

# Adding a Top-Up Sample to the National Income Dynamics Study in South Africa

Technical Paper no. 8

Nicola Branson Southern Africa Labour and Development Research Unit

# Adding a Top-Up Sample to the National Income Dynamics Study in South Africa

# Nicola Branson NIDS Technical Paper 8

#### 1. Introduction

The National Income Dynamics Study (NIDS) the first nationally representative panel study in South Africa is a national resource for economic and social research that can be used to underpin development policy. In the past ten years, the survey data has been downloaded by 4,877 registered users, used in multiple training programs and resulted in 177 known journal publications.

Key to the strength of the survey is the tracking of a cohort of nationally representative South Africans. Table 1 shows that overall baseline household response in 2008 was 72%, but response rates in predominantly white areas were only 39%1. The sample was further reduced over the years due to high attrition rates among these groups, especially between Wave 1 and 2. Table 1 in the appendix shows that 52% of the white respondents from Wave 1 and 39% of the Indian respondents were not re-interviewed in Wave 2. Further investigation of response patterns shows that attrition is highest among wealthier individuals of all race groups. Therefore, to improve representativeness of the data and increase the number of white and Indian respondents in the sample overall, in Wave 5 (2017) a sample top-up was undertaken. The aim of this resampling exercise was to interview wealthier individuals of all race groups and in doing so increase the number of white and Indian households. Given the persistent income inequalities in South Africa, the method used to achieve this aim was to draw sample clusters that were predominantly white and Indian from the Census 2011 (the most recent census). Twenty-three years after democracy neighbourhoods are integrating and, as was the case in Wave 1, not all those sampled in specific areas were likely to be either white or Indian. On the other hand, these areas remain high income, and therefore a target area for the group with the highest attrition rates in general.

Table 1: Wave 1 cluster level response rate by predominant population group and whether overlap with Wave 5 Top-up sampling frame

	Over-lapping top	Over-lapping top-up frame		ng top-up frame Not over-lapping top-up frame		Not over-lapping top-up frame		
African	54%	12	78%	277	77%	289		
Coloured	74%	4	76%	50	76%	54		
Indian	71%	4	93%	2	78%	6		
White	36%	39	52%	11	39%	50		
All	44%	59	77%	340	72%	399		

Notes to Table 1: Wave 1 Clusters are identified as overlapping the top-up sampling frame by mapping the wave 1 household coordinates to the small areas (SALS) included in the top-up sampling frame. Note there are 400 clusters in NIDS wave 1. One cluster is not included in Table 1.

<sup>&</sup>lt;sup>1</sup> And this was after 4 PSUs were replaced and additional households selected in phase 2.

This paper briefly describes the methodology, sample and response rate achieved in the wave 5 topup exercise before investigating the representativeness of the combined sample.

#### 2. Sample Design

The sample design for the top-up sample is very similar to the original Wave 1 design, with the key distinction that the sampling frame was restricted to urban residential areas (SALs) from the 2011 Census were the proportion of white residents was 50% or more, or the proportion of Indian residents was 20% or more.

Like the main sample design, the top-up sample involved two-stage sampling with stratification at the district council level. The number of SALs selected per district council were allocated proportionate to the share of the total number of households in the sampling frame in the district council<sup>2</sup>. Geographical photography was used to list households in 84 SALs and the remaining 102 SALs were manually listed in field<sup>3</sup>. 48 households were selected per SAL (in 98% of SALs) unless there were fewer than 48 households available as a result of low household density.

The sampling unit is the household defined by the original definition as all individuals listed on the roster as living under the same "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 who share food from a common 'pot' and share resources from a common resource pool. Only the primary household defined as the main owner of the residence or tenant was included in the top-up sample<sup>4</sup>, differing from the original design where all households at the dwelling point were included. All household members, both resident and non-resident at the time of the survey were included on the household roster. Resident members and those non-resident members who were out of scope were classified as continuing sample members (CSMs) and an attempt was made to interview them. They will also be included in those tracked going forward.

#### 3. Fieldwork

## a. Fieldwork procedures

The fieldwork procedures for the top-up sample emulated that of the main study but also included additional procedures. Procedures emulated included interviewers wearing branded vests and carrying name cards; giving brochures about the survey to selected households; engaging with local police stations or other security structures or gatekeepers; showing of primary access letters to sample members or gatekeepers of sample areas; the use of incentives and monitoring of data quality.

<sup>&</sup>lt;sup>2</sup> See .do file topup\_sampling\_v022019.do – note small SALs were combined (see .do file SALneighbours.do) and high-income areas were oversampled (see lines 241 and 320 in topup\_sampling\_v022019.do)

<sup>&</sup>lt;sup>3</sup> Manual listing was done in areas where individual dwelling units could not easily be distinguished using aerial geography e.g. apartment blocks.

<sup>&</sup>lt;sup>4</sup> This decision was made in field by the fieldwork company and appears to have been motivated by the fact that most dwelling units visited were single household as a result of the top-up being restricted to formal areas. It is not possible to assess the impact of this decision on the sample.

Additionally, the following procedures were employed for top-up sample areas as it became apparent that response rates were exceedingly low<sup>5</sup>: emails to local schools, police stations, body corporates, community policing forums (CPFs) and security companies informing them about the study and its' purpose; visiting schools after they were emailed further inform about the authenticity of the study and address any questions; presentations to police stations or CPFs; newspaper wraparounds; posters; mobile trailer billboards; gazebos where interviewers handed out information flyers; a YouTube video; information on the NIDS website; newspaper articles; a press release; radio interviews and a television interview. Some of these had very localised coverage (e.g. the posters) whilst others national coverage (e.g. the television interview). These were prepared in either English or Afrikaans<sup>6</sup>, or both, depending on the area and medium of information used.

#### b. Interviewers

The fieldwork for Wave 5 of NIDS began in February 2018 and was completed in December 2018. This is similar to the timespan taken to complete Wave 4, which was late September 2014 to August 2015 (with a few weeks break over late December 2014 to early January 2015). A total of 147 interviewers were employed in Wave 5. This is compared to 120 interviewers in Wave 4 of NIDS. The increase in the number of interviewers is attributable to both the natural wave-on-wave growth in the non-top-up sample as well as the inclusion of a top-up sample for Wave 5 of NIDS.

#### 4. Response

Household response in the NIDS top-up was unprecedentedly low. Table 2 shows that only 1008 households of the 8202 in-scope households located were interviewed, resulting in a household response rate of 12%. A large proportion of households refused to participate in the top-up (72%). In a further 1296 (16%) households, no one was found to be home.

Table 2: Household outcomes, comparison of baseline Wave 1 sample and Wave 5 top-up

	Wave	1 - all	Wave 5 - top-up		
	#	%	#	%	
Total dwelling units sampled	10368		8752		
Plus					
Multiple dwelling	491		0		
Total potential sample	10859		8752		
Less:					
Vacant and out of scope	523	4.8%	536	6.1%	
No-access areas	119	1.1%	14	0.2%	
Total in sample households:	10217	94.1%	8202	93.7%	
Baseline HH response:					
Non-contacts	1214	11.9%	1296	15.8%	
Refusals	1698	16.6%	5903	72.0%	
Participating households	7305	71.5%	1003	12.2%	

<sup>&</sup>lt;sup>5</sup> Some of these were only employed for some of the top-up sample areas after it became evident that response rates were much lower than anticipated, or due to cost or availability considerations. For further details see Geospace's document called 'NIDS WAVE 5 SAMPLE REFRESH AWARENESS INITIATIVE REPORT.docx'

<sup>&</sup>lt;sup>6</sup> English and Afrikaans were used as these are the dominant languages for white and Indian respondents.

Notes to Table 2: Wave 1 information sourced from an old document and the number of households do not align with current data. Wave 1 baseline response available in an access database but response outcomes differ.

Why was response so low in the wave 5 top-up? Table 1 and 3 suggest that in part the low response is to be expected in the areas targeted. Table 1 shows that response rate in Wave 1 for clusters that would overlap with the top-up sampling frame was 44% overall, and 36% in clusters defined as predominantly white in the census 2001 data. Thus, while the overall Wave 1 household response rate was 72%, these areas had much lower response. In fact, the response rates presented in wave 1 are after phase 2 of wave 1 was implemented where an effort was made to improve response in these areas.

Table 2 shows that the targeted area types (whether in Wave 1 or Wave 5) were far more likely to encounter refusals than was encountered in the full Wave 1 sample. Part of this is a function of refusals at the point of entry. Table 3 shows that response rates varied across area types and were particularly low in clusters containing estates. 38% of the areas included (69 SALs) were classified as inclusive of estates. Response is these areas was between 4-6%. However, even in areas which were characterised as normal residential or containing blocks of flats, the response rate only reached a maximum 16%.

Table 3: Wave 5 top-up cluster response rates by area type

Area Classification	Average cluster response rate	Number of clusters
In scope:		
Normal Residential (NR) Only	15%	91
NR and Flats	16%	22
NR and Estates	5%	15
NR and Estates and Flats	6%	15
Estates or Flats Only	4%	39
Mine Compounds	46%	1
In scope total		183
Out of scope:		
Prison Compound	0%	1
Holiday Resort	0%	1
Holiday Resort Farms	0%	1
Out of scope total:		3
All		186

Notes to Table 3: The fieldwork team leaders classified areas.

While the NIDS team had expected a low response rate in the top-up, we did not expect it to be as low as 12%. In addition, it is also worth noting that even once the household representative agreed to participate in the survey, individual response within the household was far lower than NIDS had experienced in Wave 1. Table 4 shows that only 73% of individuals listed in participating households agreed to respond, compared to 95% in the full Wave 1 sample, or 92% if the Wave 1 sample is restricted to areas overlapping with the top-up sampling frame. As a result, while the CSM sample was increased by 2775 individuals, only 2016 of these additions completed interviews in Wave 5.

**Table 4: Individual response** 

	W	ave 1	Wave 5
		Overlapping Wave 5 top-	
	All	up frame	Top-up
Interviewed	95%	92%	73%
Refused/not available	5%	8%	27%
Sample size	28226	2516	2775

Notes to Table 1: Wave 1 individuals were identified as overlapping the top-up sampling frame if their wave 1 household coordinates fell within a small areas (SALS) included in the top-up sampling frame.

The overall number of NIDS continuing sample members grew from 33271 to 36046 with the inclusion of the top-up sample in Wave 5. On the dimensions of population group, the inclusion of the top-up sample brings the NIDS sample weighted with the design weights (i.e. before calibration) closer to the 2017 population estimates produced by StatsSA. Given the low response rate however, the probability the respondents included are a random group of the intended sample is high. It is clear from Table 3, for example, that area type was an important predictor of response. While we adjust the design weights to account for this (see .do file 7\_r1\_design.do line 148), users should caution that the chance of the top-up sample being selective in some unobservable way is high.

### 5. Combining the Main and the Top-up samples

#### c. Weighting

Combining the main sample with the top-up sample was not straightforward as the original sample members living in areas in the sampling frame used to select the top-up sample, had a non-zero probability of being included in the top-up sample in addition to their original sample interview. To account for this we needed to adjust their individual weights downwards so as not to allow this group to be overestimated in our population estimates.

We identified households in overlapping areas using their household geographic coordinates in Wave 5 and the sampling frame boundaries. The combined cross-sectional design weight for Wave 5 including the top-up sample was constructed as:

w5_dwgt = w5_dwgt_tu	for those in the top-up sample
w5_dwgt = w5_dwgt_extu	for those in the original sample whose
	households do not fall within the top-up sampling frame
$w5\_dwgt = (w5\_dwgt\_extu+w5\_dwgt\_tu)/2$	for those in the original sample whose
	households overlap with the top-up
	sampling frame

Where w5\_dwgt\_extu and w5\_dwgt\_tu are the Wave 5 design weights adjusted for household non-response for the original sample and top-up sample respectively. These were further calibrated to the StatsSA 2017 mid-year population estimates on age-sex-race and province cells to get the calibration weights (Branson & Wittenberg, 2019).

# 6. Sample Composition

Table 5: Sex, population group and age group shares – NIDS with and without top-up and StatsSA mid-year estimates

		NII	DS		StatsSA
	Without	Top-up	With 7	Гор-ир	Mid-year
Wave 5	Design	Calibrated	Design	Calibrated	2017
characteristic	Weights	Weights	Weights	Weights	Mean
Gender					
Male	45,2%	48,9%	45,7%	48,9%	48,9%
Female	54,8%	51,1%	54,3%	51,1%	51,1%
Population group					
African	87,5%	80,8%	82,2%	80,8%	80,8%
Coloured	8,5%	8,8%	8,3%	8,8%	8,8%
Asian/Indian	1,0%	2,5%	3,3%	2,5%	2,5%
White	2,9%	7,9%	6,2%	7,9%	7,9%
Age group					
0-4	10,9%	10,4%	10,6%	10,4%	10,4%
5-9	11,9%	10,2%	11,3%	10,2%	10,2%
10-14	10,3%	9,0%	10,0%	9,0%	9,0%
15-19	9,5%	8,1%	9,4%	8,2%	8,1%
20-24	9,5%	8,9%	9,2%	8,9%	8,9%
25-29	9,2%	9,8%	8,9%	9,8%	9,8%
30-34	7,4%	9,3%	7,4%	9,4%	9,3%
35-39	5,4%	7,5%	5,6%	7,5%	7,5%
40-44	4,7%	6,0%	4,8%	6,0%	6,0%
45-49	4,1%	4,9%	4,5%	4,9%	4,9%
50-54	4,3%	4,2%	4,5%	4,2%	4,2%
55-59	3,7%	3,5%	3,9%	3,5%	3,5%
60-64	3,3%	2,8%	3,5%	2,8%	2,8%
65-69	2,3%	2,1%	2,5%	2,1%	2,1%
70-74	1,3%	1,4%	1,5%	1,4%	1,4%
75-79	1,2%	0,9%	1,2%	0,9%	0,9%
80-84	0,6%	0,5%	0,6%	0,5%	0,9%
85+	0,6%	0,5%	0,6%	0,4%	0.0%
Sample size	38166	38166	40928	40928	

Notes to Table 5: NIDS sample in Wave 5 including and excluding top-up respondents weighted either using the design of calibration weights. StatsSA 2017 mid-year estimates sourced from StatsSA website (MYPE, 2017)

Table 5 shows that the NIDS Wave 5 sample excluding the top-up weighted by the design weights (adjusted for baseline household non-response), is 87,5% African, 8,5% coloured, 1,0% Indian and 2,9% white. Including the top-up shifts this distribution to 82,2%, 8,3%, 3,3% and 6,2% for Africans, coloureds, Indians and whites respectively. The inclusion of the top up therefore brings the data far closer to the national statistics distribution of 80,8%, 8,8%, 2,5% and 7,9% (StatsSA mid-year estimates 2017).

We had hoped to achieve a sample that over represented Indians and whites in the full Wave 5 data, however this was only achieved for the Indian sub-sample and the white sample remains under representative in the data as weighted by the design weights.

The NIDS Wave 5 data, inclusive or excluding the top-up, is intended to be nationally representative of the 2017 population when weighted with the cross-sectional calibration weights. The data weighted by the calibration weights mirrors, as designed, the StatsSA estimates for both the sample including and excluding the top-up.

### 7. Assessing how survey estimates change when the Wave 5 top-up is included

Table 6 provides a comparison of characteristics according to the NIDS Wave 5 data (including and excluding the top-up sample) and three StatsSA datasets, namely the Community survey (CS) 2016, the General Household Survey (GHS) 2017 and the third quarter data from Quarterly Labour Force Survey 2017 (QLFS: Q3).

Table 6: Characteristics in NIDS Wave 5, including and excluding the top-up, compared to other StatsSA datasets

		N	IIDS		Commu	nity survey
	Withou	ıt Top-up	With 1	Гор-ир		2016
	Mean	N	Mean	N	Mean	N
Married	27%	22440	27%	24272	28%	2131436
Ever given birth	65%	22567	68%	24419	62%	920300
Highest education level	0376	22307	06 /6	24419	02 /6	920300
Degree	4%	22567	5%	24419	4%	2132336
Certificate/Diploma (without Grd 12)	7%	22567	7%	24419	2%	2132336
Certificate/Diploma (with Grd 12)	13%	22567	13%	24419	4%	2132336
Grade 12	20%	22567	21%	24419	31%	2132336
Grade 8-11	41%	22567	40%	24419	39%	2132336
Grade 0-7	11%	22567	11%	24419	12%	2132336
No schooling	3%	22567	3%	24419	5%	2132336
Geographical type						
Urban	68%	9833	70%	10841	70%	933881
Traditional	27%	9833	25%	10841	26%	933881
Farms	5%	9833	4%	10841	4%	933881
Household Size	3,03	9833	3,03	10841	3,29	933881
					GH	S 2017
Household Income	10769,75	9833	12435,58	10841	8246,34	18328
					QLFS	2017 Q3
Labour force status						
Employed	49%	21906	50%	23231	43%	43558
Unemployed strict	12%	21906	12%	23231	16%	43558
Unemployed discouraged	2%	21906	2%	23231	6%	43558
Not economically active	38%	21906	37%	23231	34%	43558

Notes: Calibrated weights used in NIDS. Person weights provided in StatsSA datasets used for Community Survey, GHS and QLFS estimates. Sample for individual outcomes is adults 15-65. All households are included in household level outcomes.

The average characteristics are remarkably similar for the two NIDS samples, except for average household income which increases from 10.7K to 12.4K with the inclusion of the top-up. There are however some marked differences with the other StatsSA datasets. Specifically, NIDS finds:

- More women have given birth, 77% versus 62% in the CS 2016
- More respondents report certificates/diplomas with matric as their highest qualification rather than just a matric as seems common in the CS 2016.
- A higher share of respondents with certificates or diplomas without matric.

- More employment and less unemployment
- Smaller household sizes
- Higher average household income

All but the last two differences can be attributed to differences in administration of the survey and survey instrument. In the NIDS, individual respondents answer their own questionnaire, while the StatsSA surveys are household level questionnaires responded to by a knowledgeable household member on behalf of individual household members. In addition, the birth history, educational attainment and labour force modules are more comprehensive in the NIDS and have been monitored, interrogated for consistency and cleaned over the panel. It is worth taking stock of the fact that these differences in survey methodology result in substantial differences, even at the mean.

Focusing on the value add of the top-up sample, we discuss the increase in household income with the inclusion of the top-up sample. First, it is noticeable that even the NIDS Wave 5 excluding the top-up has a higher average household income than the GHS data. This is particularly interesting given that one of the key motivations for doing the NIDS top-up stemmed from the concern that due to attrition of white, Indian and more affluent individuals overall, the NIDS sample had lost representativeness. This suggests that either the GHS survey may be experiencing similar difficulties in accessing these individuals and/or our calibration weights are somehow overcompensating for the loss of higher income individuals.

Including the top-up sample increases the average household income in NIDS from R10671 to R12410. The increase is in line with the intention of the survey to increase the representativeness of higher income households. Note however that until this point we had been assuming that the calibration weights were accounting for the deficit of rich households.

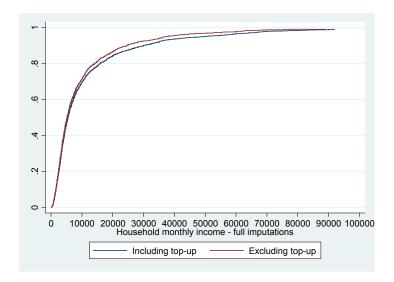


Figure 1: NIDS cumulative income distribution, including and excluding the top-up

Figure 1 looks at the change in household income across the distribution with the inclusion of the topup. The figure is broken down into sections in Appendix Figure 1 for illustrative purposes. The figure shows that the impact of the inclusion of the top-up is already visible quite low in the income distribution. For example, NIDS inclusive of the top-up already shows marginally fewer households with incomes less than R3000 per month. The difference is largest at the R50 000 mark – excluding the top-up, 3% of NIDS households earn more than R50 000, while when the top-up is included, this percentage increases to 5%.

Table 7: Wave 5 poverty and inequality measures – including and excluding Top-up – compared to GHS 2016

	NIDS Wave 5		GHS 2016
	Without Top-up With Top-up		
Headcount poverty rate	50%	48%	55%
Normalised poverty gap	24%	23%	30%
Gini Coefficient	0,64	0,66	0,64

Notes: Measures constructed using per capita household income. Upper poverty line as per Budlender et al. (2015) i.e. R1042 in 2011 rands. Calibration weights used for NIDS.

Table 7 shows how poverty and inequality measures change with the inclusion of the top-up and compare these to the GHS 2016 data. As expected, the inclusion of the top-up reduces the poverty measures. On the other hand, the Gini Coefficient, a measure of inequality increases. Comparing the values to the GHS 2016, we see that the original NIDS Wave 5 data restricted to the original sample is more closely aligned with the GHS 2016 data. One explanation could be that StatsSA is experiencing similar decreasing response in high income areas, leading the remaining sample to struggle to accurately represent the top-end of the income distribution.

### 8. Looking forward

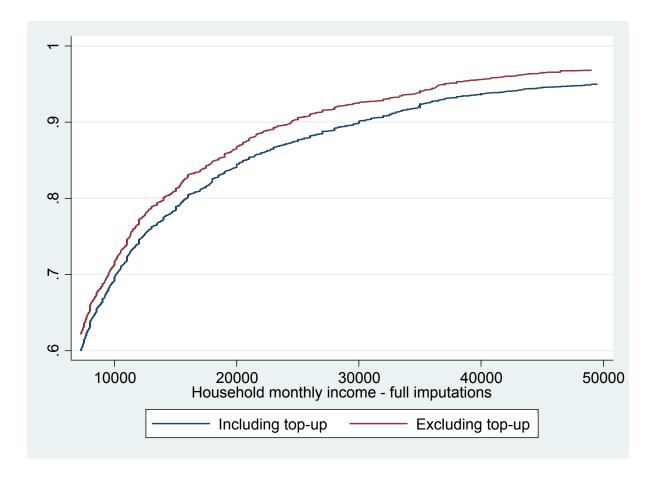
The key challenge going forward will be to retain the top-up sample. As was evident between Wave 1 and 2 for white respondents (see Appendix Table 1), the top-up sample being primarily white and Indian and high income, is most vulnerable to attrition. In addition, the delay of Wave 6 going into field means that the time since these respondents were last contacted is increasing. It would therefore be beneficial to consider contacting the top-up respondents in an unobtrusive way to remind them of the survey and the importance of their participation. Interesting results of research done using the NIDS, especially the YouTube videos, may be a good approach.

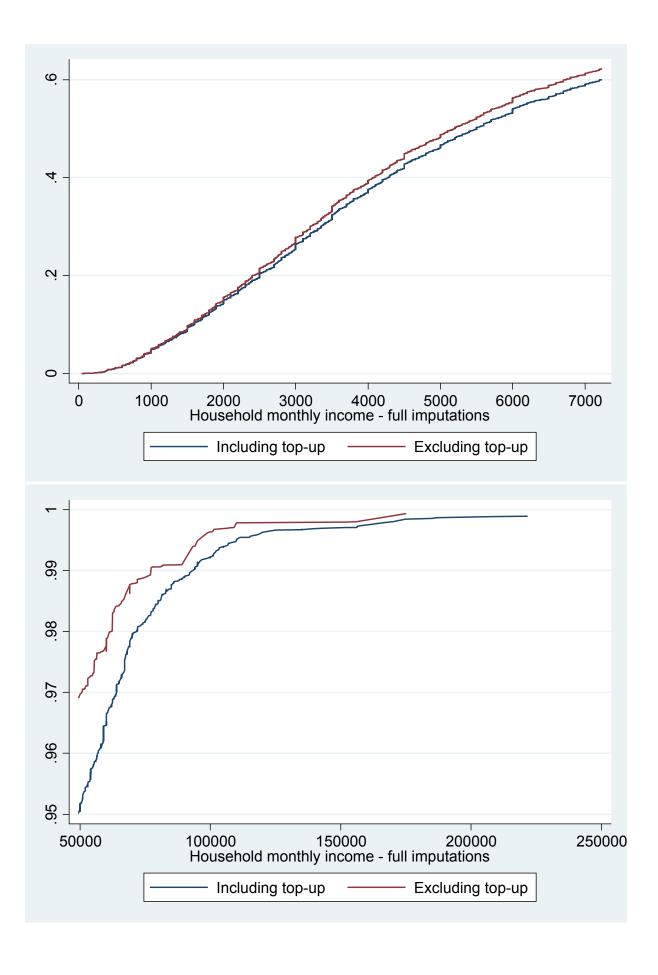
In addition, it would be worth considering a combination of telephonic, self-administered online or mailed versions of the survey as an alternate to in-person interviews. This strategy has been successfully utilised in other panel surveys around the world (Watson & Wooden, 2012). This is a strategy that could be considered for all higher income areas in NIDS, not exclusively for the top-up respondents.

# Appendix Table 1: Response rates by race across waves

	Wa	ve 1	Wa	ve 2	Wav		Wav	/e 4	Wa	<i>r</i> e 5
						JI			1	
Potential	28226	100%	27298	97%	26692	95%	24459	87%	23850	84%
Moved outside of SA			51	0%	56	0%	17	0%	17	0%
Deceased			876	3%	1478	6%	2209	9%	2797	12%
Not tracked			1	0%	0	0%	1541	6%	1562	7%
Successful	26776	95%	22063	81%	22380	84%	21706	89%	20113	84%
Refused	1450	5%	624	2%	333	1%	419	2%	899	4%
HH non response			4611	17%	3979	15%	2334	10%	2838	12%
					Δfri	can				
Potential	22206	100%	21441	97%	20955	94%	19481	88%	18990	86%
Moved outside of SA	22200	10070	26	0%	29	0%	5	0%	5	0%
Deceased			738	3%	1222	6%	1812	9%	2288	12%
Not tracked			1	0%	0	0%	908	5%	923	5%
Successful	21217	96%	18108	84%	18175	87%	17757	91%	16611	87%
Refused	989	4%	411	2%	210	1%	328	2%	480	3%
HH non response			2922	14%	2570	12%	1396	7%	1899	10%
	Coloured									
Potential	4156	100%	4048	97%	3959	95%	3631	87%	3543	85%
Moved outside of SA			6	0%	6	0%	1	0%	0	0%
Deceased			102	3%	191	5%	294	8%	379	11%
not tracked			0	0%	0	0%	230	6%	234	7%
Successful	3847	93%	3036	75%	3248	82%	3136	86%	2882	81%
Refused	309	7%	157	4%	77	2%	56	2%	212	6%
HH non response			855	21%	634	16%	439	12%	449	13%
					l a d					
Detential	429	1000/	404	0.00/		ian	220	700/	224	700/
Potential Mayard systems of CA	429	100%	421 0	98% 0%	416 0	97% 0%	338	79% 1%	334	78% 1%
Moved outside of SA Deceased			8	2%	13	3%	20	6%	24	7%
Not tracked			0	0%	0	0%	68	20%	68	20%
Successful	379	88%	255	61%	263	63%	230	68%	204	61%
Refused	50	12%	14	3%	25	6%	13	4%	50	15%
HH non response			152	36%	128	31%	95	28%	80	24%
	White									
Potential	1435	100%	1388	97%	1362	95%	1009	70%	983	69%
Moved outside of SA			19	1%	21	2%	8	1%	9	1%
Deceased			28	2%	52	4%	83	8%	106	11%
Not tracked			0	0%	0	0%	335	33%	337	34%
Successful	1333	93%	664	48%	694	51%	583	58%	416	42%
Refused	102	7%	42	3%	21	2%	22	2%	157	16%
HH non response			682	49%	647	48%	404	40%	410	42%

# Appendix Figure 1: NIDS cumulative income distribution, including and excluding the top-up, broken into subsections of the income distribution





#### References:

Branson, N & Wittenberg, M. (2019). Longitudinal and Cross sectional Weights in the NIDS data 1-5. NIDS Technical Paper 7.

Branson, N (2019). Top-up sampling dofile topup\_sampling\_v022019.do

Brophy, T (2016). Calculating SAL neighbours dofile SALneighbours.do

Geospace (2018). NIDS WAVE 5 SAMPLE REFRESH AWARENESS INITIATIVE REPORT

Statistics South Africa (2017). Mid-year population estimates 2017, Statistical release P0302. http://www.statssa.gov.za/?page\_id=1854&PPN=P0302&SCH=7048

Statistics South Africa (2016). Community survey 2016 data. Sourced from DataFirst data portal in 2018. Statistics South Africa (2017) General Household Survey 2017. Sourced from DataFirst data portal in 2018.

Statistics South Africa (2017) Quarterly Labour Force Survey 2017: Q3. Sourced from DataFirst data portal in 2018.

Watson, N., & Wooden, M. P. (2012). The HILDA survey: a case study in the design and development of a successful household panel survey. *Longitudinal and Life Course Studies*, *3*(3), 369-381.