

N.i.D.S.
NATIONAL INCOME DYNAMICS STUDY

Methodology: Report on NIDS Wave 1

Technical Paper no. 1

Murray Leibbrandt
Southern African Labour & Development Research Unit
Murray.Leibbrandt@uct.ac.za

Ingrid Woolard
Southern African Labour & Development Research Unit
Ingrid.Woolard@uct.ac.za

Louise de Villiers
Southern African Labour & Development Research Unit
Louise.DeVilliers@uct.ac.za

July 2009

Contents

1. Introduction.....	1
2. Questionnaire design	3
2.1 Pre-test	5
2.2 Finalisation of the questionnaires.....	7
2.3 Translation.....	7
2.4 Ethics Approval.....	8
3. Sample Design.....	9
3.1 Sampling frame	9
3.2 Sample of dwelling units	9
3.3 Individual respondent selection	11
4. Wave 1 Data collection	13
4.1 Publicity and pre-notification of selected households.....	13
4.2 Wave 1 Fieldworker Training.....	13
4.3 Wave 1 Fieldwork	15
4.3.1 Phase 1	15
4.3.2 Phase 2	16
5. Processes in NIDS Office	18
5.1 Logging system/ IT Systems.....	18
5.1.1 Quality control	18
5.1.2 Call backs.....	19
5.2 Data capture	20
5.3 Data storage.....	20
5.3.1 Physical storage.....	20
5.3.2 Electronic storage.....	20
6. Household and Individual response rates.....	22
7. Weights	28
8. Data Cleaning and Data preparation.....	29
8.1 Preserving anonymity in the data	29

8.2 Integration of community level data.....	30
8.3 Derived variables and imputed values.....	30
9. Panel maintenance.....	31
9.1 Panel maintenance initiatives.....	31

1. Introduction

In 2006, the South African Presidency embarked on an intensive effort to track changes in the well-being of South Africans by closely following about 28 000 people - young and old, rich and poor - over a period of years. The National Income Dynamics Study (NIDS) will be the first national panel study to document the dynamic structure of a sample of household members in South Africa and changes in their incomes, expenditures, assets, access to services, education, health and other dimensions of well-being. A key feature of the panel study is its ability to follow people as they move out of their original 7 305 households. In doing this, the movement of household members as they leave and/or return to the household or set up their own households will be adequately captured in subsequent waves.

The first “baseline” wave of NIDS was conducted by the Southern Africa Labour and Development Research Unit (SALDRU) based at the University of Cape Town’s School of Economics. The first wave of fieldwork commenced in February 2008, with data and report release in July 2009. The design of NIDS envisaged data collection every two years.

Elsewhere in the world such surveys have been invaluable in promoting understanding of who is making progress in a society and who is not and, importantly, what factors are driving these dynamics. In addition, panel data is invaluable for the purposes of evaluating and monitoring the efficacy of social policies and programmes. This is because the panel allows researchers and policy analysts to see how households and individuals are impacted when they become eligible for these programmes. Given all this, it was crucial to ensure that the base wave of NIDS provided a high quality, representative sample of South Africa in 2008 upon which to build the unfolding panel.

The 2008 NIDS questionnaires attempted to gather information on all members of the household; including those that were resident and those that were non-resident at the time of the interview. Those that were resident provide the base sample of individuals who will remain in the NIDS sample over time. Information about non-resident members is essential in understanding the household and family support systems that individuals have around them at the time of the interview.

In each household we administered a household questionnaire as well as individual questionnaires for each adult and each child in the household. In addition a range of other information, we took measurements of weight and height for all individuals and waist and blood pressure measurements for all adults in the household. Finally, we also asked individuals aged

12-59 to take a short numeracy (maths) test. Copies of these questionnaires are available on the NIDS website at <http://www.nids.uct.ac.za/nidsq.html>

This document records the processes during Wave 1 of NIDS. Specifically:

- designing of the questionnaires,
- drawing of the sample,
- data collection – including a pre-test phase and the formal fieldwork phase,
- in-office quality control and data capture,
- resultant response rates and refusal rates,
- derivation of sampling weights for use in the analysis of NIDS,
- ongoing data cleaning and data preparation for the public release of the Wave 1 data, and
- panel maintenance system that has been set up to prepare the way for the Wave 2 fieldwork.

This is an overview document and, in parts, it summarises material from more detailed technical papers.

2. Questionnaire design

From the outset, the purpose of NIDS was to shed light on a number of research questions, namely:

- Wealth creation in terms of income and expenditure dynamics and asset endowments;
- Demographic dynamics as it relates to household composition and migration;
- Social heritage, including education and employment dynamics, the impact of life events (including positive and negative shocks), social capital and intergenerational developments; and
- Access to cash transfers and social services.

These research questions must be viewed against the backdrop of the key concerns highlighted for the NIDS, which include measuring poverty in South Africa, the need to measure that all South Africans benefit from economic growth and social stability, and the concern that numbers of South Africans might end up being ‘socially excluded’, left behind or trapped in a ‘second economy’ where they are unable to benefit from economic opportunity. NIDS should shed light on the circumstances in which South Africans find themselves, how these conditions impact on their ability to improve their well-being and how government policy can play a positive role in these livelihood strategies.

These research questions were the starting point for questionnaire design because, as a minimum requirement, the questionnaires had to ensure that information was gathered that would allow for the measurement of these unfolding dynamics. Eight background papers were commissioned to inform the design of the NIDS questionnaires.¹ The authors and the papers are listed in Table 1 below.

¹ These papers are available on the NIDS website (www.nids.uct.ac.za).

Table 1: Topics & authors of background papers for NIDS questionnaire design workshop

Topic	Consultant(s)
Agriculture	Beatrice Conradie
Shocks, assets and credit	Malcom Keswell and Justine Burns
Demography	Tom Moultrie
Intra-household, social networks and social cohesion	Malcom Keswell and Justine Burns
Health	Anne Case & Cally Ardington
Education	Cally Ardington & David Lam
Migration and remittances	Daniela Casale & Dori Posel
Quantitative-qualitative integration	Andries du Toit

Based on the above papers and a thorough investigation of other panel survey questionnaires, four instruments were developed over the five months February 2007 to June 2007. These included household, adult, child and proxy questionnaires. The questionnaires were designed to take no longer than an hour for the household questionnaire, 45 minutes for the adult questionnaire and between 15 and 20 minutes for the child and proxy questionnaires. Electronic copies of the final questionnaires and the translations can be found at <http://www.nids.uct.ac.za/nidsq.html>.

Field-testing the draft questionnaire was a two step process. The first stage – “piloting” - involved trying out specific modules of the questionnaire on a small number of households to obtain an approximate idea of how well the draft questionnaire worked and how long the modules took to complete. Every section of the questionnaire was piloted by the end of July 2007. Subject area experts participated in some of these pilots. For example, the agriculture module of the household questionnaire was piloted in the rural Eastern Cape under the supervision of David Neves of the Programme for Land and Agrarian Studies at the University of the Western Cape.

Under contract from NIDS a numeracy module was drafted by Professor Patrick Griffin of the Assessment Research Centre at the University of Melbourne. This module was piloted separately and was not included in the pre-test (discussed below). The numeracy module consisted of a set of five integrated tests with each respondent being channelled into the test that matched their level of maths attainment at school or university. This module was piloted in October 2007 at several schools and on university students at UCT. The data (from over 500 respondents) was sent back to Prof Griffin for fine-tuning of the tests. The fifth test (designed for respondents with university level mathematics) was omitted after negative feedback about

the difficulty of the test and a corresponding lack of willingness on the part of respondents to attempt the test.

2.1 Pre-test

In April 2007 the tender to undertake the fieldwork for the base wave of NIDS was awarded to a field work organisation, Development Research Africa (DRA) in consortium with Take Note Trading.

Eight enumerator areas in Gauteng and KwaZulu-Natal were selected by DRA for the pre-test. These areas included all geographical types (urban formal, urban informal, tribal, rural formal) and at least one area in which we would have expected each South Africa's four race groups to be predominant.

Pre-test training was conducted in Johannesburg from 3 to 7 September 2007 and in Durban from 10 to 14 September 2007. Three field managers and 10 fieldworkers were trained in Johannesburg and 4 field managers and 16 fieldworkers were trained in Durban. Training took place over five days and included members from SALDRU and the fieldwork organisation. All households approached were given an incentive irrespective of participation.

Debriefing sessions with field managers and a sample of fieldworkers were undertaken in October 2007. This proved to be an enormously useful engagement. The NIDS team benefited a great deal from hearing directly from the field teams. The fieldworkers and field managers also appreciated the opportunity to feed back into the design of the questionnaire and to highlight issues that had caused frustration in the field.

The key issues that were highlighted during the debriefings were:

- The time taken to administer the questionnaires was not outside of the critical ranges that the NIDS team had set. However, the length of the questionnaires still seemed to cause respondent fatigue in higher income areas;
- The field teams had not anticipated how difficult it would be to locate every member of the household and had not planned as well as they should have (e.g. by visiting every household on the first two days of the field period to make appointments for later in the week);
- Respondents had become very suspicious when asked about their ownership of jewellery and firearms;
- The field teams felt that photographing the dwelling units had made respondents unnecessarily suspicious;

- Fieldworkers needed additional training on how to communicate that respondents' information would be kept confidential;
- Field managers felt that they had not been able to give enough time to quality control in the field because they were often conducting interviews themselves; and
- Field teams were working in different ways with many teams making use of two interviewers for a single respondent, which was inefficient.

A total of 159 households were interviewed at 192 dwelling units. Response rates during the pre-test were very high. The details are reported in Table 2 below.

Table 2: A breakdown of pre-test sampling and refusals

EA	Number hhs visited	Refusals	No-one at home	Total hhs completed
Gauteng Urban Informal	24	0	2	22
Gauteng Urban Formal (Predominantly African)	24	2	0	22
Gauteng Urban Formal (Predominantly Coloured)	24	3	0	21
North West Rural Formal	24	3	0	21
KwaZulu-Natal (KZN) Tribal	24	1	0	23
KZN Tribal	24	0	1	23
KZN Urban Formal (Predominantly Indian/Asian)	29	12	2	15
KZN Urban Formal (Predominantly White)	24	12	0	12

Note: The reason for 29 households being attempted in one PSU was that the study methodology requires the field team to interview all households residing at the 24 selected dwelling units and a large number of dwelling units contained more than one household.

The pre-test highlighted the fact that insufficient quality control was being done in the field. This issue was discussed with the fieldwork organisation and the parties agreed on what would be expected during the main study. In addition, supplementary funding was allocated to regional quality control in the budget for the main study.

NIDS quality control and office coding (of geo-referenced variables and occupation and sector variables) ran smoothly. Post-graduate students were hired to do these tasks and this proved to work well. The students were sufficiently interested in the material to be motivated to cross-check information across different questionnaires from the same household. Quality control took longer than expected – and thus cost more than anticipated – and new assumptions about time and cost could thus be built into the plans for the main study.

Students were also hired to perform data capture. This proved less successful than hoped. The students worked more slowly and less accurately than expected. In addition, students were often unwilling to sign up for long shifts which created additional problems given the nature of the work. It was agreed that for the main study NIDS would hire a core team of full-time data capturers that would be supplemented by part-time students.

2.2 Finalisation of the questionnaires

Changes to the questionnaire were made in the light of the findings of the pre-test and feedback from members of the Steering Committee. The major changes made were a substantial shortening of the agricultural module and the removal of the modules on remittances sent and received from the household questionnaire as this information was captured in the adult questionnaire.

The labour market module in the pre-test version of the adult questionnaire would not generate a nuanced picture of the roles of the labour market in South African livelihood activities but the price of replicating the new Stats SA labour force survey was too high. In response to this, a completely redesigned labour market module was developed and piloted. Furthermore, there was concern that the mental health module, which was based on the General Health Questionnaire- 12 Items (GHQ12), was too tightly focused on psychiatric categories rather than measuring general mental state. The latter seemed more appropriate for the NIDS as a correlate of livelihood strategies and behaviour. We sought the advice of selected members of our technical committee and replaced the GHQ12 module with CES-D 10 (Centre for Epidemiologic Studies Short Depression Scale). This module has been used in many population studies and was developed as a screen to identify general depression.

Finally, the questions on firearms and jewellery were removed from the questionnaire as these had heightened suspicion during the pre-test.

2.3 Translation

For the pre-test, it was agreed that we would translate the questionnaire into languages spoken by at least 10% of the population in either Gauteng or KwaZulu-Natal. Upon consultation with the fieldwork organisation, it was decided that the questionnaires be translated into Afrikaans, isiZulu and SeSotho. The fieldwork organisation provided the translation service for the pre-test.

NIDS took the view that it would be preferable for a professional company to undertake the translation of the questionnaires and documentation for the main study. This included the four

questionnaires, the four numeracy tests, the health information sheet, the consent form, access letter and informational pamphlet. In total then, we required 12 different documents in each of the 11 languages.

For the publicity campaign, information letters and pamphlets in the dominant language(s) of each PSU were distributed. For the fieldwork itself the strategy was to ensure that questionnaires could be administered to individuals in their language of preference, while the fieldworker completed an English set of questionnaires. The cover sheet of each questionnaire recorded the language that was used in the interview.

2.4 Ethics Approval

The Ethics Committee of the University of Cape Town operates through a set of faculty committees who then report on their work to the Senate Ethics Committee. In line with this, the Commerce Faculty Ethics Committee is the primary interface for ethical approval for NIDS. Before the pre-test, this committee assessed the informed consent forms and procedures, the pre-test questionnaires, proposed fieldwork procedures and confidentiality plans within the NIDS office. Provisional approval for the pre-test was granted. This approval allowed the pre-test to proceed.

Following the pre-test a cover letter, a completed faculty application document along with the final questionnaires and all accompanying NIDS documentation were submitted to the Commerce Faculty Ethics Committee in support of an application for formal ethical approval for Wave 1 of NIDS. The committee was satisfied that all human subject issues had been appropriately addressed and therefore granted ethical approval for the Wave 1 questionnaire and study protocols.

3. Sample Design

3.1 Sampling frame

A stratified, two-stage cluster sample design was employed in sampling the households to be included in the base wave. In the first stage, 400 Primary Sampling Units (PSUs)² were selected from Stats SA's 2003 Master Sample of 3000 PSUs. This Master Sample was the sample used by Stats SA for its Labour Force Surveys and General Household Surveys between 2004 and 2007 and for the 2005/06 Income and Expenditure Survey. Each of these surveys was conducted on non-overlapping samples drawn within each PSU.

The target population for NIDS was private households in all nine provinces of South Africa and residents in workers' hostels, convents and monasteries. The frame excludes other collective living quarters such as students' hostels, old age homes, hospitals, prisons and military barracks.

The sample of PSUs for NIDS is a subset of the Master Sample. The explicit strata in the Master Sample are the 53 district councils (DCs). The sample was proportionally allocated to the strata based on the Master Sample DC PSU allocation and 400 PSUs were randomly selected within strata. It should be noted that the sample was not designed to be representative at provincial level, implying that analysis of the results at province level is not recommended.

3.2 Sample of dwelling units

At the time that the Master Sample was compiled, 8 non-overlapping samples of dwelling units were systematically drawn within each PSU. Each of these samples is called a "cluster" by Stats SA. These clusters were then allocated to the various household surveys that were conducted by Stats SA between 2004 and 2007. However, two clusters in each PSU were never used by Stats SA and these were allocated to NIDS.

Tables 3 and 4 compare the distribution of the PSUs that are in the NIDS sample against those that are in the Master Sample. It can be seen from both tables that the sample spread per

² A PSU is defined as a geographical area that consists of at least one Enumeration Area (EA) or several EAs from the 2001 Census, when the originally selected EA was found to have less than 74 households. In some cases it has been necessary to add EAs to the original EA to meet the requirement of a minimum of 74 households per PSU. The EA or EAs added to the original EA has to be of the same settlement type as the original EA. An EA is the smallest portion of land that the country was demarcated into for the purpose of Census enumeration.

province and per geography type is quite similar between the two samples. Thus the selected sample was deemed satisfactory in this regard.

Table 3: Comparing the distribution of the PSUs per province

Province	NIDS Sample		Master Sample	
	Frequency	Percent	Frequency	Percent
Western Cape	52	13.0	385	12.8
Eastern cape	53	13.3	396	13.2
Northern Cape	27	6.8	207	6.9
Free State	31	7.8	245	8.2
KwaZulu-Natal	86	21.5	640	21.3
North West	35	8.8	259	8.6
Gauteng	48	12.0	353	11.8
Mpumalanga	30	7.5	233	7.8
Limpopo	38	9.5	282	9.4
RSA	400	100	3000	100

Table 4: Comparing the distribution of the PSUs per geography type

Geography Type	NIDS Sample		Master Sample	
	Frequency	Percent	Frequency	Percent
Farms	49	12.3	310	10.3
Tribal	131	32.8	957	31.9
Urban Formal	194	48.5	1535	51.2
Urban Informal	26	6.5	198	6.6
RSA	400	100	3000	100

It was sometimes necessary to re-list a PSU when the situation on the ground had drastically changed to an extent that the information recorded on the listing books no longer reflected the situation on the ground. In these cases, the PSU was re-listed and a new sample of dwelling units selected. However, the downside of re-listing a PSU is that the chance of sample overlap with dwelling units that are in other surveys is increased. The extent of this overlap cannot be quantified as the lists are no longer comparable. There is anecdotal evidence that sample overlap might have occurred in some PSUs.

3.3 Individual respondent selection

Fieldworkers were instructed to interview all households living at the selected address/dwelling unit. If they found that the dwelling unit was vacant or the dwelling no longer existed they were not permitted to substitute the dwelling unit but recorded this information on the household control sheet.

The household control sheet is a two page form. This form was completed for every dwelling unit that was selected in the study, regardless of whether or not a successful interview was conducted. Where more than one household resided at the selected dwelling unit, a separate household control sheet was completed for every household and they were treated in the data as separate units. In order to qualify as separate households they should not share resources or food. Lodgers and live-in domestic workers were considered separate households.

All resident household members at selected dwelling units were included in the NIDS panel, providing that at least one person in the household agreed to participate in the study. The household roster in the household questionnaire was used to identify potential participants in the study. Firstly, respondents were asked to list all individuals that have lived under this “roof” or within the same compound/homestead at least 15 days during the last 12 months OR who arrived in the last 15 days and this was now their usual residence. In addition the persons listed should share food from a common ‘pot’ and share resources from a common resource pool. All those listed on the household roster are considered household members.

All *resident* household members became NIDS sample members. In addition, non-resident members that were “out of scope” at the time of the survey also became NIDS sample members. Out-of-scope household members were those living in institutions (such as boarding school hostels, halls of residence, prisons or hospitals) which were not part of the sampling frame. These individuals had a zero probability of selection at their usual place of residence and were thus included in the NIDS sample as part of the household that had listed them as non-resident members. These two groups constitute the permanent sample members (PSMs) and should have had an individual questionnaire (adult, child or proxy) completed for them. These individuals are PSMs even if they refused to be interviewed in the base wave.

An initial sample of 9600 dwelling units was drawn with the expectation of realizing 8000 successful interviews. However, during the initial round of fieldwork for Wave 1 we did not achieve the target number of households. Therefore we went back to the field to attempt to overturn refusals in 48 PSUs and to visit 24 new dwelling units in 32 of these areas. Stats SA drew an additional 24 dwelling units from their Master Sample in predominantly White and

Asian PSUs in order to improve representation of these population groups in the data. This exercise became known as Phase 2 and is discussed in the next chapter.

4. Wave 1 Data collection

4.1 Publicity and pre-notification of selected households

During discussions with Stats SA during the latter part of 2007, it was proposed that Stats SA should play an active role in the publicity campaign for the base wave. Since Stats SA's Labour Force Survey, General Household Survey and Income and Expenditure Survey had all been conducted in these same areas it was agreed that the nine Stats SA provincial co-ordinators would identify freelance fieldworkers that had experience in conducting publicity campaigns for use in the NIDS PSUs. These publicity officers would be employed directly by the fieldwork organisation for the duration of the publicity campaign. Stats SA provided a trainer who trained the 40 freelance publicity officers in Johannesburg on 10 and 11 January 2008.

Publicity was then conducted over a ten day period with each publicity officer doing an average of one PSU per day. These publicity officers were responsible for updating the listings, making contact with police, community leaders and "gatekeepers" in the area and dropping off NIDS brochures at every selected dwelling unit.

The quality of the publicity campaign seems to have been variable, with some publicity officers doing an excellent job while others seem to have done very little. Monitoring the quality of the work was very difficult since the publicity officers worked alone.

In advance of fieldwork a publicity campaign in the media was undertaken. NIDS staff did interviews for SABC news, community radio stations and newspaper articles were published.

4.2 Wave 1 Fieldworker Training

Wave 1 fieldwork commenced with training in Durban in the last week of January 2008. Fieldwork began in one region at a time in order for any teething problems to be contained and remedied within a specific region before moving on.

Training of about 150 fieldworkers took place at the 4 regional offices, namely Durban, Johannesburg, Port Elizabeth and Cape Town. The schedule is shown in Table 5 below. The use of centralised training venues reduced travel by NIDS team members for training. It also allowed for a larger number of fieldworkers in field for a shorter period of time. In the planning phases this was thought to be the best model. Lessons learned in the field suggested that fewer, top-quality fieldworkers staying in field for a longer period of time would be a better option. Managing fewer teams at any one time reduces the logistical burden on the fieldwork

organisation and allows for tighter supervision of teams. This model will be used in subsequent waves.

The presence of 4 regional points as opposed to 9 provincial points expedited a direct NIDS presence in all phases of quality control. It also reduced transport costs for questionnaire delivery and was intended to keep tighter controls and ensure timeous delivery of completed questionnaires to NIDS.

The first week in each regional office consisted of a full week of training conducted jointly by NIDS and DRA staff. Both of the NIDS PIs, the project manager, one data manager and the NIDS post-doctoral fellow were intensively involved in training. Each fieldworker was given a package containing a training manual, copies of all letters and publicity information and two complete sets of questionnaires to be used for training and annotated for the field. Daily tests were held and marks were recorded in order to allow the fieldwork organisation to assess the quality of all fieldworkers. A qualified nurse assisted with the anthropometric training and a special training session was organised at which measurements were taken of babies and young children.

During week 2, the KwaZulu-Natal field teams went out to complete 1 PSU per team while training continued in Gauteng. During week 3, the NIDS data quality manager trained the fieldwork organisation office staff on in-office quality control (QC) and data editing protocols. The KwaZulu-Natal field teams were then given a one day refresher training session on problems that had been picked up on during quality control. The same process was followed for Gauteng, with teams going out during week 3 and quality control training and feedback occurring in week 4. The Port Elizabeth office QC staff were trained in Durban. The fieldwork organisation elected to train the Cape Town QC staff themselves by sending the fieldwork co-ordinator from Durban to Cape Town in week 8.

Table 5: Fieldworker training dates

Province	# of PSUs	# of Teams	Training sessions	Week
Kwa Zulu-Natal	86	11	Session 1	1 (28 Jan)
Gauteng	48	7	Session 2	2 (4 Feb)
Mpumalanga	30	5	Session 2	2
Limpopo	38	9	Session 2	2
Free State	31	5	Session 2	2
North West	35	5	Session 2	2
Eastern Cape*	53	5	Session 3	5 (25 Feb)
Northern Cape	27	3	Session 3	5
Western Cape	52	10	Session 4	6 (3 Mar)
	400	60	4 Sessions	6 weeks

**One team was trained in KZN*

As with the pre-test, once fieldwork in each PSU was complete, the bundles of questionnaires for each household in the PSU were sent to the fieldwork organisation regional office for internal quality control. Once the quality was deemed satisfactory, the completed household bundles were couriered to the NIDS offices at SALDRU for another round of quality control, coding and data capture. We will describe these NIDS office processes in detail later in this chapter.

4.3 Wave 1 Fieldwork

4.3.1 Phase 1

As part of their bid, DRA proposed the fieldwork model that distributed the work load across four regions rather than 9 provinces. This made use of their established regional system which aggregated the provinces into 4 regions. Logistics associated with the fieldwork for Wave 1 were coordinated from these regions.

Fieldwork teams consisted of a fieldwork manager and between three and four fieldworkers. Field managers were responsible for identifying the correct dwelling unit to be approached based on the listing, introducing fieldworkers to households and taking GPS co-ordinates for all selected dwelling units. Field managers also co-ordinated the use of anthropometric equipment within a PSU. Each field team spent 5 days in a PSU. In that time they were meant to identify each of the selected dwelling units, interview all respondents and overturn refusals. In the case of non-response, the fieldworker had to visit the dwelling unit three different times on three different days before it was accepted as a non-response, i.e. no one at home. Overturning soft refusals was also the responsibility of the field manager.

Based on advice from the Stats SA representative on the Steering Committee, wherever possible households and fieldworkers were matched based on language and race. Due to the insufficient number of fieldworkers trained it was not always possible to match fieldworkers to households in this way. There is some evidence that potential respondents did not respond well to fieldworkers who could not speak their language or who were from a different population group and therefore the response rates were negatively affected. For anthropometric measurements only women were allowed to measure women.

As per the training timetable presented above, the fieldwork began in early February 2008. It was initially scheduled to be complete in May, although NIDS had always built in June as a potential spill-over month. However, there were even more delays than envisaged and fieldwork for the first phase was only completed in July 2008. The completion of an initial round of fieldwork in all PSUs resulted in about 6500 successful households from the 9600 dwelling units that were sampled.

4.3.2 Phase 2

Given the fact that the target number of households (8000) was not realised, we started investigating the feasibility of expanding the fieldwork period and going back into the field. There were a number of options available at the time. Our Steering Committee asked us to prepare a detailed report on non-response and refusals as the basis for a decision. This was done and, based on extensive input from our Steering Committee, we decided on the following strategy: Firstly, we asked Stats SA to draw replacement PSUs for 9 areas in which no interviews were conducted during the first phase of fieldwork. Secondly, it was decided that we would re-visit (with the intention of overturning refusals) all predominantly white PSUs in Gauteng, Mpumalanga, Limpopo and the Western Cape, all Indian/Asian s in Gauteng and KwaZulu-Natal and all Coloured PSUs in Gauteng.

The additional responses generated by this strategy were precious to the panel and made the re-visitation campaign worth pursuing. However, the PIs and the NIDS Steering Committee were of like mind that we needed to further boost the number of responses in urban formal white and Asian/Indian PSUs. Therefore we asked Stats SA to draw a new sample of 24 dwelling units in each of the predominantly white and Asian/Indian PSUs. We interviewed these new households when we visited the PSUs to overturn refusals.

From a sampling point of view this amounted to an *ex ante* decision to oversample by predominant racial group in three provinces. This *ex ante* simplicity was deemed to be good as it implied that the derivation of sampling weights would be easier.

The fieldwork for phase two was put out for tender. In the tender it was clear that the winning fieldwork company would be required to undertake a new publicity campaign and ensure that they could put fieldwork teams into the field that were appropriately matched with the PSUs that they would be working in. The fieldwork company that won this tender was Take Note Trading. As part of the consortium that had won the bid for the Phase 1 fieldwork, they had also been involved in phase one fieldwork.

Due to the timing and short notice of phase 2 fieldwork, NIDS arranged to have the training of all fieldworkers conducted in Cape Town. Some of the fieldworkers had worked on the NIDS project before, but it was decided that all fieldworkers had to attend training again to ensure that the information was communicated in a consistent way.

Fieldworker training took place from Monday 15 September 2008 to Thursday 18 September 2008. Field work started the following weekend. For the sake of efficiency the fieldwork company did not do in-office quality control during Phase 2, completed questionnaires were sent directly to the NIDS offices. The same in-office processes used in phase one were followed once questionnaires were received at the NIDS offices.

Phase 2 was implemented in the same way as phase 1 and the same protocols were followed in field. Flags have been inserted in the data to differentiate phase 1 and phase 2 households and participants from each other.

The result of the phase 2 fieldwork was that 1856 households were (re)visited during field work. In the majority of cases these were "difficult" PSUs that had already proven themselves to be tricky to access and the majority of the households that were being visited had previously refused to participate. Low response rates were, therefore, expected. The re-visit exercise resulted in an additional 807 successful households, which is a 43% response rate. This response is very creditable.

After the additional fieldwork phase NIDS achieved 7305 participating households, consisting of 28255 individuals.

5. Processes in NIDS Office

5.1 Logging system/ IT Systems

The procedure for receiving bundles at the NIDS offices evolved over the life of the study. Initially a manual system was used until the automated SQL system was developed. The automated system led to substantial efficiency and transparency improvements.

The logging system is used to trace the whereabouts of a bundle as it moves from storage to quality control, back to field or to data capture and back to storage. Using the SQL system, we are able to track how many bundles have been received and how many are with quality control or data capture or back in the field. Quick analysis can also be done on response rates by province. This allows the office team to quickly alert the field work organisation of low response areas.

When the questionnaires were received each bundle (household pack) would be checked to ensure that the household control sheet matched the contents of the bundle (i.e. there are the correct number of adult, child, proxy questionnaires) and all document numbers were correctly copied onto the household control sheet. The entire household control sheet was captured at NIDS. In the case of outright (household-level) refusals, the only document in the bundle will be the household control sheet. This document provides key information for understanding sample bias.

5.1.1 Quality control

From the outset, the plans for the base wave incorporated a substantial quality control (QC) and data editing (preparation for data capture) operation at UCT. This recognised the fact that such a QC operation, along with our independent call-backs and re-visits, were our primary means of assessing and ensuring the quality of the fieldwork.

NIDS employed 32 quality controllers on a part-time basis. Where possible, quality controllers were allocated bundles grouped by fieldworker. In this way, repeated errors and possible cheating were more easily identified. Questionnaires that were incomplete or had serious errors were sent back to field. There was an agreed protocol with the fieldwork organisation about what constituted a “back-to-field”. The back-to-fields were limited to missing questionnaires, cases where the wrong questionnaire was completed, questionnaires which has missing sections or those where a signature on the consent form was missing. Occasionally a bundle would be

returned to the field work company if the information supplied by the respondent during an office call back significantly differed from what was recorded on the questionnaire.

The quality controller also prepared the questionnaires for data entry by ensuring that all fields were completed; for example by inserting missing codes where questions had been skipped. Once this was done, the pack was passed on to a “coder” who coded the geo-referenced data (e.g. migration history) and the occupation and sector variables.

5.1.2 Call backs

The NIDS call back strategy encompassed both in-field and telephonic call backs. The purpose of in-field call backs was to verify that the fieldworkers were administering the questionnaires in a professional manner, that the correct households were being interviewed and to obtain feedback from the respondents on the questionnaire and the fieldworkers. In-field call backs also had the advantage that contact could be made with households that had refused and the NIDS representative could gain greater insight into the factors that influenced refusal. Telephonic call backs were used to obtain feedback from the respondents on the questionnaire and the fieldworkers and to query any problematic information detected during the QC process. A Respondent Evaluation Questionnaire (REQ) was administered when doing call backs. This questionnaire included questions regarding the professionalism of the fieldworker, the questionnaire and some more general questions.

In-field call back visits were conducted in April in KwaZulu-Natal (Durban and Pietermaritzburg), Gauteng (Pretoria and Johannesburg) and the Eastern Cape (Port Elizabeth) over a period of 4 days per centre. The Western Cape in-field call backs were not completed due to time constraints.

The in-field call backs yielded a similar picture in each region. The upper income households were very hesitant to participate. The invasion of privacy, concerns regarding the legitimacy of the study and time constraints were cited as reasons for non-participation. Response rates were significantly higher among poorer households. Unemployment proved to be the major contributor to the availability of respondents to participate. Also, poorer households were less suspicious of the legitimacy of the study. Previous community studies that were perceived to have yielded improved service or facility provision in those areas also contributed to greater participation within those communities.

Telephonic call backs were done by our full-time survey assistants. Two random telephonic call backs were done per PSU. If potential problems were identified during these call backs, additional call backs were done to other households in the area. In addition, call backs were

done in cases where information was missing or incomplete but the questionnaire was not sufficiently poor to merit sending it back to field. For example, if a person's gender differed on the roster versus their individual questionnaire a telephonic call back would be done. Finally, telephonic call backs were done for every questionnaire that was sent back to field. We verified that a follow up visit was in fact done and ensured that the fieldworkers behaved courteously and professionally.

5.2 Data capture

Initially it was our intention that data capture would be done in house. However, by early March 2008 it had become evident that data capture was proceeding too slowly for us to meet our deadlines. We decided that the risk of continuing with in-house capture was too great and we put out a tender for the work. Citizen Surveys was awarded the tender. We paid the data capture organisation a per unit cost. They had good systems in place to double capture all questionnaires and then reconcile any differences with reference to the questionnaires. We received regular data dumps that enabled us to check captured data against hard copies of the questionnaires.

Given the late start, data capture was a potential bottle neck. However, the data capture organisation quickly became more proficient at capturing the questionnaires. They also installed new software and worked night shifts to complete the work. Mopping up of data capturing errors was time consuming because bundles had to go off-site to be re-captured and the controls on both sides were quite onerous to ensure that physical documents did not go missing. Here we paid a particularly heavy, but unavoidable, price for the decision to undertake a phase two fieldwork campaign.

5.3 Data storage

5.3.1 Physical storage

The physical questionnaires are stored at NIDS. While Wave 1 is "live", i.e. data cleaning is in process, the documents are kept within the NIDS offices. Once Wave 1 has been "closed" the documents will be moved to a secure area.

5.3.2 Electronic storage

The electronic version of the public release data is hosted by DataFirst. Contact and other confidential information is kept secure in NIDS databases.

DataFirst prepared and documented the public release data in accordance with the internationally recognised Data Documentation Initiative (DDI) standards. The data is stored on the DataFirst servers and disseminated to users through the DataFirst online web catalogue. The catalogue can be viewed at;

<http://www.datafirst.uct.ac.za/home/index.php/metadata-and-downloads>

Users wanting to download the data will be required to fill in an online form which asks for their contact details as well as a brief summary of the type of research they intend conducting with the data. They will also be required to agree to a confidentiality statement before being allowed to proceed. This information is collected and stored on the DataFirst server. Once users have completed the access form they will be e-mailed a unique download link from which they can download the data and documentation. Versions of the data will be made available in a format suitable for the statistical packages, STATA, SPSS, SAS, S-Plus, R as well as in ASCII format.

6. Household and Individual response rates

As mentioned before, response rates in phase 1 were disappointing and phase 2 was embarked upon to realise a more acceptable base wave sample. A detailed analysis of household level and individual level response rates follows. Item non-response rates are not addressed here. Such non-response is flagged in the data and is appropriately discussed in the context of specific analyses in the Discussion Paper series.

Household response rates were calculated using the number of visited dwelling units as the denominator and the number of participating households as the numerator. In the instances where response rates are given by race the predominant race group of the PSU is assigned to all households in that PSU. This is done because, by definition, non participating households were not interviewed and we did not gather information about the race of their members from the questionnaires.

Every effort was made to correctly identify all resident household members at the time of the interview. For different reasons not all resident household members were interviewed. For 1754 adults who were unavailable proxy questionnaires were completed. For a further 1250 adults no questionnaires were completed. For these individuals we only have the information supplied in the household roster, i.e. date of birth, education, etc. They are however panel members and we will attempt to make contact with them in the next wave.

Over the combined field work periods NIDS fieldworkers knocked on 10,642 household doors. Of these households, 7305 agreed to participate and the interview was completed. This equates to a 69% response rate. The total sample for NIDS consists of 409 PSUs. Of those, 9 were replaced in phase 2 because the whole PSU was inaccessible in phase 1. They are therefore excluded from the rest of the calculations.

Figure 1 below presents the breakdown of national response rate by province. It can be seen that KwaZulu-Natal had the best response rate at 81%, while Gauteng, Free State and Western Cape had the worst response rates at 60%. Thus, there was fairly wide variance in response rates at the provincial level. Provinces vary by predominant geotype and race and, in understanding what lies behind this provincial variance, it is useful to look at the breakdown of response rates according to these demographic markers. Figures 2 and 3 present the pictures for geotype and race respectively.

It can be seen that the response rate in urban formal areas was 60%. In all geotypes other than this, response rates of 70% or more were achieved with particularly high response rates in rural informal and tribal areas. Despite the phase 2 strategy, Figure 3 makes it clear that white

response rates are low at 36%. In sharp contrast the response rates for all other racial groups are much higher with the figures for Indian, Coloured and African being 66%, 73% and 76% respectively.

Table 6a below reports the number of households per race group that were achieved in phase 1 and phase 2 of the fieldwork. This is an important table as it makes it clear that the phase 2 strategy more than doubled the number of white households in the sample. Thus, although the response rate of white households was disappointing across both phases of fieldwork, the phase 2 fieldwork was very successful in bolstering the number of white households and individuals that constitute the base sample of NIDS going forward. The same is true of Indian households and individuals.

As mentioned earlier, we have no directly gathered information on the race of the members of non-respondent households and we had no choice but to look at household non-response using our sampling data base to tell us the predominant race group per PSU. Table 6a is calculated in this way and in this sense it is consistent with the preceding analysis in this section. However, given that it reports respondents rather than non-respondents, we can compare these figures to those in the NIDS data. As such, it offers an interesting point of comparison between the PSU-based figures and the actual data that we gathered in the two phases of fieldwork. Table 6b presents the same racial breakdown as 6a based on the NIDS data. There are fewer households in 6b than 6a because there is no race data for 20 households.

Table 6b shows that, relative to predictions using the predominant race of the PSU, we realised notably fewer whites and Indians, notably more Africans and the expected number of Coloureds in the NIDS sample. This is to be expected. We are 15 years into our new democracy and, even though racial desegregation of residential areas has been slow, there has been some movement of Coloured and African households into wealthier residential areas but very little movement the other way.

These inaccuracies in the PSU based exercises were predictable and were discussed in the Steering Committee during the planning of the phase 2 fieldwork. One of the international members of the committee summed up this discussion well by reminding all that, while we do want adequate racial representivity in NIDS, it is an income dynamics study and we therefore want to make sure that we have adequate representivity across South African income classes at the time of the base wave. The dissonance between the figures in Table 6a and 6b provide some comfort that the base wave of NIDS and particularly the phase 2 campaign has achieved this.

Figure 1: Response rates by province

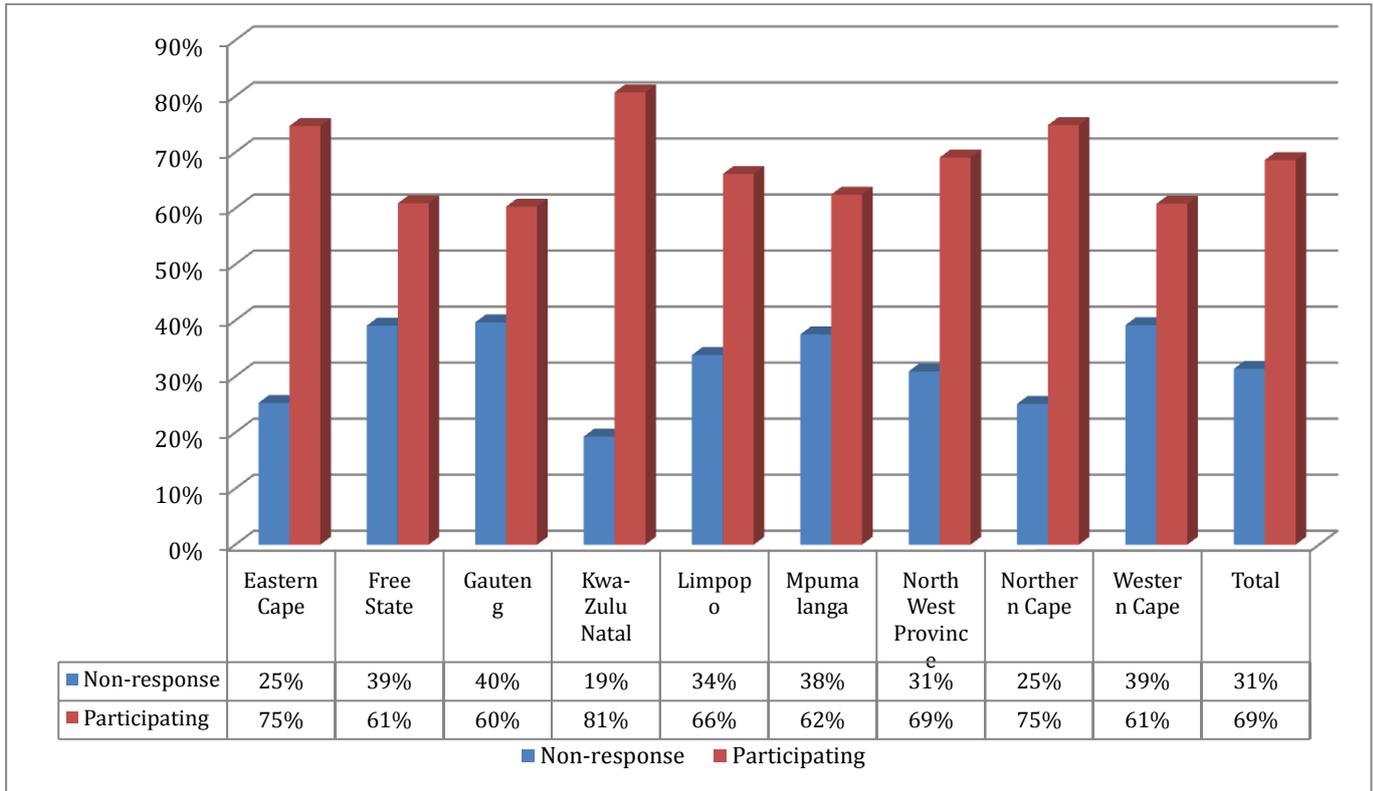


Figure 2: Response rates by geotype

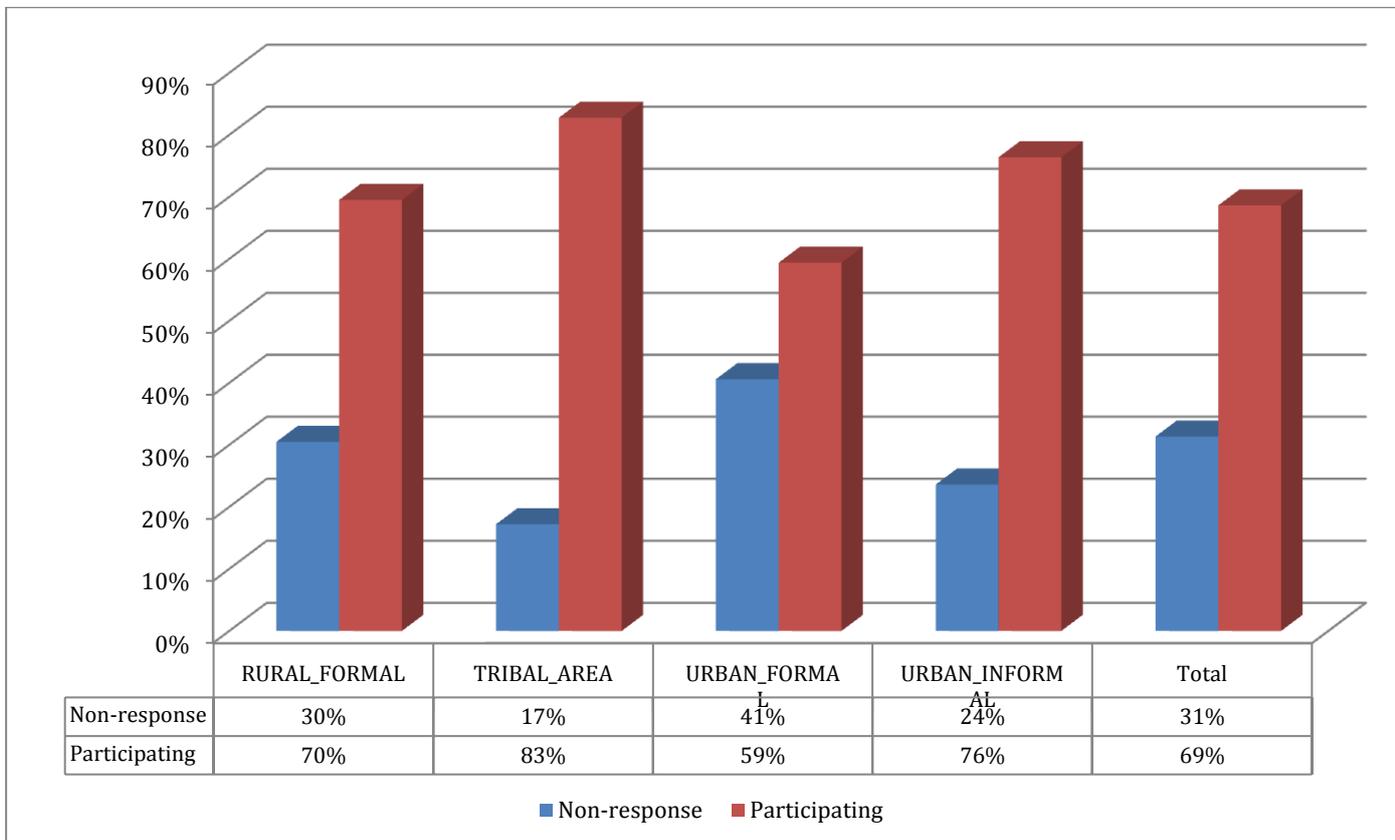


Figure 3: Response rates by race

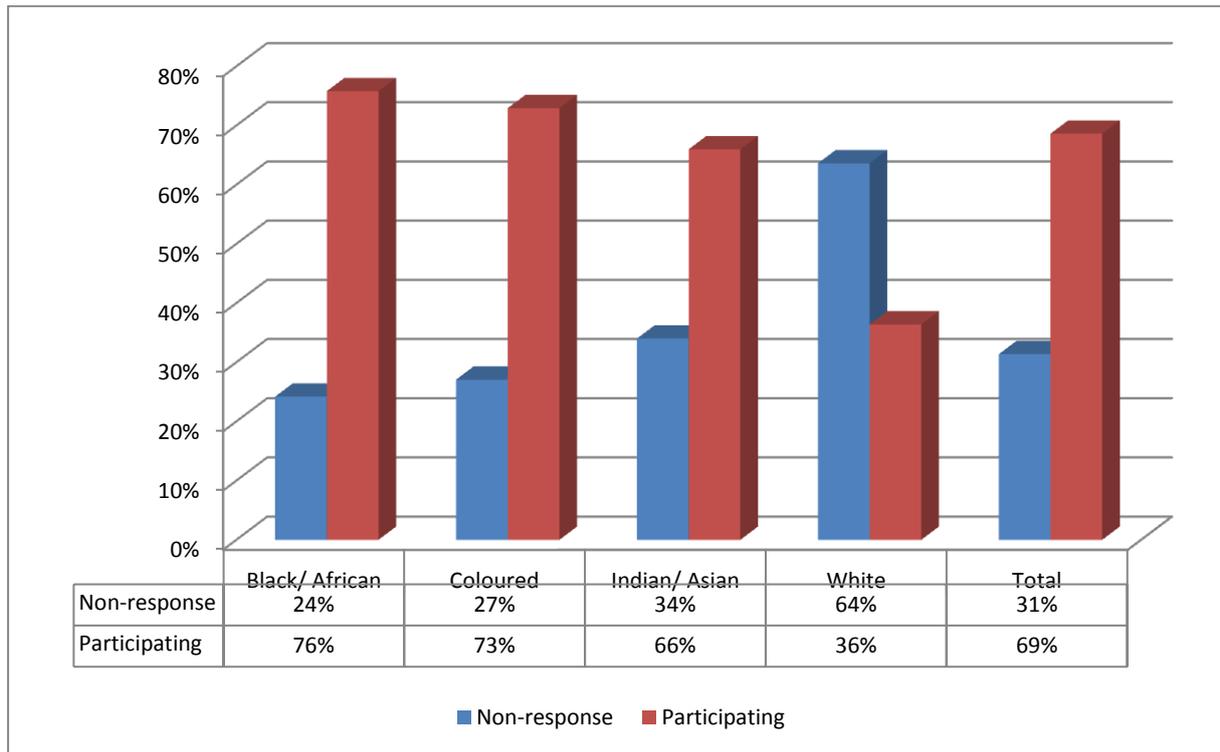


Table 6: Number of households per race group from phase 1 and phase 2.

a) By predominant race group in the sampled PSUs

	Phase 1	Phase 2	Total
Black/ African	5,225	272	5,497
Coloured	951	76	1,027
Indian/ Asian	32	98	130
White	290	360	650
Total	6,498	806	7,304

b) By actual race of respondent in questionnaire

	Phase 1	Phase 2	Total
Black/African	5,202	396	5,598
Coloured	938	77	1,015
Indian/Asian	48	62	110
White	283	270	553
Other	8	0	8
Total	6,479	805	7,284

Most of the contestable assumptions in working on NIDS data arise from three kinds of non-response. Firstly there is household non-response. We have discussed such non-response in

detail above. Our discussion of sampling weights below makes it clear that the key decisions about dealing with such non-response are imbedded in the derivation of these weights. Analysts are likely to assume that the recommended weights take care of such non-response.

Secondly there are non-respondents within responding households. Table 7 below shows the distribution of this unit non-response across responding households. Just over 88% of the 7303 households in the achieved sample had zero non-response. This is encouraging sign in terms of the extent of bias from unit non-response as only about 12% of households are affected at all. In addition, less than 1% of households had a response rate lower than 50%. The 14 households that have 100% non-response to the adult questionnaire are still counted as a responding household because household rosters were completed by those households.

Roughly 6,7% of the sample of adults from the achieved sample of households, did not respond and, for these individuals, we have information only from the household roster. Relatedly, there are adults who were unavailable for interview and for whom proxy questionnaires were completed. Proxy questionnaires make up 9,4% of the adults from the achieved sample of households. For such people we have more information than that contained in the household roster but not complete information. Also, it is an open question as to how the data quality differs given that the questions are not answered by the adult themselves.

Table 7: Intra-household adult non-response rate

	Freq.	Percent	Cum.
0%	6,438	88.16%	88.16%
0% - 25%	105	1.44%	89.59%
25% - 49%	321	4.40%	93.99%
50% - 74%	391	5.35%	99.34%
75% - 100%	34	0.47%	99.81%
100%	14	0.19%	100.00%
Total	7,303	100%	

Each analyst needs to make assumptions about how to deal with non-responding individuals varying from assuming that those that do not respond have no income to assuming that such non-response can be imputed based on the characteristics of the individual (e.g. race, age, sex, geotype etc) that are known from the household roster.

Finally, among individuals or households who do respond to the survey there is item non-response. For example, where an individual professes to earn income from a particular source but does not give the number, we define this as item non-response. Similarly, if a household says that it spends money on phone bills but does not give an amount.

There is very little that can be said about such item non-response in general. The reports on the key variables in NIDS will report on item response. Many derived variables, such as aggregate incomes and expenditures, include imputed values for such non-response. In these cases it is essential that the imputation decisions are made explicit.³

³ The imputations decisions for the derived income variable and the derived expenditure variables are discussed in NIDS Technical Reports No. 3 and No. 4 respectively.

7. Weights

Before analysis and report-writing on the NIDS data could begin it was necessary to calculate sampling weights. Professor Martin Wittenberg at the University of Cape Town was asked to calculate these weights for NIDS. Technical Paper Number 2 *Calculating the NIDS weights* details the methodologies and assumptions made when calculating the weights.

This is essentially a two stage procedure. In the first stage, the design weights were calculated as the inverse of the probability of inclusion. In the second, the weights were calibrated to the 2008 midyear estimates. Two sets of weights are thus provided, the design weights and the post-stratification weights.

The basis of the calculation of the design weights is the information that Stats SA provided to NIDS about the process of two-stage sampling from the Master sample. Two sets of calculations were necessary in deriving the design weights. First there is a calculation of the probability of sampling each PSU and, second, there is a calculation about the probability of including each specific household in each PSU in the NIDS sample. The latter corrects for household non-response.

The second set of weights are the post-stratification weights. These weights adjust the design weights such that the age-sex-race marginal totals in the NIDS data match the population estimates produced by Stats SA for the Mid Year Population Estimates for 2008. In addition, we imposed the constraint that the population distribution by provinces should correspond to that released in those population estimates and that the total weights should add up to the estimated total population of 48,687,000. Finally, a further constraint imposed was that the weights should be constant within households.

8. Data Cleaning and Data preparation

8.1 Preserving anonymity in the data

It is the responsibility of the NIDS team to ensure that respondent identities are protected. During the interviews information was collected that would enable tracking and contacting of respondents for subsequent waves of data collection. However, some of this information is excluded from the public release data set to preserve the anonymity of the respondents. The type of data collected but excluded from public release fall into three broad categories.

Firstly, the names and addresses of the respondents will never be released. This information is kept separately as part of the panel maintenance system.

Secondly, we will not release the detailed geographical information about the respondent's current location or information regarding their migration history. The names of suburbs and towns are not released, but district council identifiers are provided.

Thirdly, detailed information regarding schools, clinics and occupations were collected. These "text"-fields will not be released as they could potentially be used to identify respondents.

NIDS has three versions of data:

1. Public
2. Internal
3. Secure

The Public data does not have any personalised information or school and hospital coding. The Internal data has all the components of the public data as well as coding for schools and hospitals and more detailed information on the geographic location of households (such as the sub - and main - places within which households fall) and the occupations and sectors of employment of panel members. This dataset, or parts thereof, will only be made available for analytical purposes to certain researchers on special request. The Secure data will include all information collected on respondents and will be used only for operational purposes. The Secure data will thus never be made available to researchers but will be used for the purpose of revisiting respondents and any related fieldwork needed to track the NIDS panel households.

8.2 Integration of community level data

The questionnaires were designed in such a way as to facilitate the integration of community and administrative data. This particular project will be one that is embarked on after the public release of the data. Respecting the anonymity of respondents, GPS information and clinic or school data could be used to calculate distances to nearest school or clinic. There is great research and policy potential for such links with administrative data.

8.3 Derived variables and imputed values

Some of the preparation of the NIDS data for public release requires the calculation of derived variables. In deriving these variables one is moving beyond cleaning and preparation of primary data. Two important examples are the calculation of total household income or total household expenditure. Both require the aggregation of all income sources or all expenditure categories for each household and, in addition, require some assumptions about the treatment of missing incomes or expenditures.

The key principle for NIDS is that it is the anonymised primary data that forms the basis of the public release NIDS data. Therefore, any analyst will be able to start from the primary data and aggregate data for themselves and make their own assumptions about how to treat missing data. However, we have provided some derived variables in the public release data. Work that is done within NIDS to impute missing data for individual variables and/or to derive new variables are placed in derived data files that are clearly separate from the primary NIDS data files. In addition the programmes used to derive these variables and clear documentation about the decisions are available to users.

9. Panel maintenance

People who were resident at the selected dwelling units at the time of the interview are the permanent sample members and they will be tracked as they move around South Africa. Children born to female permanent sample members will themselves become continuing sample members.

Over time people move and household compositions change. Given that the estimated time between field work periods is about 2 years it is important to keep respondents interested in the study, while not placing an undue burden on them. During the period between field work phases, we also need to keep track of the information that we learn about households.

In order to achieve this, we have designed software to track our interactions with respondents. The first interaction is the interview. Information about subsequent interactions such as phone calls and post cards are recorded, including the date of the interaction, the person who made the contact and the outcome. There is also provision to record events such as a death within a household. Most importantly we are able to update contact details for individuals and households. The software allows households to split and to keep tracking permanent sample members as they move away. Having an up to date database containing this sort of information before going into field for Wave 2 will greatly assist field work planning.

9.1 Panel maintenance initiatives

As part of the panel maintenance system NIDS has undertaken a number of initiatives and have a couple more planned. Firstly, NIDS commissioned an End of Year card in November 2008. The card had a greeting in all 11 languages. An interface was designed where each participating household's postal address could be manually checked for accuracy and any obvious errors. Where necessary, telephonic call backs were made to find out the correct postal address for households. The envelope also included a change of address post card that respondents could return to NIDS at no cost in order to notify us of any change in their details.

During 2009 we plan to do more telephonic call backs and to visit areas where we have few contact details for households. Some of these initiatives have already begun and are bearing the fruits of our labour. After the public release households will also be sent a brochure with a summary of the key findings and to inform them that we will be returning in 2010.