WORLD BANK HOUSEHOLD SURVEYS

FOR THE AFRICA MIGRATION PROJECT

SOUTH AFRICA MIGRATION PROJECT SUMMARY REPORT

February 2011

HSRC
Human Sciences
Research Council

centre for poverty employment and growth

Human Sciences Research Council

February 2011

Produced by:	Centre for Poverty Employment and Growth
Contact:	Dr Miriam Altman Executive Director, CPEG
E-mail:	MChitiga@hsrc.ac.za
Tel:	+27 12 302 2402

Contents

1 Introduction

The Human Sciences Research Council (HSRC) carried out the Migration and Remittances Survey in South Africa for the World Bank in collaboration with the African Development Bank. The primary mandate of the HSRC in this project was to come up with a migration database that includes both immigrants and emigrants. The specific activities included:

- A household survey with a view of producing a detailed demographic/economic database of immigrants, emigrants and non migrants

- The collation and preparation of a data set based on the survey

- The production of basic primary statistics for the analysis of migration and remittance behaviour in South Africa.

Like many other African countries, South Africa lacks reliable census or other data on migrants (immigrants and emigrants), and on flows of resources that accompanies movement of people. This is so because a large proportion of African immigrants are in the country undocumented. A special effort was therefore made to design a household survey that would cover sufficient numbers and proportions of immigrants, and still conform to the principles of probability sampling. The approach that was followed gives a representative picture of migration in 2 provinces, Limpopo and Gauteng, which should be reflective of migration behaviour and its impacts in South Africa.

This report details the processes and procedures followed in carrying out the Migration and Remittance Survey in South Africa and the production of the database. It describes concisely the choice of study areas, selection and training of fieldworkers, the survey instrument, pilot survey, sampling, data collection, data entry and cleaning, and some summary statistics.

The survey fieldwork began in earnest in mid-November 2009 and ended on the 23rd of December 2009, with all the questionnaires checked and received by the end of that day.

2 Study area

Following discussions with the World Bank team, the South Africa household migration survey was restricted to Limpopo and Gauteng provinces.

Figure 1 Map of South Africa's Nine provinces

It is instructive to note that there generally 5 migration corridors in South Africa. Three of these include the two provinces (Gauteng and Limpopo) covered by this survey. Limpopo is the main corridor for migration from African countries to the north of South

Africa while Gauteng is the main port of entry as it has the largest airport in Africa. Gauteng is a destination for internal and international migrants because it has three large metropolitan cities with a great economic activity and reputation for offering employment, accommodations and access to many different opportunities within a distance of 56 km. Limpopo and Gauteng provinces are also home to more than 33% of South Africans.

SOUTH AFRICA POPULATION BY PROVINCE 2010

Province	Population	% of total
Eastern Cape	6 743 800	13.5%
Free State	2 824 500	5.7%
Gauteng	11 191 700	22.4%
KwaZulu-Natal	10 645 400	21.3%
Limpopo	5 439 600	10.9%
Mpumalanga	3 617 600	7.2%
Northern Cape	1 103 900	2.2%
North West	3 200 900	6.4%

Western Cape	5 223 900	10.4%
TOTAL	49 991 300	100%

Source: Statistics South Africa

These two provinces therefore were expected to accommodate most African migrants in South Africa, co-existing with a large host population.

3 Recruitment and Selection of Fieldworkers

Experienced fieldworkers conversant in the languages spoken in the selected Enumerator Areas (EAs) were recruited for Gauteng and Limpopo provinces, since interviews were to be conducted in the languages preferred by respondents.[1] All fieldworkers recruited for this project had prior survey experience, having been involved in at least two prior projects for HSRC and/or CASE in the past twelve months.

The fieldworkers were recruited from CASE's updated national database of fieldworkers and worked under a coordinator and team of supervisors. This database includes fieldworkers, who are evaluated after every project, and is updated constantly (based on this evaluation) to ensure that it is as current as possible. The number of recruited fieldworkers was determined by the number of interviews to be conducted, given the timeframe of the project. Experience and proximity to sample areas were also key factors in the recruitment of fieldworkers.

A total of 50 fieldworkers, including supervisors, were directly involved in collecting data

---

[1] See the HSRC's Migration and Remittances Training and Pilot Survey Report

for this project. Additionally, drivers were recruited for transporting fieldwork teams to sample areas. The roles and responsibilities were as follows:

- The fieldwork Coordinator was responsible for overseeing the whole process of data collection, as well as ensuring high quality of the data collected. In addition, he was responsible for the proper deployment of the fieldworkers.

- Supervisors were responsible for the physical identification of the EAs with the households listed, the selection of the exact households to be interviewed within the selected EA, and for checking the quality and consistency of the information in the questionnaires. They were also tasked with negotiating access to EAs and notifying communities about the existence and purpose of the study.

- Fieldworkers were responsible for randomly selecting respondents within the selected households and for conducting face-to-face interviews. In addition, they were to notify the respondents that the HSRC might visit them later for data checking.

Data collection in affluent areas was carrying out by a white fieldwork teams. Whilst this should not have been desirable in terms of likely biases or interpretations in responses, it was important to ensure maximum co-operation of interviewees and easier access to these areas.

A list of all the fieldworkers who participated in data collection in Gauteng and Limpopo province is in the Fieldwork Report.

4 Training

The training of Fieldworkers was conducted on the 8th and 9th of October 2009 in a central facility. At that training, a number of issues related to the questionnaire and the framing of questions were raised. A detailed discussion of training procedures is found in the training manual. In short, the purpose of the training was to:

- Explain the background of the study,

- Provide and share a common understanding of what the study required,

- Train the fieldworkers on how to administer the questionnaire, including establishing rapport with the respondent,

- Describe roles and responsibilities of fieldworkers and supervisors,

- Describe the specific procedures to be followed during the data collection period,

- Explain practical details about the process of submission of completed questionnaires and data capture,

- Describe how to avoid misunderstandings and ensure good working relationships among fieldworkers and between fieldworkers and supervisors.

The training of fieldworkers was conducted by the CASE Research Project Manager, the Fieldwork Manager, a Fieldwork Coordinator, and HSRC representatives in Johannesburg, Booysens Hotel in Gauteng on 8 and 9 October 2009.

Several suggestions and changes to the questionnaire were proposed by the trainees and were addressed before commencement of the pilot survey.

5 Pilot Test of the Instrument

The pilot was conducted between 11 and 13 October 2009 in 8 different sites, 4 in Gauteng Province (GP) and 4 in Limpopo Province (LP). A total of 50 face-to-face interviews were conducted, 25 in each province.

The pilot was carried out without paying attention to randomness in the selection of the households. The enumeration areas in which the pilot was conducted were selected out of the ones that were not to be included in the main survey sample. Of the 50 piloted questionnaires, 43 were sent for data entry to check on the data entry program (the remaining 7 were found to not be of adequate quality when checked). Questions that needed changes in wording were modified, which led to the final questionnaire used.

6 Sample design and data constraints

To design a probability household survey, it is advisable to have appropriate information on the characteristics of the population for the smallest possible spatial entities. This is evidently absolutely necessary in situations where it is planned to conduct some kind of listing operation in the last stage sampling units. In many countries, including South Africa, these smallest area units are census enumeration areas (EAs). It is important to draw the sample of EAs in such a way that greater preference is given to selecting areas with higher proportions of immigrants since this will facilitate finding them and also thereby improve the efficiency of fieldwork. However, recent census data on immigration at this spatial level do not exist. The most recent population census undertaken by Statistics South Africa (Stats SA) was the census of October 2001, but data from this census are available only down to the so-called *sub-place* level, and hence for groups of EAs combined. The last census for which migration data are available at the EA level was the Census of 1996, which is too many years ago to reflect the current situation. Although a large household survey, called the Community Survey (CS) 2007, has since been undertaken, its migration data are available only at the local government level.[2] These three data sources are described briefly below.

---

[2] Although the 2007 migration-level data (i.e. on numbers or proportions of migrants) is available at the local government level, data on place of origin of the last move is available only at a provincial/country-region level.

The spatial coverage of other household surveys with migration-level components, such as the former six-monthly Labour Force Surveys (from 2000 to 2007) and the HSRC's own 2001–02 Migration Survey, is not suitable for the purposes of probability sampling that must be based on observed migration levels.

The 2007 CS data are the most recent and comprehensive, providing information on the numbers of (a) *lifetime inter-provincial migrants* (born in a different province), (b) *lifetime international migrants* (born in a different country), (c) *recent intra-provincial migrants* (who had moved from another place in the same province during the period 11 October 2001 to 10 October 2007), (d) *recent inter-provincial migrants* (who had moved from another province during the period 2001-2007), and (e) *recent international immigrants* (who had moved from another country during the period 2001-2007

Of the above five categories of migrants, only two could be identified (and only down to the sub-place (SP) level of disaggregation) in the 2001 Census: (a) *recent intra-provincial migrants* (who had migrated within the same South African province during the period 11 October 1996 to 10 October 2001) and (b) *recent inter-provincial migrants* (who had migrated from one South African province to another during the period 1996-2001).[3] Note there was no data at all on international migrants (viz., immigrants) from the 2001 census of population.

On the other hand, the 1996 census provided data for all five categories of migrants of interest here at the level of the EA, namely for: (a) *lifetime inter-provincial migrants* (born in a different province), (b) *lifetime international immigrants* (born in a different country), (c) *recent intra-provincial migrants* (who had moved from another place in the same province during the period 1 January 1992 to 10 October 1996), (d) *recent inter-provincial migrants* (who had moved from another province during the period 1992-1996), and (e) *recent international immigrants* (who had moved from another country during the period 1992-1996).

---

[3] The migrant categories for which 2001 data was <u>not</u> available were: (a) lifetime intra-provincial migrants, (b) lifetime inter-provincial migrants, (c) lifetime international migrants, and (d) recent (1996-2001) international migrants.

Therefore, migration data for South Africa are available for 2007 only at the level of local governments or municipalities from the 2007 CS; for smaller areas called "sub places" (SPs) only as recently as the 2001 census, and for the desired EAs only back so far as the Census of 1996. In sum, there was no single source that provided recent data on the five types of migrants of principal interest at the level of the Enumeration Area, which was the area for which data were needed to draw the sample since it was going to be necessary to identify migrant and non-migrant households in the sample areas in order to oversample those with migrants for interview (two-phase sampling—see Bilsborrow et al. 1997).

7 Migrant clusters

In an attempt to overcome the data limitations referred to above, it was necessary to adopt a novel approach to the design of the sample for the World Bank's household migration survey in South Africa, to identify EAs with a high probability of finding immigrants and those with a low probability. This required the combined use of the three sources of data described above.

The starting point was the CS 2007 survey, which provided data on migration at a local government level, classifying each local government cluster in terms of migration level, taking into account the types of migrants identified. The researchers then spatially zoomed in from these clusters to the so-called sub-places (SPs) from the 2001 Census to classifying SP clusters by migration level. Finally, the 1996 Census data were used to zoom in even further down to the EA level, using the 1996 census data on migration levels of various typed, to identify the final level of clusters for the survey, namely the spatially small EAs (each typically containing about 200 households, and hence amenable to the listing operation in the field).

A higher score or weight was attached to the 2007 Community Survey municipality-level (MN) data than to the Census 2001 sub-place (SP) data, which in turn was given a

greater weight than the 1996 enumerator area (EA) data. The latter was derived exclusively from the Census 1996 EA data, but has then been reallocated to the 2001 EAs proportional to geographical size. Although these weights are purely arbitrary since it was composed from different sources, they give an indication of the relevant importance attached to the different migrant categories. These weighted migrant proportions (secondary strata), therefore constituted the second level of clusters for sampling purposes.

In addition, a system of weighting or scoring the different persons by migrant type was applied to ensure that the likelihood of finding migrants would be optimised. As part of this procedure, *recent* migrants (who had migrated in the preceding five years) received a higher score than *lifetime* migrants (who had not migrated during the preceding five years). Similarly, a higher score was attached to *international* immigrants (both recent and lifetime, who had come to SA from abroad) than to *internal* migrants (who had only moved within SA's borders). A greater weight also applied to *inter-provincial (internal)* than to *intra-provincial* migrants (who only moved within the same South African province).

How the three data sources were combined to provide overall scores for EA can be briefly described. First, in each of the two provinces, all local government units were given migration scores according to the numbers or relative proportions of the population classified in the various categories of migrants (with non-migrants given a score of 1.0. Migrants were assigned higher scores according to their priority, with international migrants given higher scores than internal migrants and recent migrants' higher scores than lifetime migrants. Then within the local governments, sub-places were assigned scores assigned on the basis of inter vs. intra-provincial migrants using the 2001 census data. Each SP area in a local government was thus assigned a value which was the product of its local government score (the same for all SPs in the local government) and its own SP score. The third and final stage was to develop relative migration scores for all the EAs from the 1996 census by similarly weighting the

proportions of migrants (and non-migrants, assigned always 1.0) of each type. The final migration score for an EA is the product of its own EA score from 1996, the SP score of which it is a part (assigned to all the EAs within the SP), and the local government score from the 2007 survey.

Based on all the above principles the following set of weights or scores was developed:

1	EA 2001 data for 1996 EAs: Lifetime intra-provincial migrant proportion weight	1.000
2	EA 2001 data for 1996 EAs: Lifetime inter-provincial proportion weight	1.125
3	EA 2001 data for 1996 EAs: Lifetime international proportion weight	1.250
4	EA 2001 data for 1996 EAs: Recent (1992-1996) intra-provincial proportion weight	1.375
5	EA 2001 data for 1996 EAs: Recent (1992-1996) inter-provincial proportion weight	1.500
6	EA 2001 data for 1996 EAs: Recent (1992-1996) international proportion weight	1.625
7	SP data for 2001: Lifetime intra-provincial migrant proportion weight (NO DATA)	1.750
8	SP data for 2001: Lifetime inter-provincial proportion weight (NO DATA)	1.875
9	SP data for 2001: Lifetime international proportion weight (NO DATA)	2.000
10	SP data for 2001: Recent (1996-2001) intra-provincial proportion weight	2.125
11	SP data for 2001: Recent (1996-2001) inter-provincial proportion weight	2.250
12	SP data for 2001: Recent (1996-2001) international proportion weight (NO DATA)	2.375
13	MN data for 2007: Lifetime intra-provincial migrant proportion weight (NO DATA)	2.500
14	MN data for 2007: Lifetime inter-provincial proportion weight	2.625
15	MN data for 2007: Lifetime international proportion weight	2.750

16	MN data for 2007: Recent (2001-2007) intra-provincial proportion weight	2.875
17	MN data for 2007: Recent (2001-2007) inter-provincial proportion weight	3.000
18	MN data for 2007: Recent (2001-2007) international proportion weight	3.125

In sum, we multiplied the proportion of populations of each migrant type, or their incidence, by the appropriate final corresponding EA scores for persons of each type in the EA (based on multiplying the three weights together), to obtain the *overall* score for each EA. This takes into account the distribution of persons in the EA according to migration status in 1996, the SP score of the EA in 2001, and the local government score (in which the EA is located) from 2007. Finally, all EAs in each province were then classified into quartiles, prior to sampling from the quartiles (see column showing number of EAs in each quartile for each province in the table below).

8 Sampling within the clusters

From the EAs so classified, the sampling took the form of selecting EAs, i.e., primary sampling units (PSUs, which in this case are also Ultimate Sampling Units, since this is a single stage sample), according to their classification into quartiles. The proportions selected from each quartile are based on the range of EA-level scores, which are assumed to reflect weighted probabilities of finding desired migrants in each EA. To enhance the likelihood of finding migrants, much higher proportions of EAs were selected into the sample from the quartiles with the higher scores compared to the lower scores (disproportionate sampling). The decision on the most appropriate categorisations was informed by the observed migration levels in the two provinces of the study area during 2007, 2001 and 1996, analysed at the lowest spatial level for which migration data was available in each case. Because of the differences in their

characteristics it was decided that the provinces of Gauteng and Limpopo should each be regarded as an explicit stratum for sampling purposes. These two provinces therefore represented the *primary explicit strata*. It was decided to select an equal number of EAs from these two primary strata.

The migration-level categories referred to above were treated as *secondary explicit strata* to ensure optimal coverage of each in the sample. The distribution of migration levels was then used to draw EAs in such a way that greater preference could be given to areas with higher proportions of migrants in general, but especially immigrants (note the relative scores assigned to each type of person above). The proportion of EAs selected into the sample from the quartiles draws upon the relative mean weighted migrant scores (referred to as proportions) found below the table, but this is a coincidence and not necessary, as any disproportionate sampling of EAs from the quartiles could be done, since it would be rectified in the weighting at the end for the analysis. The statistical technique that was used to achieve this is called "controlled selection in probability sampling". An approach originally developed by Goodman and Kish (1950)[4], which has since been widely applied and tested (see, for example, http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1065737&blobtype=pdf, www.amstat.org/sections/srms/proceedings/papers/1992_065.pdf, http://adb.sagepub.com/cgi/content/abstract/15/4/397). For a description of the

---

[4] Goodman, R. & Kish, L. 1950. Controlled selection—A technique in probability selection. *Journal of the American Statistical Association*, Vol. 45, pp. 350–372.

technique or its application, see for example Kish (1965)[5], US Bureau of the Census (1975)[6], Heeringa and Hess (1983)[7], and

Stoker (1983, 1985)[8]. This technique has also been extended to a so-called two-dimensional optimal controlled selection method by Tiwari and Nigam (1998)[9].

The resultant proportions of migrants then led to the following *proportional allocation of sampled EAs* (Quartile 1: 5 per cent (instead of 25% as in an equal distribution),

---

[5] Kish, L. 1965. *Survey methods*. New York: Wiley.

[6] U.S. Bureau of the Census. 1978. 1978. *The Current Population Survey: Design and Methodology*. Technical Paper No. 40. Washington, DC: U.S. Government Printing Office.

[7] Heeringa S.G. & Hess, I. 1983. *More on controlled selection*. Section on Survey Research Methods, 1983 Proceedings. American Statistical Association.

[8] Stoker, D.J. 1983. *Steekproefneming in die praktyk [Sampling in practice]*. Geleentheidspublikasie Nr. 4 [Occasional Paper No. 4]. Pretoria: RGN [HSRC]; Stoker, D.J. 1985. *Die tegniek van beheerde seleksie in steekproefneming [The technique of controlled selection in sampling]*. Geleentheidspublikasie Nr. 27 [Occasional Paper No. 27]. Pretoria: RGN [HSRC].

[9] Tiwari, N. & Nigam, A.K. 1998. "On two-dimensional optimal controlled selection". *Journal of Statistical Planning and Inference*, 69(1):89-100.

<u>Quartile 2</u>: 15 per cent (instead of 25%), <u>Quartile 3</u>: 30 per cent (instead of 25%), and <u>Quartile 4</u>: 50 per cent (instead of 25%).[10]

It was agreed that a sample size of at least 2 000 households would be required to elicit the required information. It was agreed further that only six (6) households would be selected in the final level of clusters, i.e., the PSUs or the EAs, to reduce clustering effects, viz., and the possible impact of spatial interdependence of survey responses. This gave a required total of 334 EAs (2 000 / 6 = 333.33) to be selected. The final distribution of EAs in the sample was therefore as indicated in the right-most column of the table below:

Province	Quartile	Total EAs	Sample EAs

---

[10] The quartile-based distribution of EAs in the universe of all EAs in the study area was found to be as follows:

<u>Quartile 4</u> (Maximum, 100%):Weighted migrant proportion = 0.374378 (37.438%)

<u>Quartile 3</u> (75%):   Weighted migrant proportion = 0.219704 (21.970%)

<u>Quartile 2</u> (Median, 50%):   Weighted migrant proportion = 0.168519 (16.852%)

<u>Quartile 1</u> (25%):   Weighted migrant proportion = 0.056408 (5.6408%)

<u>Minimum</u> (0%):   Weighted migrant proportion = 0.014158 (1.4158%)

   (a) <u>Quartile 1 (0-25%)</u>: 1,4158% – 5,6408% (4 782 EAs): Instead of an equal proportion of about 83.5 *a sample of 17 EAs was drawn -- all from 4 782 Limpopo EAs* – giving a disproportional number of selected EAs (17 instead of 83.5)
   (b) <u>Quartile 2 (25%-50%)</u>: 5,6409% – 16,8519% (4 782 EAs): Instead of about 83.5, *a sample of 50 EAs was drawn* **--** *3 from 2 331 Gauteng EAs and 47 from 2 451 Limpopo EAs*
   (c) <u>Quartile 3 (50%-75%)</u>: 16,8520% – 21,9704% (4 782 EAs): Instead of about 83.5, *a sample of 100 EAs was drawn -- 7 from 4 590 Gauteng EAs and 93 from 192 Limpopo EAs*
   (d) <u>Quartile 4 (75%-100%)</u>: 21,9705% – 37,4378% (4 781 EAs): Instead of about 83.5, *a sample of 167 EAs was drawn -- 157 from 4 726 Gauteng EAs and 10 from 55 Limpopo EAs*

	1	0	0
	2	2 331	3
Gauteng	3	4 590	7
	4	4 726	157
	Total	11 647	167
	1	4 782	17
	2	2 451	47
Limpopo	3	192	93
	4	55	10
	Total	7 480	167
	1	4 782	17
	2	4 782	50
Total	3	4 782	100
	4	4 781	167
	Total	19 127	334

An explicit, disproportional stratification of provinces (primary strata) and incidence migrant proportions (secondary strata) was therefore used as a basis for the selection of EAs. The disproportionate distribution of these selected EAs was to be rectified afterwards through the use of EA weights during all data analyses.

9 Sampling within final level clusters (EAs)

Within each sample EA selected following the procedures above, an approximate listing of dwellings was undertaken by the survey team, and updated maps (showing streets/roads, potentially eligible dwellings, and other easily identifiable features for orientation purposes) were produced.

Within the EA a starting point was chosen and the household was interviewed. Observing the sampling interval the second household was interviewed only if there was a migrant within the household[11]. Following the agreed methodology a maximum of 20 households were visited in each EA. When there were more than one household at a particular visiting point, only one was randomly selected. In the case of a block of flats, townhouse complex or retirement village, it was important to regard every occupied flat/unit as a potential visiting point of the interval. In the case of single-sex workers' hostels, each room or dormitory constituted a visiting point and every occupied bed in a selected room/dormitory represented a (single-person) household.

The sampling process was according to the following plan.

- Enumerator Areas were randomly selected using the approach outlined earlier

- Maps of the selected EAs were obtained the from Statistics South Africa (STATS SA),

- For each EA, the fieldwork supervisor/team identified the physical boundaries from the map and ensured that the map and the physical location were congruent,

- The fieldwork supervisor/team counted the number of houses/dwellings within each EA. Call this **Nile,**

---

[11] Please see the full survey methodology report.

- 20 households per EA were visited, so the sampling interval was calculated as *Nile*/20. For example, if *Nile*=200 houses/stands, the sampling interval was *Nile*=200/20=10. This means that every 10th house/stand was visited,

- The supervisor identified a random starting point, such as a school, a shop, a library, or some similar public point. If none could be identified then one dwelling was identified,

- From this randomly selected starting point, every 10th house/dwelling was visited. In other words, systematic sampling was used to identify the required sample,

- The actual household interviewed was selected following the procedure below:

   o interviewer approached the first household (call that Household #1) and completed the interview irrespective of whether there are migrants in the household,

   o Households #2 to #15 were interviewed only if there was at least one international migrant in the household,

   o Households #16 to #20 were interviewed irrespective of whether there were migrants in the household,

   o If there were migrants in the first six households visited, the interviewer stopped and did not visit any more of Households. The other households were just noted,

   o This meant that for each EA, 20 households were in the sample frame, but a maximum of six would be interviewed,

   o If the dwelling unit replacements were required, e.g., if some households refused to be interviewed, then interviewers selected the next house to the right, followed if necessary by the next house to the left, and so on.

In addition, fieldworkers also had to fill-in a recording sheet. The purpose of the recording sheet was to make sure that fieldwork teams recorded all the household they visited, recording addresses as well as the status of the household, i.e. whether the household had an international migrant or not.

10 Data collection during the main survey

10.1 Refresher training for fieldwork teams

The commencement of data collection was delayed for approximately three weeks due to refinement of the questionnaire based on the training feedback and the pilot survey. Gauteng refresher training session was held on 10 November 2009 while the Limpopo one was held on 12 November 2009.

10.2 Deployment of Fieldwork Teams

Fieldworkers and supervisors involved in the project were grouped into fieldwork teams, with each team consisting of one supervisor and four to five fieldworkers. Each team was allocated a particular number of Enumerator Areas. Data collection started on 13 and 14 November 2009 in Gauteng and Limpopo provinces, respectively, and ended on the 23rd of December 2009.

A feedback session was held by each fieldwork team after each interviewer in the team had conducted approximately three interviews. The challenges that were most prevalent were the usual encountered in surveys, and included the following:

- Potential respondents were not at home during the first visits so it was necessary to make many appointments and call backs.
- Negotiating access in high affluent areas was challenging.
- Administering the questionnaire took longer than expected (more than an hour) due to its length and complexity.
- Respondents had to be constantly reminded of the difference between migrants,

non-migrants and return migrants.

· Sensitivity to the migration issue led many potential household respondents to deny either that they were international migrants or that they had within their household an international migrant.

10.3   Supervision of Fieldwork

All fieldworkers in the project were directly supervised throughout the fieldwork, i.e., supervisors were in the field at all times during data collection. Completed questionnaires were checked by supervisors immediately after the interview--whether all the relevant questions were answered/coded and for consistency. Supervisors also conducted random callbacks on the completed questionnaires:

· checking if the interview had actually taken place
· checking whether the interview was conducted with the respondent recorded on the questionnaire
· checking whether the people listed on the household grid were correctly identified as members or non-members of that household
· verifying whether the migration status of household members was correctly listed on the questionnaire

The Fieldwork Coordinator was responsible for quality control during the data collection phase of the project. This included checking a percentage of questionnaires from every fieldwork team, and also conducting some callbacks. The Fieldwork Coordinator checked for data consistency and whether routine instructions were followed in the administering of the questionnaire. The callbacks on completed interviews were to verify whether the interview was conducted with the recorded respondent and whether the instructions in randomly selecting the household and the respondent were followed. Very few errors were found. See also below

10.4   Accessing Targeted Areas

Fieldwork teams did not encounter major challenges in accessing sample areas and households. However, there were a number of refusals by households, especially in high affluent areas in both Gauteng and Limpopo provinces. Lack of interest in the topic was the most common reason given by sample respondents for refusing to participate in the study. To minimize this, fieldwork teams reminded potential respondents of the importance of the study and of them participating before they could record those cases as refusals.

10.5 Substitution of Enumerator Areas

Only 1 EA substitution was necessary during data collection based on refusals. Three other EAs that did not have dwellings were replaced. The EAs involved and the reasons for substitution were:

1. EA 77409214 in Ormonde (affluent area) was substituted for by a similar, nearby EA because most potential respondents refused to participate, denying access.
2. EA 77409008 in Anchorville was substituted because it is an industrial area.
3. EA 91400011 and EA 91400013 in Swartklip SP were substituted because they are in a mining area and no people reside within these two EAs.

10.6 Sample Realisation

A total of 2026 interviews were conducted in 340 EAs. This represented 26 more households than the target of 2000 and 6 more EAs than the target of 334. This was done by randomly carrying out extra interviews in order to ensure that the final sample would not fall short of the target. Careful note was made of where these extra interviews were carried out, so this information could be used in calculating EA weights.

11 Challenging issues

The most prominent challenge was the timing of the survey. In the recent past South

has experienced 2 major xenophobic disturbances. The survey was conducted the period between the 2 episodes. Coupled with a large incidence of illegal migration, this meant that respondents were very wary of providing information (that they might have felt would expose their identity) and it took a lot more effort to solicit information than would have been the case in 'normal' circumstances. Among the challenges raised by the fieldwork teams during the data collection phase of the project were:

- o fieldwork teams were worried that they were not picking up enough migrant respondents in most EAs;
- o some sampled respondents were skeptical of participating in the study because they thought the study was targeting undocumented migrants;
- o many respondents refused to respond to the "Household Use of Financial Services" section;
- o some international migrants refused to participate even though the purpose was explained carefully again; and
- o complaints about the length of the questionnaire were raised by a substantial number of respondents.

12 Data Entry and Cleaning

Data entry was carried out while interviews were continuing in the field, but proceeded slowly so that most was done after the fieldwork had been completed. A double data entry procedure was used in which a questionnaire was entered twice by different persons to improve accuracy.

Challenges encountered included:
- o Many questions in the questionnaire allowed multiple responses but were not marked as such;
- o Some questions included two parts but were not marked as such; and
- o Some households had more members than the space allocated.

Thus, data processors had to alter the data entry template to accommodate these situations, causing delays in data processing, e.g. for question 1.2, the data entry template was initially designed for nine household members, but later was adapted to allow for more than nine..

13 Summary of Findings[12]

The data set was received from data entry persons on 27 January 2010 and immediately data cleaning started. A sample size of 2 026 households was realized in 340 enumeration areas (EA's). Due to lack of valid population registers of households in each EA, the listing process found that the sample included 2 EAs that had fewer than 20 households (EA numbers 77303021 (with 14 households) and 91200319 (with 7 households)).

It was found that in some dwelling places, more than one household was eligible to be interviewed. In these instances, interviewers randomly chose the respondent household, but noted the number of eligible households. This information was then used for weighting purposes.

Table 1 shows the sample realization by province.

Province	Number of households visited	Percent
Gauteng	1 022	50
Limpopo	1 004	50

---

[12] More detailed findings can be accessed from the WBSA Migration Unweighted file attached.

Total	**2 026**	**100**

Table 1: Number of households visited by Province

The household questionnaire was made up of grids. Thus questions were asked about each household member and the data for each household and its members was entered as a single observation. However, this format was not conducive for analysis so the format of the data was modified so that each household member was listed as a standalone observation. Thus, 7,768 household members files were created. However, in each household, there was a key respondent who was asked all the questions about the other household members.

Person number in the household	Gauteng	Limpopo	Total
1	1 022	1 004	2 026
2	862	896	1 758
3	635	732	1 367
4	426	581	1 007
5	257	402	659
6	148	248	396
7	87	156	243
8	48	99	147

9	33	57	90
10	16	29	45
11	4	10	14
12	4	5	9
13	1	4	5
14	0	2	2
Total	**3 543**	**4 225**	**7 768**

Table 2: Household members by person number and Province

A total of 368 household members were reported to have left at some point in the past to live in another country or some other place in this country, whether urban or rural, for at least 6 months, without returning to the household, as shown in Table 3 (Q6) below.

Province	Males	Females	Missing	Total
Gauteng	53	37	4	94
Limpopo	146	128	0	274
Total	**199**	**165**	**4**	**368**

Table 3: Number of emigrants by Province

121 of these former household members were reported to have sent money to their former households, as shown in Table 4 below.

Province	Males	Females	Total
Gauteng	30	33	63
Limpopo	27	31	58
Total	**57**	**64**	**121**

Table 4: Former household members who sent money to household in last 12 months

There were a total of 72 return migrants in the data, as shown below.

Province	Number of return migrants	Percent
Gauteng	34	47

Limpopo	38	53
Total	**72**	**100**

Table 5: Number of return migrants by Province

A total of 1 268 immigrants were recorded, as presented in Table 6 below.

Province	Number of immigrants	Percent
Gauteng	826	65
Limpopo	442	35
Total	**1268**	**100**

Table 6: Number of immigrants by Province

Other findings include:

Total number of households in the sample   2026

Total number of households members in the sample   7768

Total number of immigrant households   330

Total number of emigrant households   246

Number of non-migrants   4874

Number of internal migrants   1850

Household Characteristics

Household Head's place of birth

		Number	Percent
South Africa	Urban	802	39.6
	Rural	886	43.7
Foreign born		330	16.3
Missing		9	0.4
Total		2,027[13]	100

Lifetime migration

Lifetime international migration	Frequency	Percent
Not an international migrant	7026	90.5

---

[13] In one household, the respondent was insistent that the household was co-headed by 2 individuals (the husband and the wife), hence the number of household heads is larger than the number of respondent households.

Lifetime international migrant from SSA	700	9.0
Lifetime international migrant from outside SSA	42	0.5
Total	7768	100.0

14 Problems Encountered

1. Conducting a survey of this nature in SA is difficult in the best of times, and this proved to be no exaggeration. Access was particularly a problem in affluent areas and in xenophobia-sensitive places. Just mentioning the word migrant led to suspicion due to the fact that there exists a large illegal migration in SA. Some respondents recommended that there have been a televised or print media advert preceding the fieldwork to generate confidence that the information gathered would not be used either against the respondent or their immediate neighbours.

2. Related to 1. above is the fact that there are a relatively large number of illegal migrants in SA, some of whom claim to not be migrants from abroad.

3. The timing of the fieldwork was a problem: December is not a good month to conduct such work. Many people were already on holiday, and some, including migrants (both internal and international) had already left their residences. Those that were found were often irritated by the timing. Some interviews were conducted quite literally on Christmas Eve. This can be likened to conducting a survey on the eve of Thanksgiving!

4. Some respondents felt that some questions were too exact for their liking. This affected questions that involved personal information, such as salary levels or expenditure patterns. A suggestion to use range categories seemed to be what most respondents who had a problem with this seemed to prefer.

5. Due to delays in the finalisation of the questionnaire, there was a long time lag between the initial training and fieldwork, which necessitated re-training of the fieldworkers.

6. Although respondents were willing to answer most questions, there was a larger than expected tendency to refuse to answer certain questions. Thus in some cases, respondents would initially refuse to even give names of household members but later in the interview they would provide data about those same members. In other households, a list of members would be given but not much more information was provided after that. How much this was caused by the timing issue above or the xenophobic matters mentioned in 1. above is not clear.

ATTACHMENTS

- Clean Data Set