Statistics South Africa

Preferred supplier of quality statistics

Guide to the Quarterly Labour Force Survey

August 2008

Report number: 02-11-01

Statistics South Africa • Mbalo-mbalo dza Afrika Tshipembe • Tinhlayo-tiko ta Afrika-Dzonga • Dipalopalo tsa Aforika Borwa • Ezezibalo zaseNingizimu Afrika
Dipalopalo tša Afrika Borwa • Dipalopalo tsa Afrika Borwa • Ubalo lwaseMzantsi Afrika • Telubalo eNingizimu Afrika • iNanimbalo leSewula Afrika • Statistiek Suid-Afrika

email: info@statssa.gov.za
www.statssa.gov.za

170 Andries Street • Private Bag X44, 0001 Pretoria, South Africa
Tel: +27(12) 310 8911, Fax: +27(12) 321 7381

Table of contents

1. Background

The Quarterly Labour Force Survey (QLFS) is a household-based sample survey conducted by Statistics South Africa (Stats SA). It collects data on the labour market activity of individuals aged 15 years or older who live in South Africa.

Starting in 2005, Stats SA undertook a major revision of the Labour Force Survey (LFS). This revision resulted in changes to the survey methodology, the survey questionnaire, the frequency of data collection and data releases, and the survey data capture and processing systems. The redesigned labour market survey is the QLFS which was launched in 2008.

2. Objective of this guide

The objective of this guide is to provide the information necessary for users of the QLFS data to understand all aspects of the QLFS survey process and their impact on the data.

Other documents that provide more detail on various elements of the re-design process are available on the Stats SA website: www.statssa.gov.za/qlfs/index.asp

3. History of the labour force survey

The October Household Survey (OHS)

The principal vehicle for collecting labour market information for the whole country over the period 1994–1999 was the annual October Household Survey (OHS). However, the OHS also collected information from respondents about a diverse range of issues relating to: births and deaths, health, crime, education and training initiatives as well as the services and amenities available to the dwelling(s) in which households lived etc. Changes were made to the OHS sample design for successive surveys. Essentially, the OHSs were independent cross-sectional surveys that had different sample designs. Over the years, the labour market component of the OHS questionnaire was also changed to accommodate both national requirements in terms of providing information to inform policymakers and international requirements that conformed to the standards of the International Labour Organisation (ILO).

The Labour Force Survey (LFS)

The first LFS was conducted in 2000 and since then it has been undertaken on a six-monthly basis in March and September each year. The LFS is more focused on labour issues than its predecessor (the OHS) since the bulk of the non-labour questions were channeled to the General Household Survey (GHS).

As with the OHS, the LFS sample is representative of all provinces and strata (which are District Councils) within provinces. However, since 2000, Stats SA has used a Master Sample of 3 000 Primary Sampling Units (PSUs) from the population census as the sampling frame for the LFS. As a result, and unlike the OHSs, the sampling methodology was consistent in each round of the survey. The intention was that the selected dwelling units would remain in the sample for five consecutive surveys, with one-fifth of these dwelling units rotating out each round of the survey. The dwelling unit approach is adopted because households are mobile and cannot easily be tracked. The unit of sampling is therefore the dwelling unit and the unit of observation is the household.

The Quarterly Labour Force Survey (QLFS)

The decision to redesign all aspects of the LFS emanated from criticisms by data users and these are documented in the report written by International Monetary Fund (IMF) consultants in June 2005[1]. These criticisms related to the scope, coverage, timeliness and frequency of the survey.

In addressing these issues, Stats SA decided to embark on a quarterly cycle for the collection of labour market information. Increasing the frequency of the survey, coupled with the additional requirement to release results in a timely fashion required the following:

- Continuous data collection.
- Automated data processing system.

A new Master Sample and listing procedures have been developed, new fieldwork procedures have been implemented, and a shorter core questionnaire and an end-to-end data processing system has also been developed. These are summarised in this document and greater detail is provided in various documents on the Stats SA website: www.statssa.gov.za/qlfs/index.asp.

[1] Stats SA website @ www.statssa.gov.za/qlfs/index.asp.

4. Objective of the QLFS

The objective of the QLFS is to collect quarterly information about persons in the labour market, i.e., those who are employed; those who are unemployed and those who are not economically active.

This information will be published as core labour market indicators[2] four weeks after the end of each quarter and an annual report and supplementary data will be published six months after the end of each calendar year.

5. The QLFS survey design

Introduction

The QLFS sample covers the non-institutional population except for workers' hostels. However, persons living in private dwelling units within institutions are also enumerated. For example, within a school compound, you would enumerate the schoolmaster's house and teachers' accommodation because these are private dwellings. Students living in a dormitory on the school compound would therefore be excluded.

Survey requirements and design

The Labour Force Survey frame has been developed as a general purpose household survey frame that can be used by all other household surveys irrespective of the sample size requirement of the survey. The sample size for the QLFS is roughly 30 000 dwellings and these are divided equally into four rotation groups, i.e. 7 500 dwellings per rotation group.

The sample is based on information collected during the 2001 Population Census conducted by Stats SA. In preparation for the 2001 census, the country was divided into 80 787 enumeration areas (EAs). Some of these EAs are small in terms of the number of households that were enumerated in them at the time of Census 2001. Stats SA's household-based surveys use a Master Sample which comprises of EAs that are drawn from across the country. For the purposes of the Master Sample the EAs that contained less than 25 households were excluded from the sampling frame, and those that contained between 25 and 99 households were combined with other EAs to form Primary Sampling Units (PSUs). The number of EAs per PSU ranges between one and four. On the other hand, very large EAs represent two or more PSUs.

The sample is designed to be representative at the provincial level and within provinces at the metro/non-metro level. Within the metros, the sample is further distributed by geography type. The four geography types are: urban formal, urban informal, farms and tribal. This implies that for example, that within a metropolitan area the sample is designed to be representative at the different geography types that may exist within that metro.

The current sample size is 3 080 PSUs. It is equally divided into four sub-groups or panels called rotation groups. The rotation groups are designed in such a way that each of these groups has the same distribution pattern as that which is observed in the whole sample. They are numbered from one to four and these numbers also correspond to the quarters of the year in which the sample will be rotated for the particular group.

The sample for the redesigned Labour Force Survey is based on a stratified two-stage design with probability proportional to size (PPS) sampling of primary sampling units (PSUs) in the first stage, and sampling of dwelling units (DUs) with systematic sampling in the second stage.

Sample rotation

The sampled PSUs have been assigned to 4 rotation groups, and dwellings selected from the PSUs assigned to rotation group "1" are rotated in the first quarter. Similarly, the dwellings selected from the PSUs assigned to rotation group "2" are rotated in the second quarter, and so on. Thus, each sampled dwelling will remain in the sample for four consecutive quarters. It should be noted that the sampling unit is the dwelling, and the unit of observation is the household. Therefore, if a household moves out of a dwelling after being in the sample for, say 2 quarters and a new household moves in then the new household will be enumerated for the next two quarters. If no household moves into the sampled dwelling, the dwelling will be classified as vacant (unoccupied).

Each quarter, ¼ of the sampled dwellings rotate out of the sample and are replaced by new dwellings from the same PSU or the next PSU on the list. A total of 3 080 PSUs were selected for the redesigned LFS, and 770 have been assigned to each of the four rotation groups.

[2] The core QLFS questionnaire was specifically designed for this purpose. Supplements will be developed in due course depending on the needs of users.

Weighting

The sampling weights for the data collected from the sampled households are constructed so that the responses could be properly expanded to represent the entire civilian population of South Africa. The weights are the result of calculations involving several factors, including original selection probabilities, adjustment for non-response, and benchmarking to known population estimates from the Demographic Division of Stats SA.

The base weight is defined as the product of the provincial Inverse Sampling Rate (ISR) and the three adjustment factors, namely, adjustment factor for informal PSUs, adjustment factor for sub-sampling of growth PSUs and an adjustment factor to account for small EAs excluded from the sampling frame (i.e. EAs with less than 25 households).

Non-response adjustment

In general, imputation is used for item non-response (i.e blanks within the questionnaire); edit failure (i.e invalid or inconsistent responses) and weight adjustment to account for the non-respondent households (e.g. refusal, no contact, etc). The sampled dwellings with no eligible households, e.g. foreigners only, or no households, (i.e. vacant dwellings), do not contribute to the survey. The eligible households in the sampled dwellings can be divided into two response categories: respondents and non-respondents.

The non-response adjusted weight is the product of the base weight with the non-response adjustment factor given above. If the PSU level non-response rate is too high the non-response adjustment is applied at the VARUNIT level, where two VARUNITs have been created by grouping PSUs within strata level. PSU level non-response adjustment is applied only if the corresponding adjustment factor is less than 1,5.

Final survey weights

The final survey weights are constructed using regression estimation to calibrate to the known population counts at the national level population estimates (which are supplied by the Demography Division) cross-classified by 5-year age groups, gender and race, and provincial population estimates by broad age groups are used for calibration weighting. The 5-year age groups are: 0–4, 5–9, 10–14, 55–59, 60–64 and 65 and over. The provincial level age groups are: 0–14, 15–64 and 65 years and over.

Estimation

The final survey weights are used to obtain the estimates for various domains of interest, e.g. number of persons employed in agriculture in the province of Western Cape, number of females employed in manufacturing, etc. The estimates of ratios are obtained as ratios of the estimated totals. Thus, survey estimates for any estimation domain can be computed using the set of final weights for the respondents in the domain of interest.

Reliability of the survey estimates

Because estimates are based on sample data, they differ from figures that would have been obtained from complete enumeration of the population using the same instrument. Results are subject to both sampling and non-sampling errors. Non-sampling errors include biases from inaccurate reporting, processing, and tabulation etc., as well as errors from non-response and incomplete reporting. These types of errors cannot be measured readily. However, to the extent possible, non-sampling errors can be minimised through the procedures used for data collection, editing, quality control, and non-response adjustment. The variances of the survey estimates are used to measure sampling errors. The variance estimation methodology is discussed in the next section.

Variance estimation

The most commonly used methods for estimating variances of survey estimates from complex surveys, such as the QLFS, are the Taylor-series Linearization, Jackknife Replication, Balanced Repeated Replication (BRR), and Bootstrap methods (Wolter, 2007)[3]. We implemented the replication method for the QLFS mainly because of simplicity[4]. The QLFS sampled 3 080 PSUs by selecting an even number of 4 or more PSUs from within strata. The Jackknife method would be applicable for the sample design with more than two PSUs per stratum but this would result in 3 080 replicates, which would be computationally very intensive. The Fay's BRR method on the other hand is applicable when two primary sampling units (PSUs) are sampled from each stratum. Therefore, we decided to use Fay's BRR method by collapsing PSUs into two groups of PSUs within each stratum.

[3] Wolter, K. M. (2007), *Introduction to Variance Estimation, 2nd Edition,* Springer-Verlag: New York.

[4] Note that variance estimation methodology was not implemented for the LFS.

5.6.2 Other measures of precision

In practice, the sampling variance itself is hardly ever reported. Instead, users find it more useful to rely on one of the derivatives of the sampling variance, such as the standard error, the coefficient of variation, the margin of error, or the confidence interval. These are all related expressions, and it is quite easy to go from one to the other using simple mathematical operations.

Standard error

The standard error of an estimator is the square root of its sampling variance. This measure is easier to interpret since it provides an indication of sampling error using the same scale as the estimate whereas the variance is based on squared differences. If $\hat{\theta}$ is the estimate of a given population parameter θ (e.g., true employment but unknown) and $v(\hat{\theta})$ is the corresponding estimate of its variance, then the standard error of the estimate is defined as $se\left(\hat{\theta}\right) = \sqrt{v\left(\hat{\theta}\right)}$.

Coefficient of variation

It is more useful in many situations to assess the size of the standard error relative to the magnitude of the characteristic being measured. The ***coefficient of variation*** $\left(cv\right)$ provides such a measure. It is the ***ratio of the standard error of the survey estimate to the value of the estimate itself expressed as percentage.*** It is very useful in comparing the precision of several different survey estimates, where their sizes or scale differ from one another.

Confidence intervals

The 95 per cent confidence interval is the interval such that there is a 95 per cent probability (chance of 19 out of 20) of the unknown population parameter θ being within the interval. The 95 percent confidence interval is given by $\hat{\theta} \pm 1.96 \times se\left(\hat{\theta}\right)$. The lower limit of the interval is $\hat{\theta} - 1.96 \times se\left(\hat{\theta}\right)$, and the upper limit of the interval is $\hat{\theta} + 1.96 \times se\left(\hat{\theta}\right)$. The width $1.96 \times se\left(\hat{\theta}\right)$ is known as half-width of the 95 per cent confidence interval. The smaller the half-width of the confidence interval, the more precise is the survey estimate.

Design effects

Most surveys are based on complex designs involving stratification, and clustering due to multi-stage designs. Moreover, the weighting involves non-linear adjustments (e.g., non-response and weight calibration adjustments, etc.). It is crucial that these features of the complex survey design be accounted for in the variance estimation (Choudhry and Valliant, 2003). The ***design effect*** compares the variance of the estimate from the sample design that was actually implemented to the variance of the estimate that would have been obtained from a simple random sample (SRS) design. ***Design effect*** is another way to evaluate the efficiency of a sample design and the procedure used to develop the survey estimates. Design effect is defined as the ratio of the variance of an estimate for a complex sample design and the variance of the estimate under the SRS design with the same sample size. Kish (1965) introduced the concept of design effect to deal with complex sample designs involving stratification and clustering. Stratification generally leads to a gain in efficiency over simple random sampling, but clustering leads to deterioration in the efficiency of the sample design due to positive intra-cluster correlation among units in the cluster (PSUs in the case of QLFS). To determine the total effect of any complex design on the sampling variance in comparison to the alternative simple random sample design, the design effect $\left(deff\right)$ is defined as:

$$Deff = \frac{sampling \;\; variance \;\; of \;\; a \;\; complex \;\; sample \;\; design}{sampling \;\; variance \;\; of \;\; simple \;\; random \;\; sample \;\; design}$$

A design effect can be derived for any sampling design and estimator, provided we can compute a sampling variance. It is important to note that the design effect is associated with both the design and the estimator; therefore, for a given survey, the design effect can vary substantially lot from one variable to another.

Effective sample size

Another concept that is often used is effective sample size defined as the actual sample size that was selected for the complex design divided by the corresponding design effect. The effective sample size can be interpreted as the sample size that would be needed for an SRS design to obtain the same variance as that obtained with the complex design (i.e. the design that was actually implemented).

6. Determining labour market status

Introduction

This section focuses on the international recommendations for determining labour market status. The method of deriving the key labour market components; employed; unemployed and not economically active is discussed, as well as the issues surrounding the inclusion of non-market production activities as employment. The implementation of the framework in the South African context is discussed in the section that follows (Section 7).

The labour market consists of a supply side and a demand side. The labour supply of the population, referred to as the economically active population or labour force, has two components: employed persons and unemployed persons. The labour demand of enterprises consists of two components, filled posts (jobs) and unfilled posts (vacancies).

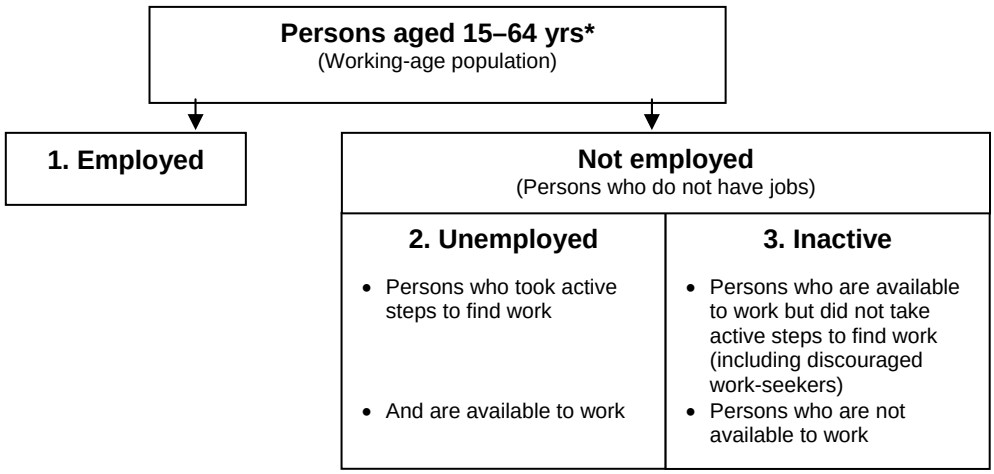
As noted by the ILO[5], statistics of the economically active population, employment, unemployment and underemployment serve many purposes. They provide measures of labour supply, labour input, the structure of employment, and the extent to which the available labour time and human resources are actually utilised or not. Such information is essential for macro-economic and human resources development planning and policy formulation. When collected at different points in time, the data provide the basis for monitoring current trends and changes in the labour market and the employment situation, which may be analysed in connection with other economic and social phenomena to evaluate macro-economic policies. The unemployment rate, in particular, is widely used as an overall indicator of the current performance of a nation's economy.

The labour force framework

The labour force framework is the basis for the joint measurement of employment and unemployment according to the international guidelines[6]. The requirements of the framework ensure that the three categories - employed, unemployed and economically inactive - are exhaustive and mutually exclusive. The framework also requires that precedence is given to employment (even if only for one hour during the reference week) over unemployment and to unemployment over economic inactivity. Thus, a person who is both working and seeking work is classified as employed.

Figure 1 shows that the working age population is divided into two broad labour market groups – persons that are **employed** and those that are **not employed**. Persons that are not employed are further divided into those who are **unemployed** and those who are **inactive**.

Figure 1: The labour force framework



*Note: The international guidelines suggest that the appropriate age cut-off for inclusion in the labour market is determined by the national legislation of member countries The QLFS nevertheless collects labour market activity data for persons aged 65 years and older.

[5] Hussmanns, R (ILO Bureau of Statistics) 2007. Measurement of employment, unemployment and underemployment –Current international standards and issues in their application.

[6] Hussmanns, R.; Mehran, F.; Verma, V. 1990: Surveys of economically active population, employment, unemployment and underemployment: An ILO manual on concepts and methods, ILO, Geneva, 1990.

Non-market production activities

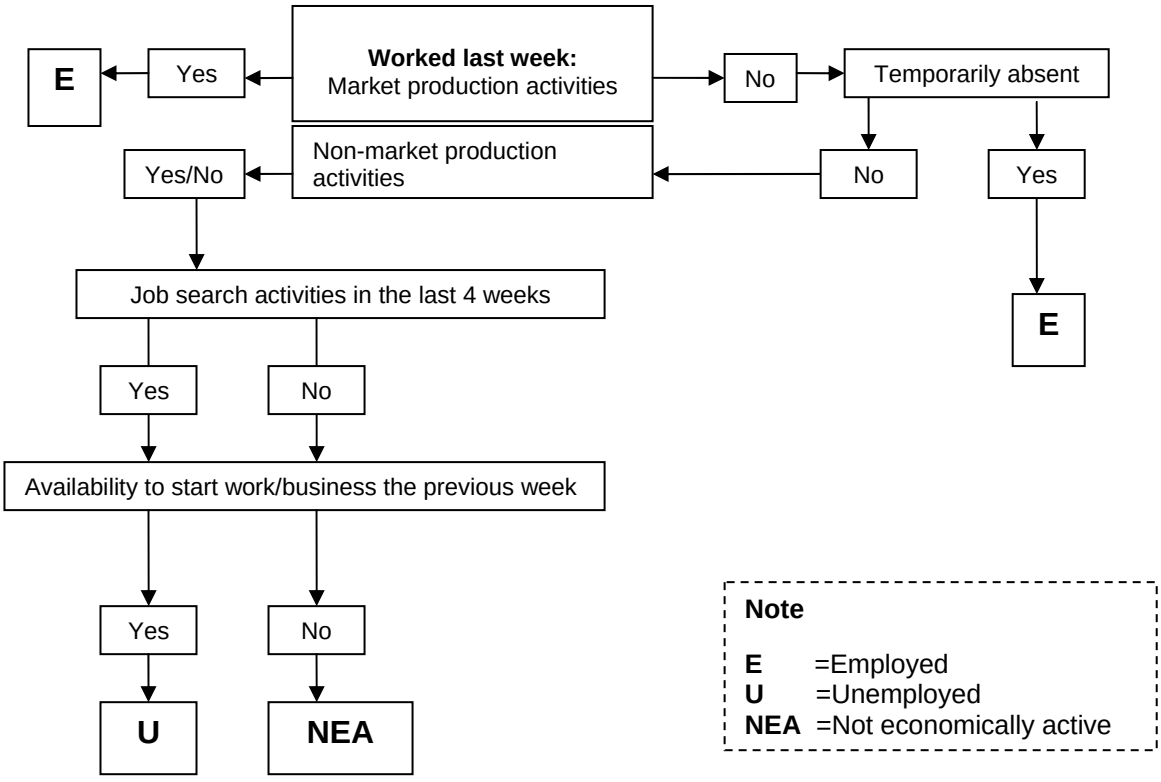
Non-market production activities are those carried out solely for the consumption/use of the household (subsistence farming, hunting, fishing etc.,). "So as to cover market production completely, it is necessary to include some non-market production as well, as it is in practice often impossible to measure the market component separately when the same persons or households are engaged in both types of production. For example, at the time when the production of an agricultural crop takes place, it is often impossible to determine how much of it will be used for own final consumption, for storage, or for sale or barter.

The international standards mention, however, that these persons should be considered employed only if such production comprises an important contribution to the total consumption of the household…The important contribution provision also serves to exclude from the employed population persons who may, for example, be growing some vegetables in their backyards but whose subsistence does not significantly depend on it. However, its implementation in labour force surveys is difficult, due to the lack of information on total household consumption in such surveys. For this reason, some countries have resorted to assessing the importance of household production for own final use on the basis of information on the volume of labour inputs to such production (number of hours worked)"[7]. The solution to this problem in the South African context is discussed in Section 7.2.

7. Implementing the labour force framework in the QLFS

Figure 2 shows the key skip patterns followed in the QLFS questionnaire to classify persons into the major labour market categories that are consistent with the international guidelines.

Figure 2: Labour force classification



NOTE: The following persons are not considered as employed and are therefore routed to questions in Section 3 of the QLFS questionnaire to determine whether they are unemployed or inactive:

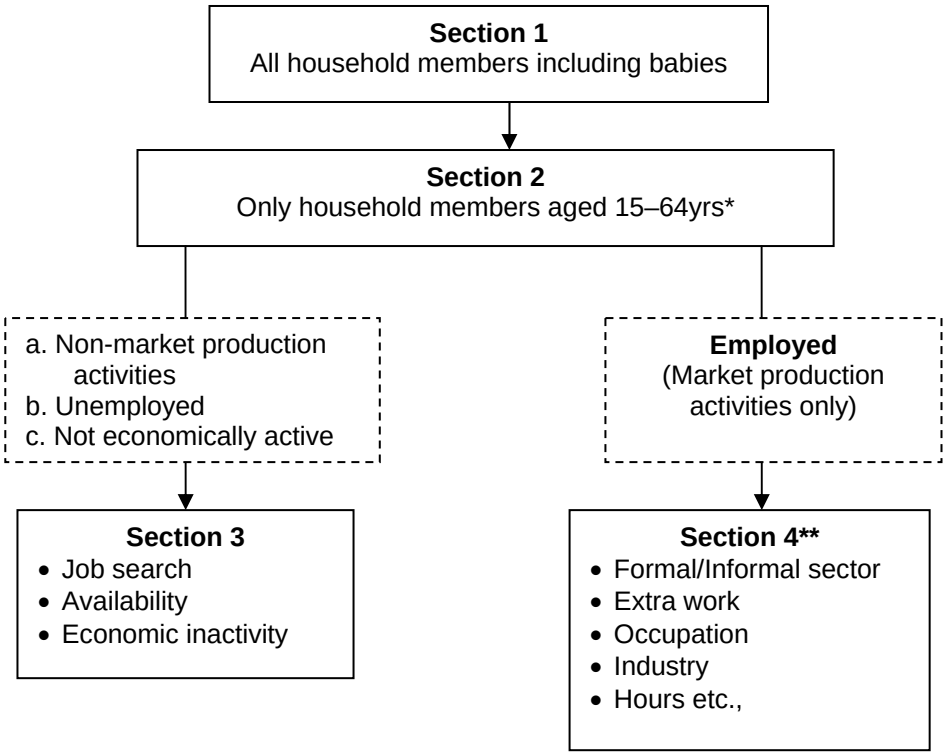
1. Persons helping unpaid in their household business that were temporarily absent from such work in the reference week.
2. Persons who were absent from work in the reference week because of seasonal factors.
3. Persons who were absent from work in the reference week because they had a new job or business to start at a definite date in the future.

[7] Hussmanns R; (ILO Bureau of Statistics), 2007. Measurement of employment, unemployment and underemployment – Current international standards and issues in their application.

Employment

The conceptual framework is reflected in the structure of, and skip pattern embedded in the key employment questions in the QLFS questionnaire (Figure 2). A clear distinction is made between market and non-market production activities. Market production activities are grouped into one question while non-market production activities are identified in a separate question. In terms of the latter, provision is also made for all types of non-market production activities (including fetching wood/water/dung for household use and producing other goods such as clay pots etc., for household use).

Figure 3: Structure of the QLFS questionnaire



*Note: The international guidelines suggest that the appropriate age cut-off for inclusion in the labour market is determined by the national legislation of member countries The QLFS nevertheless collects labour market activity data for those aged 65 years and older.
** See the detailed questions in Appendix 1.

As shown in Figure 3, in the QLFS questionnaire, to reduce respondent burden, persons identified as engaged in market production activities do not answer additional questions about whether or not they were engaged in non-market production activities during the reference week. Instead, they are routed directly to questions about various aspects of their work situation in Section 4.

Persons employed in market production activities are those (aged 15-64 years) who during the reference week, even if it was for only one hour, did any of the following:

a) Worked for a wage, salary, commission or payment in kind (including paid domestic work).

b) Ran any kind of business, big or small, on their own or with one or more partners.

c) Helped without being paid in a business run by another household member. Persons helping unpaid in such businesses who were temporarily absent in the reference week are not considered as employed, they are routed eventually to questions about: job search activities; their desire to work; and their availability to work - to determine whether they are unemployed or inactive.

d) Were temporarily absent from their jobs or businesses to which they would definitely return. They could have been absent in the reference week but definitely returning to their job if the reason given for their absence was any of the following: ill-health, vacation leave, caring for family or others; maternity or paternity leave, other family/community obligations; strike/stay-away/lockout; problems with transport; bad weather; study or training leave; unrest; temporarily laid off; other reasons. However, as indicated in Figure 2, there are two groups of individuals that may have been temporarily absent from their jobs in the reference week who are not considered as being employed as follows:
   i) Persons who were absent from work in the reference week because of seasonal factors.
   ii) Persons who were absent from work in the reference week because they had a new job or business to start at a definite date in the future.

Such individuals (i and ii above) are routed to questions in Section 3 of the QLFS questionnaire to determine whether they are unemployed or inactive.

Non-market production activities in the QLFS

Individuals who are engaged in non-market production activities, such as subsistence farming, fetching wood and collecting water, making clay pots, etc., **for the use or consumption of their own household** are routed to questions in Section 3 that determine whether or not they are unemployed or economically inactive (Figure 3).

As noted earlier, the international guidelines require that non-market production activities be included as employment if they make a substantial contribution to the total consumption of the household. At this juncture, in light of the difficulty of establishing robust criteria for their inclusion as employment, Stats SA identifies persons engaged in non-market production activities separately. Among them, the two components are then identified as follows:

- the unemployed (i.e. persons engaged only in non-market production activities, who actively sought work and were available for work in the reference period).
- the not economically active (i.e. persons engaged in non-market production activities who did not engage in any type of job-search activity during the reference period).

Unemployment

Unemployed persons are those (aged 15–64 years) who:

a) Were not employed in the reference week.
b) Actively looked for work or tried to start a business in the four weeks preceding the survey interview. And
c) Would have been able to start work or would have started a business in the reference week.

Persons who stated that they had not looked for work in the reference period because they had already arranged to take up a job or to start a business at some later date are not required to have actively sought work in the reference period. They are included as unemployed if they would have been available to start work/business in the previous week.

Not economically active

Not economically active persons are those (aged 15–64 years) who:

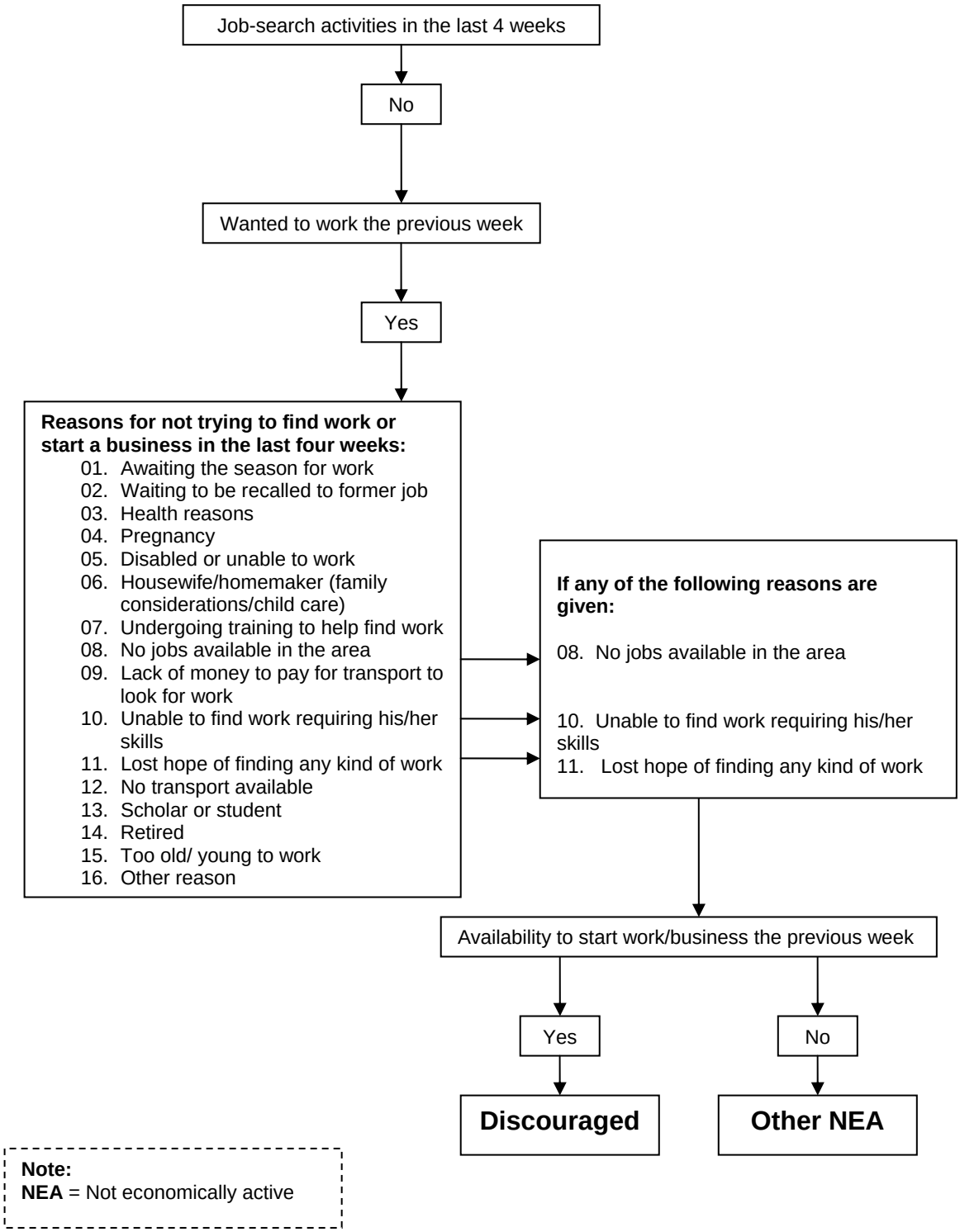
a) Did not work in the reference week;
b) Did not look for work or try to start a business in the four weeks preceding the survey, or;
c) Were not available to start work or a business in the reference week.

Figure 4 shows that among those who fall into the not economically active category, a sub-set is identified as “discouraged work-seekers” on the basis that they wanted to work and the main reason why they did not try to find work or start a business in the reference period was any of the following:

i) No jobs available in the area.
ii) Unable to find work requiring his/her skills.
iii) Lost hope of finding any kind of work.

Against this background, other labour market indicators that describe various aspects of the employed, the unemployed and the not economically active are discussed in greater detail – in “Section 10.4 Derived variables” – of this report.

Figure 4: Discouraged work-seekers



8. Data collection

Introduction

To facilitate continuous data collection, training and fieldwork monitoring from the regional offices across the country, permanent field staff (332) were appointed to conduct the QLFS. In addition, 95 contract staff have been recruited to assist in areas such as listing[8] and the capturing of publicity[9] forms (Table 1).

The appointment of a permanent fieldwork force for the QLFS marks an important break with the past practice of Stats SA regarding fieldwork for household-based surveys. For the first time, a household-based survey – the QLFS - will benefit from the continuity and institutional memory that develops through adopting this approach, which allows for: refresher training; performance monitoring and where necessary remedial action; and interviewer/respondent rapport to grow over the course of the four interviews.

Table 1: QLFS field operations

Provincial staff	
Permanent:	
9	Provincial survey coordinators
9	Administrative support officers
54	District survey coordinators
27	Provincial quality monitors
233	Survey Officers
Contract:	
95	Listers, data capturers

Two computer systems have been implemented to track and monitor the enumeration process:

i) Real time Management System (RTMS)
ii) Assignment planning system

Real Time Management System (RTMS)

RTMS is an important management tool for personnel involved with field operations to monitor progress. The applications used by field operations are as follows:

RTMS Stage 1: Recording of publicity at gate keeper level (police, indunas, street committees, farmers' unions etc.,) for the Master Sample.
RTMS Stage 2: The tracking and monitoring of PSU listing books from the provinces to Head Office (HO) as well as through the processing phase,
RTMS Stage 3: Recording of publicity at dwelling unit and household level.
RTMS Stage 4: The tracking and monitoring of the completed questionnaires from the provinces to HO as well as through the processing phase.

The inclusion of a barcode that is linked to a unique number on each QLFS questionnaire allows the tracking of questionnaires throughout the survey process.

The unique number is made up as follows:

a) PSU number (the first eight digits).
b) Segment number (the next three digits).
c) Dwelling unit number (the next five digits).
d) Household number (the next two digits).
e) Questionnaire number is the last digit.

[8] The initial stages of sample selection end with the selection of Primary Sampling Units (PSU). Once the PSUs have been selected a list needs to be made of all of the dwelling units in that PSU. From this list a sample of specific dwelling units is selected.

[9] The QLFS's publicity program is vital in reducing non-response. Publicity starts prior to a PSU being brought into the sample (i.e., before the first interview takes place). Publicity officers visit the PSUs to with two purposes. They make local officials aware that the survey will be taking place in their area and to seek their endorsement. They also visit each dwelling unit in the sample to inform those living there that they will be interviewed shortly for the LFS, to answer any questions that they may have about the survey and to encourage them to participate.

The barcode on a completed questionnaire is scanned in the relevant regional office prior to it being sent by courier to the data processing centre in Pretoria where it is then logged into the store management system. On the basis of the link between the barcode and the unique number, each questionnaire can then be easily tracked throughout the system.

Assignment planning system

Assignment planning is a systematic approach to distributing the workload of QLFS field-staff. Its application allows management to track and monitor publicity and data collection as well as to provide rapid feedback to QLFS field-staff. A unique identifier – the assignment number – is pre-printed on all listing booklets, publicity forms and questionnaires that belong to the specific workload. This number incorporates information about: the Province (1st digit); the District Survey Coordinator (DSC) responsible for a specific district/region within that province (the next two digits); and lastly the Survey Officer(s) who report to that specific DSC.

Data collection

Household members living in approximately 10 000 dwelling units in 1 025 PSUs are interviewed in each of the three months within a quarter. Key information from completed questionnaires is captured by data capturers in the regional office using the RTMS. This includes whether or not the interview was successful, thus allowing follow-up by senior staff in the event of refusals.

The QLFS data collection strategy is based on a '0110' approach. The first digit "0" represents the first week of any given month. During this week set-up interviews/publicity and listing maintenance is conducted. Data collection is conducted during the middle two weeks ("11") of each month (except in January and December). The last week ("0") is dedicated to completing the work allocation assigned for that month including the listing of growth areas in the sampled PSU identified during the first week. In summary, the "0"represents a non-data collection week and the "1" represents the two weeks of data collection.

Pre-data collection activities

Training and refresher training

The QLFS training initiative currently in place has two elements. Firstly, existing field-staff attend refresher training sessions that are conducted on a quarterly basis in the regional offices on issues relating to listing, listing maintenance, and data collection. This ensures that field-staff are aware of any new developments regarding the administering of the core QLFS questionnaire as well as Master Sample related issues. Secondly, turnover training is conducted for new field-staff. Each quarter time is set aside for training activities and all training includes field-practice sessions prior to actual field work.

The training materials that have been developed in support of the new training initiative are:

- Publicity manual for listing and data collection;
- Survey Officer's manual;
- Quick reference guides for publicity, listing and data collection;
- Listing and maintenance procedures manual;
- Training guide;
- Quality assurance manual;
- Evaluation exercises; and
- Survey Officer's operations calendar.

The choice of training materials is driven by an analysis of the error patterns revealed by the edit and imputation module of the Head Office Processing System.

Logistics

The quarterly nature of the survey requires that as soon as questionnaires and listing booklets are completed in the various regions of each of the nine provinces, they are sent to the Data Processing Center (DPC) in Pretoria. This facilitates the continuous capture of questionnaires and ultimately the ability to release results four weeks after the end of the quarter. Completed questionnaires are therefore sent by courier to the DPC twice each week.

Publicity

Publicity or set-up interviews are undertaken with stakeholders each month. These are aimed primarily at creating awareness about the survey and the presence of Stats SA field-staff in the sampled PSUs.

Field-staff first visit various gatekeepers / access controllers (police stations, indunas, farmers' unions, street committees, body corporate members of gated communities etc.,) to obtain permission to work in the area. Having obtained permission, all dwelling units in the sample are then visited to set up interviews, with the respective household members, which will be conducted during the subsequent two weeks of data collection. The initial visits to sampled dwelling units are always conducted in person (face-to-face). Since each sampled dwelling unit is visited every quarter (for one year), subsequent publicity visits are conducted telephonically where there is scope for this action.

Listing and listing maintenance

The design of the new Master Sample warranted the implementation of innovative procedures that would ensure that the sampled PSUs were listed and maintained appropriately to guarantee that the subsequent sample of dwelling units is representative. In addition to the conventional 'Serpentine' listing (where field-staff list structures and features in the PSU on the left), two new procedures were introduced. These are:

- **The process of blocking in urban formal and some tribal areas**: This procedure requires that roads and streets are used to create the boundaries of blocks. Each of these blocks are then listed sequentially, ensuring that corner houses / structures are not erroneously omitted or duplicated.

- **The process of segmentation in urban informal settlements**: There are two phases to segmentation. Firstly, segment boundaries are created (using gravel roads, foot paths, dirt roads etc.,) and dwelling unit counts are recorded for each identified segment. Secondly, segment(s) are selected by methodologists at head office from which a sample of dwelling units is drawn for enumeration. This procedure was introduced to ensure that the sampled dwelling units could be easily identified to enable field-staff to return for subsequent visits.

9. Data processing

Introduction

The purpose of data processing is to ensure that the information collected from the sampled primary sampling units, dwelling units and households (i.e. the boxes containing QLFS questionnaires) are physically received, stored and processed. The aim is to produce a clean dataset that has all the information contained in the questionnaires. Except for the scanning system, all other elements of the data processing system were developed in-house.

One important innovation that is central to the smooth operation of the entire system is the development of barcodes that are linked to a unique number on each questionnaire. This information provides the link between the information recorded in the Master Sample database and other processes such as editing and imputation as well as weighting and variance estimation.

Processing phases

QLFS data processing is continuous, starting on the second week of every month. Data processing for each quarter must be completed by the first Friday of the subsequent month to ensure that the four-week deadline for publication of the QLFS results is met.

The phases listed below occur sequentially.

Receiving of questionnaires

The contents of the boxes containing questionnaires sent from the regional offices are verified when received at the DPC. The questionnaire barcodes captured in the provinces are captured again at the DPC to ensure that all questionnaires have been received.

Primary preparation

The purpose of primary preparation is to ensure that all questionnaires are correctly stacked and positioned prior to being guillotined.

Guillotining

The purpose of the guillotine process is to cut off the spines of the questionnaires in order to have pages separated for scanning.

Secondary preparation

The purpose of secondary preparation is to ensure that the questionnaires are correctly stacked and positioned for scanning. At the same time, quality assurance takes place on the work done during the primary preparation and guillotining processes.

Scanning

The purpose of scanning and recognition is to convert the questionnaires into an electronic format and Tagged Image File Format (TIFF) images.

Verification

The purpose of scanning verification is to manually correct un-interpretable characters, missing data and errors detected by validation rules.

Electronic coding

Industry and occupation codes are assigned using the electronic coding system which converts the respondents' industry and occupation descriptions into numeric codes based on Standard Industry Classification (SIC) and South African Standard Occupation Classification (SASCO). If the system fails to assign a code for either industry or occupation, the coding is assigned manually.

Automated editing and imputation

QLFS uses the editing and imputation module to ensure that output data is both clean and complete[10]. There are three basic components, called functions, in the Edit and Imputation Module:

Function A: Record acceptance
Function B: Edit and imputation
Function C: Clean up, derived variables and preparation for weighting

Function A: Record acceptance [11]

This function is divided into three phases:

First phase: *Pre-function A*

The first phase ensures that the records contain valid information in selected Cover Page questions required during edit and imputation and during the subsequent weighting and variance estimation. Any blanks or other errors that need to be corrected are done here before processing of the record can proceed.

Second phase: *Function A record acceptance*

The second phase ensures that there is enough demographic and labour market activity information to ensure that editing and imputation can be successfully completed.

Third phase: *Post Function A clean up*

This phase ensures that certain data are present where there is evidence that they should be. This for example, involves:

- Ensuring that if there is written material in the job description questions then there are corresponding industry and occupation codes for them.
- Ensuring that partial blanks or non-numeric characters that appear in questions where the Survey Officer is required to enter numbers are validated.
- Ensuring that where there is written material in the space provided for "Other – specify" that the corresponding option is marked.

Function B: Edit and imputation

Having determined in Function A that the content of the record would support extensive editing and imputation, this function carries out those activities. Editing is the detection of errors in the captured questionnaire. Imputation is the correction of the detected errors.[12]

[10] Details of the editing and imputation system are provided at Stats SA website: www.statssa.gov.za/qlfs/index.asp.
[11] Each record represents the information collected about a single respondent regardless of age. While it is possible to link all of the records belonging to a household (this is, in fact, done just prior to weighting), the records themselves only contain information about the characteristics and labour market activities of the respondent represented by a given record.

Function C: Clean up, derived variables and preparation for weighting

Function C includes all of the "post E&I clean up" functions such as "Off-path cleaning", "Result Code validation", verification of the presence of industry and occupation codes, and the generation of all derived variables.

Master Sample

All completed PSU listings, new and updated, are sent to the DPC for processing. The output of the processing of the listings is the dwelling frame that is used for sample selection.

Electronic systems for processing

Electronic data processing systems have been developed to ensure that the key QLFS results are published four weeks after the end of data collection each quarter. The system is fully automated and includes the seven sub-systems discussed in detail in the subsections that follow (9.3.1 to 9.3.6): Real Time Management System (RTMS)

RTMS serves two important functions. Firstly, it is a management tool for personnel involved with field operations to monitor progress. Secondly, it provides an important link between field operations and data processing as follows:

- Ensures that publicity information at the PSU and dwelling unit level can be rapidly assessed, thus allowing for speedy intervention should the need arise.
- Enables the tracking and monitoring of PSU listing books and questionnaires from the provinces to HO as well as through the processing phase.

Stores Management System (SMS)

Stores management ensures that completed QLFS questionnaires received from the field are physically stored and managed throughout all stages of the data processing process. This system is also used to control and monitor the flow of PSU boxes throughout the data processing process.

Scanning system

The scanning system converts the information contained in the QLFS questionnaires into a digital format. The system has been developed to scan, interpret and verify the contents of the QLFS questionnaires. Multiple TIFF images are uniquely stored using the questionnaire barcode. The system is fully integrated with other stages of data processing after data verification has been completed and the data transferred to the final centralised database. At this stage, the data is ready for editing and imputation.

Electronic coding system

Industry and occupation descriptions recorded in the QLFS questionnaires are electronically transferred into the electronic coding system which matches these descriptions with those already stored in a "Coding Knowledge Base". If a match is found, the code is assigned in the QLFS database. When a match is not found, the code is manually assigned in the electronic coding system and then transferred to the "Coding Knowledge Base" for use in future quarters.

Updating and image retrieval system

The system retrieves the information from the central database as well as the images stored during the scanning phase. The purpose of this system is to rectify missing and inconsistent information identified through Function A with the aim of ensuring that the data processing output is ready as input for the rest of the editing and imputation system.

Master Sample capturing system

The objective of the system is to provide an electronic dwelling unit frame from which samples for the QLFS and other household-based surveys are drawn. All information about the PSUs in the Master Sample that is recorded during fieldwork is captured, stored, retrieved and updated on a continuous basis to ensure that the dwelling unit frame used for sample selection is up-to-date.

Head Office Processing System (HOPS)

This is an integrated system that comprises the following:

- Record acceptance (Function A above);
- Editing and imputation (Function B above);
- Off-path cleanup and derived variables (Function C above);
- Weighting (see Section 5 above); and
- Tabulation.

[12] In some statistical programs the word "editing" is used to refer to both "editing" and "imputation as they are described in this note.

10. Questionnaire design and analysis

Questionnaire design

The core QLFS questionnaire has a cover page and two parts. Part 1 comprises of seven questions: two general questions to establish household membership etc.; and five socio-demographic questions (age, sex, martial status, population group, educational attainment). Persons aged 15 years and older are then screened to answer Part 2 of the questionnaire which has three sections. Each questionnaire allows for up to six household members to complete Part 2 on an individual basis.

Questionnaire structure and content

The inclusion of questions in the core questionnaire was guided by the need to have a minimum set of questions that would enable robust analysis of key labour market patterns on a quarterly basis, while at the same time not burdening respondents with a lengthy questionnaire. In light of this, two types of questions were included in the QLFS core questionnaire as follows:

i) Classification questions: Those required in determining labour market status.
ii) Descriptor questions: Those that provide insight into key labour market patterns.

In addition to the fieldwork and processing details required on the cover page of each questionnaire, the core QLFS questionnaire has in total 59 questions in Section 1 through to Section 4. The sequencing and phrasing of each question was tested through behind-the-glass observations and focus-groups in all the official languages. The questionnaire was also translated into all the official languages for reference during fieldwork. Four field tests using the new questionnaire were conducted during the period April 2006 to December 2007.

The contents of the QLFS questionnaire are summarised as follows[13]:

Cover page of the QLFS core questionnaire

The cover page contains details that enable the tracking of the questionnaire and monitoring of fieldwork at Head Office, the Provincial offices and District offices.

Section 1 of the QLFS core questionnaire

This section has socio-demographic questions which are completed for all household members regardless of age.

Section 2 of the QLFS core questionnaire

The questions in this section determine those individuals, aged 15–64 years, who are employed and those who are not employed.

Section 3 of the QLFS core questionnaire

This section determines which respondents are unemployed and which respondents are not economically active.

Section 4 of the QLFS core questionnaire

This section contains questions about the work situation of respondents who are employed. It includes questions about the number of jobs at which the respondent works, the hours of work, the industry and occupation of the respondent as well as whether or not the person is employed in the formal or informal sector etc.,

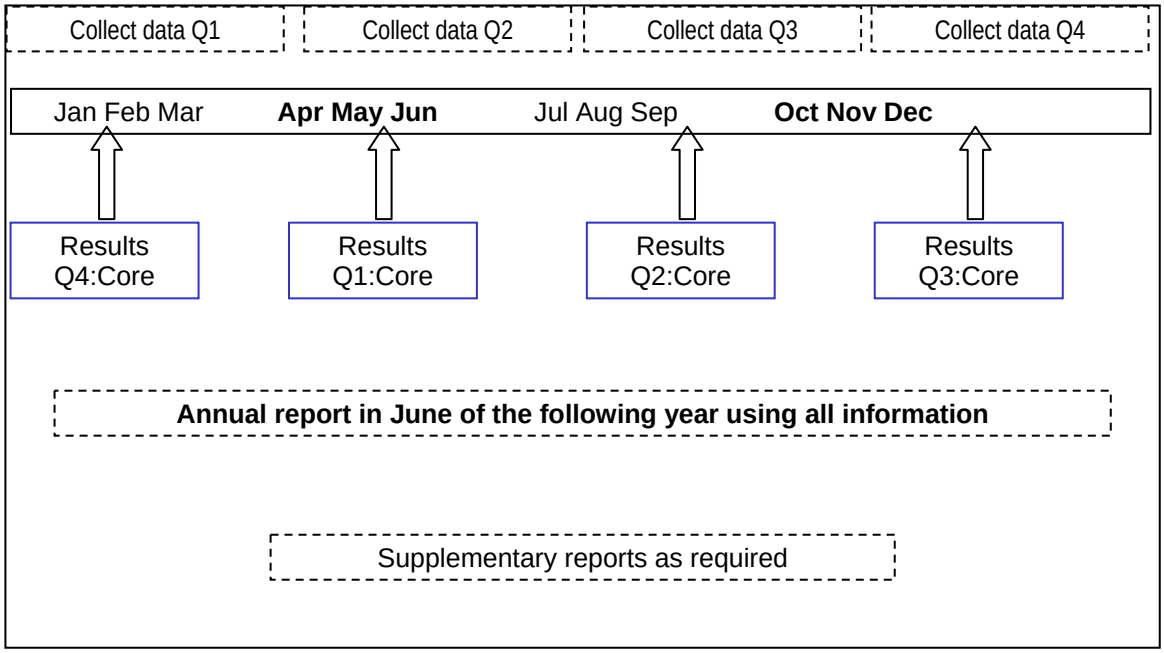
Analysis

Analysis of the core QLFS results is undertaken on a quarterly basis using a standard tabulation plan than includes all questions included in the core. As noted earlier, in each quarter there is a 75 per cent overlap of dwelling units, since 25 per cent of dwelling units in the Master Sample are rotated out each quarter. This ensures stability in the results (lower variance in estimates of change) and produces more robust estimates. Figure 5 shows that in any given calendar year, data collected in the first quarter (January to March) are analysed and published by end-April. In turn, data collected in the second quarter (April to June) are published end-July etc. In addition, an annual report that provides detailed analysis of the quarterly changes will be published six months after the end of each year.

The requirement to reduce the core QLFS questionnaire to a minimum length to facilitate the quarterly publication of results has consequences for the analysis of questions that were previously in the questionnaire. These questions are under review and will be subject to stakeholder consultations. They may well be re-introduced as a separate supplement on an annual basis. In addition, other supplements will also be undertaken as the need arises.

[13] See Appendix 1 for the actual questionnaire.

Figure 5: The analysis and publication of results



Derived variables

Classification questions

The three most important derived variables: employed; unemployed and not economically active; are discussed in detail in Section 6 of this guide in the context of the conceptual framework. Non-market activities and discouraged work-seekers are also discussed in Section 6. The classification questions used to derive the three major labour market components are as follows[14]:

Q2.4 Employment in market production activities
Q2.5 Temporary absence
Q2.7 Main reason for absence
Q3.1 Job search or starting a business in the past four weeks
Q3.3 Future starts
Q3.9 Availability to start work last week
Q3.10 Availability to start business last week

Other derived variables

Other derived variables that are used in the quarterly analysis of the QLFS are as follows:

Under-utilisation of labour

Under-utilised labour is calculated as the sum of the following three categories:

1. Unemployed;
2. Under-employed (persons who worked less than 35 hours during the reference week but wanted to work more hours and were available to work more hours); and
3. Discouraged work-seekers.

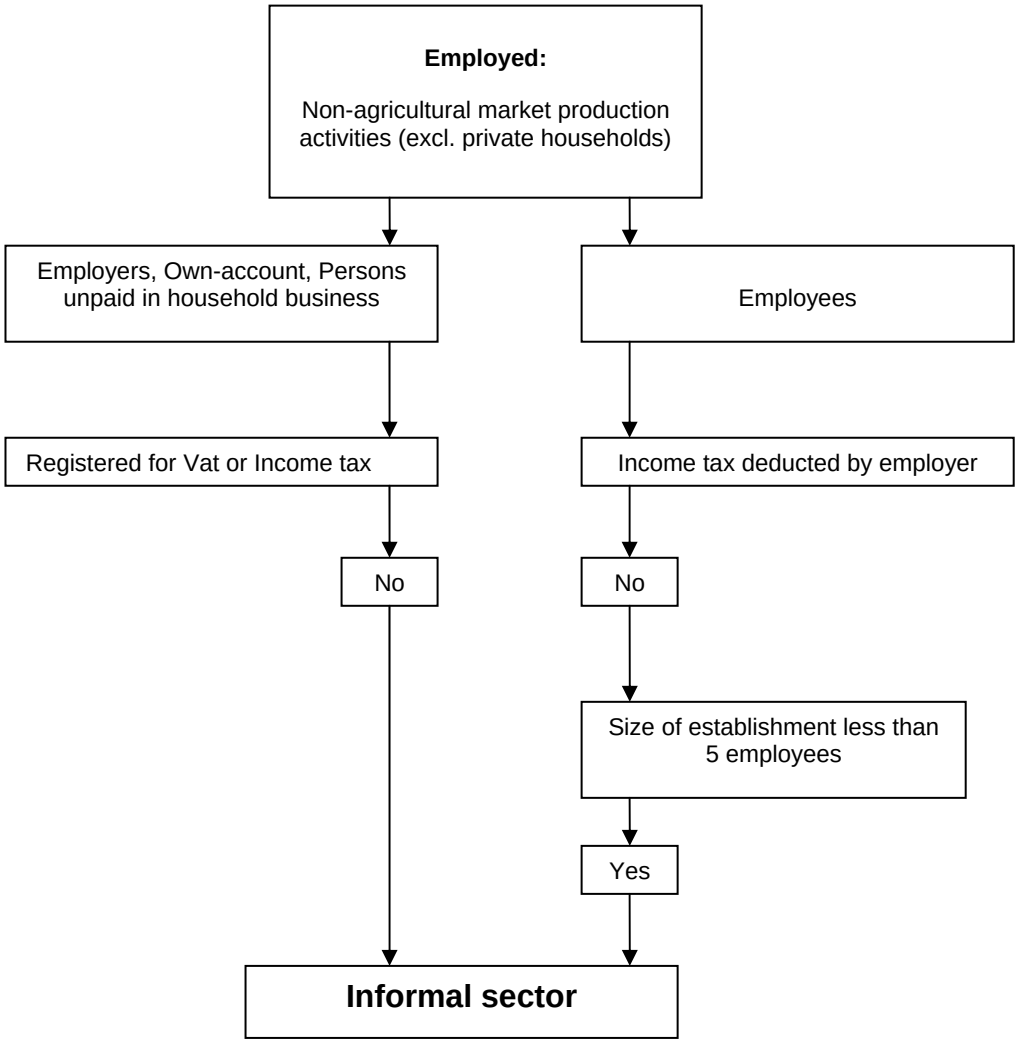
The informal sector

The definition of the informal sector is based on registration and the size classification of enterprises (Figure 6). The informal sector has the following two components:

1. Employees who are not registered for income tax and who work in establishments that employ less than five persons;
2. Employers, own-account workers and persons helping unpaid in their household business who are not registered for either income tax or value-added tax.

[14] See Appendix 1 for the exact questions used in the QLFS questionnaire.

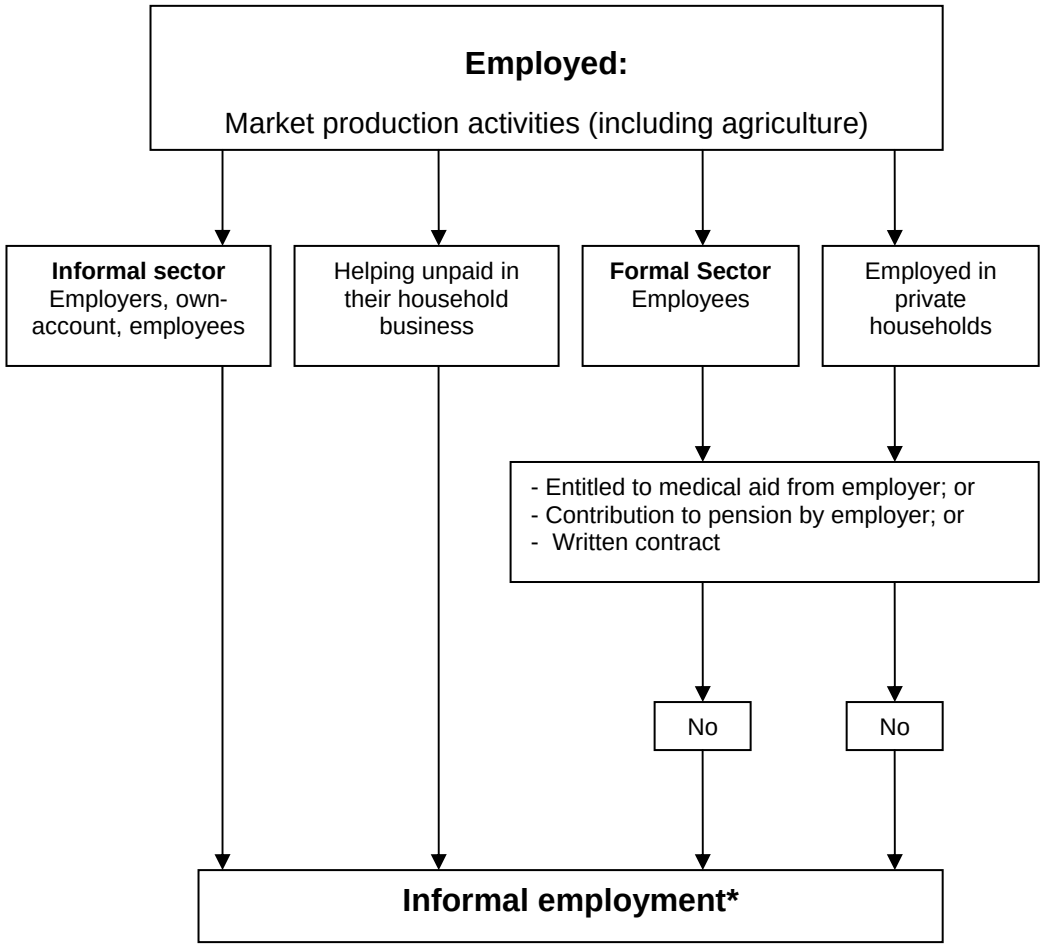
Figure 6: Deriving the informal sector in the QLFS



Informal employment

This indicator identifies persons who are in precarious employment situations. It includes all persons in the informal sector and persons helping unpaid in their family business. It also includes employees in the formal sector and persons employed in private households who are NOT entitled to basic benefits from their employer such as a pension or medical aid and who also do not have a written contract of employment (Figure 7).

Figure 7: Deriving informal employment in the QLFS



Time-related underemployment

This indicator is measured as a sub-set within employment for those who work less than 35 hours in the reference week and wanted to work additional hours and were available to do so.

Unemployed new entrants to the labour force

New entrants into unemployment are identified as persons who were unemployed during the reference period that had never worked before and were currently looking for work.

Unemployed re-entrants to the labour force

Re-entrants into unemployment are unemployed persons who worked before and who were currently looking for work, and whose main activity before looking for work was any of the following: managing a home; going to school; other reason.

Unemployed job losers

Job losers are unemployed persons were working when they became unemployed and had lost their job; they were laid off; the business in which they had previously worked had been sold or had closed down.

Unemployed job leavers

Unemployed job leavers are those among the unemployed who were working when they became unemployed and had stopped working at their last job for any of the following reasons:

- Caring for own children/relatives;
- Pregnancy;
- Other family/community responsibilities;
- Going to school;
- Changed residence;
- Retired; or
- Other reasons

11. Tabulation and dissemination of results

Tabulation

The main focus of the QLFS tabulation is to provide levels and quarter to quarter changes in absolute numbers and expressed as percentages.

The tables in the QLFS tabulation plan will be published in a standard format each quarter. These tables are populated at various levels of importance: national, province and population group. Small values will not be published since since these estimates are not reliable. Key labour market indicators and quarterly changes (in absolute numbers and as percentages) are presented for: employment in market production activities, unemployment and economic inactivity etc.. These tables also include derived variables such as the unemployment rate (where non-market production activities are excluded from employment), the labour force participation rate, the employment-to-population ratio (absorption) and the number of discouraged work-seekers. Persons engaged in non-market production activities by province are reported separately. Additional tables focusing on other aspects of the labour market are also included in the QLFS tabulation as follows:

Employed:
Age; highest level of education; marital status; time related under-employment; industry; occupation; informal sector; informal employment etc.,

Unemployed:
Age, highest level of education; marital status; long-term unemployment; new entrants; re-entrants; job-leavers; job-losers; previous occupation and industry etc.,

Not economically active:
Age, highest level of education; marital status; reasons for inactivity etc.,

Dissemination of results

The main products associated with the release of QLFS data are as follows:

1. A quarterly statistical release which will focus on quarter to quarter changes of key labour market indicators.

2. An annual report (from June 2009) that will provide in-depth analysis of the data collected throughout the year.

3. Unit record data sets for the QLFS will also be made available four weeks after the release of results each quarter.

In addition, revised labour market indicators will be published for the LFS to provide historical continuity. Link factors will be computed on the basis of an overlap between the QLFS and the LFS for key labour market variables. The LFS March series will be back cast to March 2001 and published in August 2008 and the September series will be back cast to September 2001 and published in March 2009.

12. Quality improvements in the QLFS

The major changes in the QLFS redesign translated into quality improvements that are summarised as follows[15]:

Methodology:

a. Development of a new Master Sample with an extended life that can serve the needs of an expanded household surveys programme.
b. Implementation of a sophisticated rotation scheme to avoid overlap with other surveys.
c. Development of variance estimation procedures to produce tools for the assessment of data quality. In addition, the new sample will yield estimates with lower sampling variance even though the two sample sizes are the same.

Field operations:

a. Recruitment of permanent field staff to ensure consistency in the application of fieldwork procedures in each round of the QLFS.
b. Development of training materials to facilitate a good grasp of the concepts, definitions and procedures used in the QLFS.
c. Assignment planning to use field-staff with optimum efficiency.
d. The use of questionnaire barcodes to enable the continuous tracking of questionnaires throughout the survey process.
e. The use of translated core questionnaires into the official languages to facilitate better training and to increase the accuracy of communications with respondents.
f. Implementation of different approaches to listing depending on the settlement type to guarantee complete sample coverage.
g. Upgrade of regional offices to accommodate additional field staff.

Questionnaire design and analysis:

a. Development of a short core questionnaire (59 questions) to reduce respondent fatigue and facilitate dissemination of results on a quarterly basis.
b. Development of automated tables for analysis to ensure minimised manual intervention in the production process.
c. Analysis of key results on a quarterly basis to monitor labour market changes on a regular basis.
d. Provision of link factors to enable historical continuity with the LFS.
e. In-depth analysis of quarterly results on an annual basis.

Data Processing:

The development of an end-to-end real time system that includes:

a. Real time management system (RTMS) for tracking and monitoring publicity and data collection.
b. Store management system to enable the tracking of questionnaires throughout the data processing phases.
c. Scanning and image retrieval system to enable the rapid verification of questionnaire content.
d. Editing and imputation system to ensure minimised human intervention in the production process.
e. Fully automated bulk coding system for industry and occupation to ensure consistency in the coding process.
f. Master Sample capturing system to provide an up-to-date electronic database from which unique samples for each household-based survey can be drawn.

13. Linking the old and the new labour market indicators

Introduction

As discussed earlier, both the LFS and the QLFS are instruments used for collecting labour market information. However, the redesign of all survey processes to accommodate the requirements of a quarterly design resulted in changes to the survey methodology, the sample frame, the frequency of data collection and data releases, the questionnaire, and the data capture and processing systems.

[15] Details of all quality improvements are provided in "Quality improvements in the QLFS by Workstream" at the Stats SA website @ www.statssa.gov.za/qlfs/index.asp.

It is not possible to disentangle the various factors throughout the survey process that contribute to the differences between the results of the LFS and those of the QLFS. However, the structure of the new core QLFS questionnaire is likely to have played a major role. The new training initiatives, the redesign of the Master Sample and the continuity gained by having permanent field-staff are also important contributors. In addition, the seamless process of data processing that requires minimal manual intervention and ensures that questionnaires are processed continuously such that results can be published four weeks after the end of the quarter are also likely to have made a substantial contribution.

Link factors

The revised labour market series for key indicators enable historical continuity with the LFS. These historical series are based on link factors computed on the basis of the overlap between the LFS conducted in March 2008 and the QLFS conducted in Jan–March 2008. A similar process will be undertaken to historically link the QLFS that will be conducted in July–September 2008 with the LFS to be conducted in September 2008.

Revision of historical data

The purpose of historical revision is to make the LFS estimates from 2002 to 2007 comparable to the QLFS data starting in 2008. "Comparable" means that measures of change that cross the 2007/2008 threshold are as valid as the revision technique can make them.

Conceptually, revising historical LFS data means that the revised LFS data for, say, March 2002 represent Stats SA's best estimate of what the QLFS would have shown had it been conducted in, say, Q1, 2002.

The historical revision is carried out in three phases. In the first phase a set of high-level variables is chosen (e.g., employed formal sector, employed formal sector, etc). For each of these variables, ratios of the QLFS estimate to the corresponding LFS estimate are calculated. One set of ratios is obtained from the Q1 overlap of the QLFS (Q1, 2008) and the LFS (March 2008). The second set is from the Q3 overlap of the QLFS (Q3, 2008) and the LFS (September 2008).

In the second phase, LFS records that contain blanks in the variables for which the ratios were calculated are purged. Then the historical LFS unit record (micro data) files are re-weighted. This weighting incorporates population estimates that reflect the population as of the middle of the quarter (as opposed to the last month in the quarter).

The corresponding vectors from the LFS going back to March 2002 are multiplied by the ratios described above to obtain historical control totals. These vectors are then adjusted so that they are consistent with the population control totals as well as with each other (e.g., the various disaggregations of employment add to the same total).

In the third phase the files are re-weighted again. This time, in addition to the conventional control totals, control totals reflecting the application of the 2008 ratios, are also included.

Historically revised LFS data are then obtained by running tabulations on these re-weighted files.