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 a household questionnaire was administered as well as individual questionnaires for each adult and each child in the household.

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. The authors and the papers are listed in Table 1 below.

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. These included household, adult, child and proxy questionnaires.

Under contract from NIDS a numeracy module was drafted by Professor Patrick Griffin of the Assessment Research Centre at the University of Melbourne. 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 at several schools and on university students at UCT. 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.

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.

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. For further information on the pre-test, finalisation of

the questionnaire and ethics approval see the full methodology document (NIDS technical paper no. 1) is available on the NIDS website http://www.nids.uct.ac.za/

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) were selected from Stats SA's 2003 Master Sample of 3000 PSUs. This Master Sample was the sample used by Stats SA for other household surveys 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.

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 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. 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 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.

For further details on the Wave 1 field worker training refer to the main methodology document (NIDS technical paper no. 1) available on the NIDS website http://www.nids.uct.ac.za/

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.

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, the feasibility of expanding the fieldwork period and going back into the field was investigated. 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: First, we asked Stats SA to draw replacement PSUs for 9 areas in which no interviews were conducted during the first phase of fieldwork. Second, 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.

Field work started in the third week of September 2008 the following weekend. 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.

For further details on the Wave 1 field work refer to the main methodology document (NIDS technical paper no. 1) available on the NIDS website http://www.nids.uct.ac.za/

4. Processes in NIDS Office

For details on the NIDS logging/IT systems, quality control, data capture and cleaning procedures refer to the main methodology document (NIDS technical paper no. 1) available on the NIDS website http://www.nids.uct.ac.za/

5. Accessing the NIDS Data

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

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

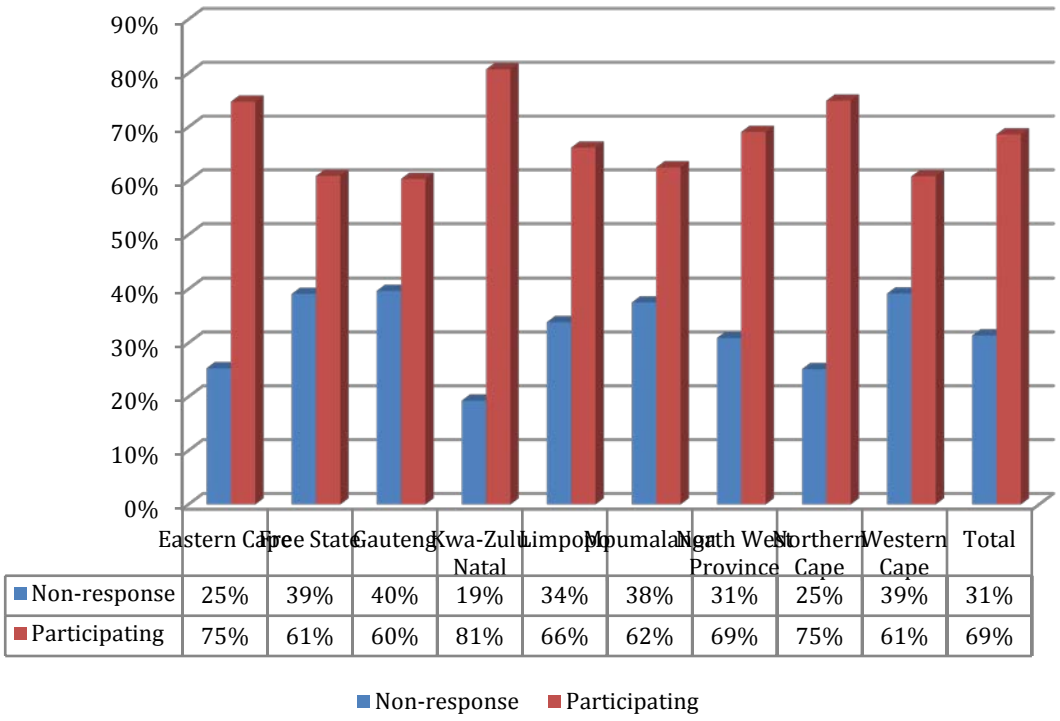
constitute the base sample of NIDS going forward. The same is true of Indian households and individuals.

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.

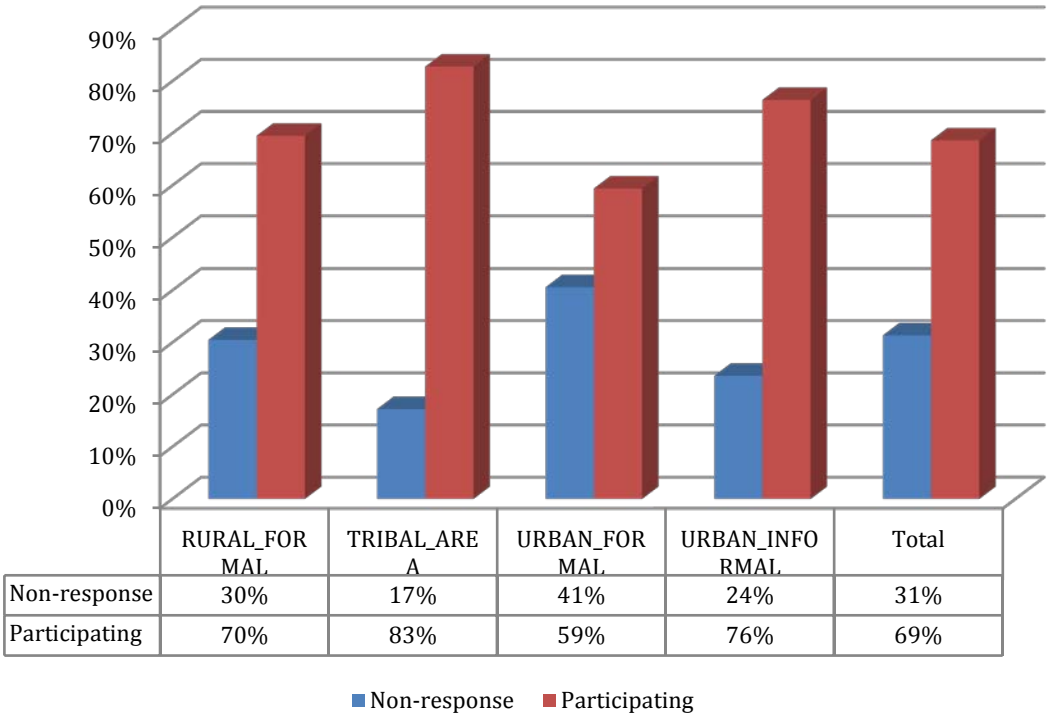
These inaccuracies in the PSU based exercises were predictable and were discussed in the Steering Committee during the planning of the phase 2 fieldwork. One of the international members of the committee summed up this discussion well by reminding all that, while we do want adequate racial representivity 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



	Eastern Cape	Free State	Gauteng	Kwa-Zulu Natal	Limpopo	Mpumalanga	North West Province	Northern Cape	Western Cape	Total
Non-response	25%	39%	40%	19%	34%	38%	31%	25%	39%	31%
Participating	75%	61%	60%	81%	66%	62%	69%	75%	61%	69%

Figure 2: Response rates by geo-type



	RURAL_FORMAL	TRIBAL_AREA	URBAN_FORMAL	URBAN_INFORMAL	Total
Non-response	30%	17%	41%	24%	31%
Participating	70%	83%	59%	76%	69%

Figure 3: Response rates by race

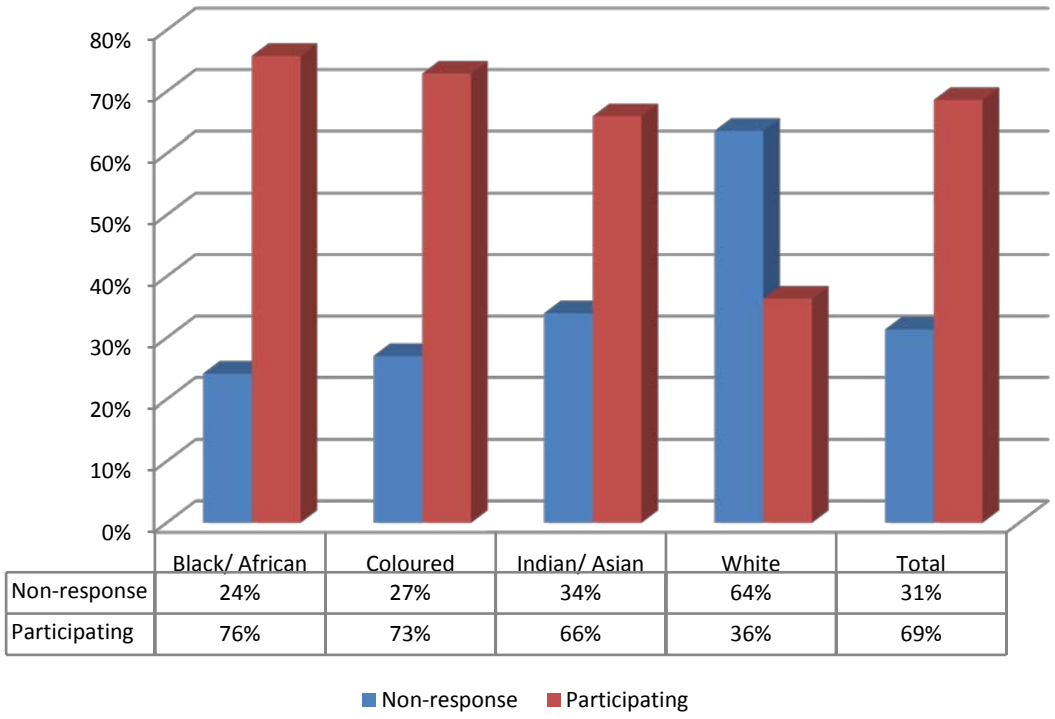


Table 6: Number of households by race group - phase 1 and phase 2

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.

Table 7: Intra-household adult non-response rates

	Freq.	Percent	Cum.
0%	6438	88.16%	88.16%
0%-25%	105	1.44%	89.60%
25%-49%	321	4.40%	94.00%
50%-74%	391	5.35%	99.35%
75%-100%	34	0.47%	99.82%
100%	14	0.19%	100.00%
Total	7303	100.00%	

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.

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.

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 preparation

8.1 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.

8.2 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 is 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.

For further details of the panel maintenance initiative undertaken by the NIDS team refer to the main methodology document (NIDS technical paper no. 1) available on the NIDS website http://www.nids.uct.ac.za/