For the field team the following sub – systems have been developed:

- ❖ Workload distribution – assignment planning
- ❖ Field Edit Module (FEM)– editing and imputation rates per Survey Officer
- ❖ Progress reporting – at 3 different levels; national, provincial and district

The assignment planning system was implemented to allow survey management to track the workload distribution and performance of field staff. The assignment number was uniquely designed for each District Survey Coordinator with their supervisory units (Survey Officers). This unique number links all field operations instruments (questionnaires, listings etc.), activities (data collection) and systems (reporting) to the concerned field staff.

The FEM was developed to track Survey Officers performance in terms of error and imputation rates. In preparation for the refresher training sessions the management team from head office review reports generated by the system and tailor design training material for the different regions based on the finding from the FEM – reference is their performance for the past quarter for collection.

With regards to progress reporting, a series of real time reports can be generated at any particular moment during the collection period by management. These reports can be viewed at national, provincial and district level. The real time reporting allows management to make informative decisions on a regular basis. Management at any level is thus able at any point in the collection cycle able to scrutinize performance in the field and give constructive feedback as well.

- Method of data collection

Household members living in approximately 10 000 dwelling units are interviewed in each of the three months within a quarter. Data is collected on a continuous basis following a “0110” approach, where each month:

- During 1<sup>st</sup> week , the “0” represents publicity and maintenance of the Master Sample listings
- The middle two weeks of the month “11” represent the data collection period
- The last “0” represents the wrap up week for the monthly activities

The survey has been designed to incorporate a series of factors that usually contribute towards increasing response errors. Some of the issues that have been looked into are that of reducing respondent burden – the sampled dwelling unit stays in the sample for 4 visits (one interview per quarter). The redesigned questions also contribute greatly toward reducing the levels of respondent fatigue. Response rates in the QLFS from 2008 up to quarter 1 2011 has averaged 92%.

The project team has also looked into issues of ensuring that the correct sampled dwelling units (DUs) are visited. The Master Sample frame has been geo – referenced with coordinates for all sampled DUs. Pilot tests are currently underway where setup interviews (publicity) in a sample of provinces are using GPS devices to identify sampled and complete the appointments for collection digitally.

- Training

It is also widely known fact that survey officers who conduct the interviews can also be a source of response error. Knowing this, the project team undertook to conduct an intensive centralized training session. All permanent field staff underwent these training sessions covering all aspects of data collection from listing and listing maintenance of the Master Sample, publicity (gatekeeper awareness campaigns) as well as training on the content of the core questionnaire. The initial training was conducted in centralized training venues where all field staff for the different provinces were trained together and were evaluated extensively on all aspects of collection. In these sessions, all field staff were trained on asking and answering the questionnaire in a consistent standardized manner. An intensive questionnaire translation guide was also developed for the eleven official languages of the country. The translations were tested extensively through focus group discussions, amongst other observational studies. All attempts were made to ensure that survey officers also understand the meaning behind the questions so as to ensure no misinterpretation of the responses. The field staff are also exposed at a district level to “Refresher training” sessions during the first week of every quarter. Some training material that has been developed:

- Survey Officer manual
- Listing and Listing Maintenance manual
- Quality assurance manual
- Quick reference guides for data collection, listing and setup interviews

## 2. Questionnaire design and analysis

Due to the criticisms that were received from the international community especially with regards to the content, layout and frequency of the questionnaire a complete re-design of the questionnaire was undertaken.

The redesign influenced issues around:

- Questionnaire length,
  - Questionnaire structure,
  - Questionnaire content,
  - Classification of questions,
  - Results, frequencies and timelines
- Questionnaire length
 

The Development and Output team in the Household Labour Market Statistics Division (HLMS) re-designed the previous questionnaire. The previous LFS questionnaire had approximately 140 questions. Some of the questions were related to issues outside of the labour market indicators. The questionnaire was streamlined to have a core content focusing on only 59 questions. The revised questionnaire was tested extensively through observational studies (behind the glass testing, doorstep testing as well as focus group discussions.)
  - Questionnaire structure
 

In the previous questionnaire information for all household members were recorded in 1 single questionnaire. The re-designed questionnaire has a combination of six individual questionnaires in collated into one book. The revised questions were also sequenced in a respondent friendly manner having opening statements about what the survey aims to achieve as well as establishing rules of association for household membership – who qualifies to be interviewed.

The demographic questions are well explained establish the sequence for completing the questionnaire with household members who qualify. This household sequence number is kept for all individuals in the household. A “Household Record Form” (HRF) was developed for the field staff to ensure consistency of household members’ information throughout their active involvement in the survey – 4 quarters of the year. There were some impediments to the form, some of which were related to household / individual non – response, where nobody is at home during the collection period. Other impediments would relate to household member turn-over (migrant workers), survey staff interviewing at the wrong sampled dwelling unit, survey officers incorrectly applying the rules of association (household membership). In an attempt to remedy these impediments, basic information from the initial interview is recorded on the HRF and used as reference when interviewing household members in subsequent interviews.

- Questionnaire content

Previously field staff had to work with a questionnaire where the market and non market questions were all asked together. In the revised questionnaire, the market and non market activities are separate questions. Persons who are involved in market activities are not asked about non market questions. Those persons involved in non market activities are asked questions about unemployment and economic inactivity. The team also created a separate question for employees involved in the informal sector. These revisions greatly reduced the number of errors that the field staff committed especially confusing market and non market related employment activities.

- Classifications

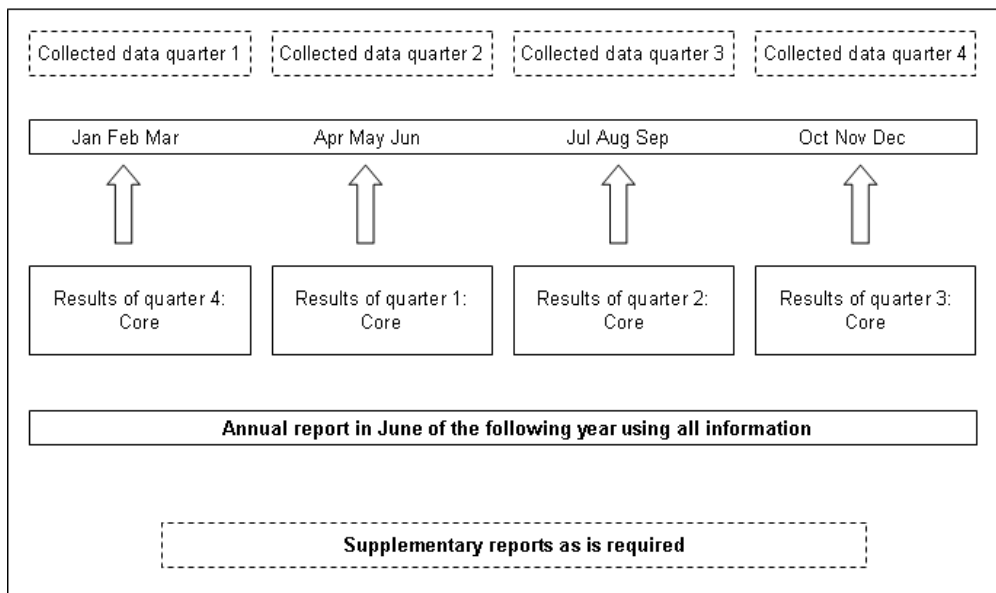
According the QLFS and LFS differences at a glance document, unemployment and discouraged work seekers was much more clearly defined where:

- Unemployment referred to persons who had not worked in market related activities in the calendar week before the interview, who had looked for a job or tried to start a business in the past four calendar weeks, or were available to start a business or a job in the week before the interview.
- Discouraged work seekers was defined as anyone who has not looked for a job but is available for work and cites the following reasons:
  - That there are no jobs in the area
  - There are no jobs requiring his / her skills
  - Or that he / she has lost hope in finding work

- Results – frequency and timeliness

A detailed analysis of the data is done every quarter and the results are released quarterly. The figure below shows that in every given calendar year, data collected in the first quarter (January to March) are analyzed and published by end of April etc.

Figure 1: The analysis and publication of results for the QLFS



### 3. Data processing

Previously different surveys required different scanning and editing programmes to be generated. The questionnaires for different surveys required intensive pre and post editing before the final data was released. The re-engineering also looked at having an end to end processing system that would use an integrated approach for all surveys focusing on the way questionnaires are designed, the systems that should be used for the processing of the questionnaires as well as the timely processing of the questionnaires to meet the project deadlines.

- Questionnaire design

In the previous questionnaire the tick boxes for responses had a blacked border, these lines created noise on the imagery when scanned at lower interpretation rates. To account for these distortions, all tick boxes were designed to have an opaque drop-out.

- Systems

An integrated Store Management System was developed that linked all processes from questionnaire printing all the way through to data processing. The Instrument Tracking System that is used by the Survey Operations team is maintained by the Corporate Data Processing division. The systems are designed to be end-to end and have been automated in an attempt to reduce human error in processing. The processing phases have also been re-engineered to ensure that the four week deadline for publication of the QLFS results is met. An automated Editing and Imputation module (E&I) was also built into the processing systems. The three basic functions to the E&I, these are:

- Function A - record acceptance, where blank or missing information is removed before processing can proceed.
- Function B - editing and imputation where errors in the questionnaire are detected through logic rules and these are then corrected with the imputation function
- Function C – includes all post-cleaning up of derived variables and preparation for data weighting.

- Timeliness

Questionnaires that are completed in the field are processed continuously as they arrive from the field. This continuous processing of the questionnaires allows for the data to be released four weeks after the enumeration quarter.

### Conclusion

The effort to reduce response error will remain an integral part of the data collection process. There will always be new innovative approaches introduced and for any field management team, the ideal is to meet a balance that caters for optimal productivity with quality that is not compromised. In alignment with the broader Stats SA vision, the Survey Operations team at Stats SA will continue to investigate innovative initiatives in data collection.

**Figure 1**

The analysis and publication of results for the QLFS

**REFERENCES**

American Statistical Association. 1993. How to Conduct Pretesting. *Survey Research Methods*. American Statistical Association.

Biemer, P.P., R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman, eds. 1991. *Measurement Errors in Surveys*. John Wiley and Sons, New York.

**Guide to the Quarterly Labour Force Survey** Statistics South Africa. Pretoria: Statistics South Africa, 2008

Payne, S.L. 1951. *The Art of Asking Questions*, Princeton University Press, Princeton, New Jersey.

Tourangeau, R., L.J. Rips and K. Rasinski. 2000. *The Psychology of Survey Response*, Cambridge University Press, Cambridge, U.K.