

Wave 1 Changes between V5.3 & V6.0

This document highlights the final numbers for Wave 1 data in V5.3 (previous public release data sets) compared to Wave 1 V6.0 and also documents major changes in variable names.

i) Change in numbers between Releases

Table 1 below shows the number of individuals and households in each data file. There has been no change in the numbers in all data files since the previous release.

Table 1: Summary of N values

Dataset	Wave 1 V5.3	Wave 1 V6.0	Change
Individual Questionnaires			
Adult	16 871	16 871	0
Child	9 605	9 605	0
Proxy	1 750	1 750	0
Indderived	28 226	28 226	0
Admin Data	28 226	28 226	0
Household Questionnaires			
hhderived	7 296	7 296	0
HHQuestionnaire	7 296	7 296	0
Household Roster Breakdown			
CSM	28 229	28 229	0
TSM	2 915	2 915	0
Total	31 144	31 144	0
Resident	28 226	28 226	0
Non-Resident	2 918	2 918	0
Deceased	0	0	0
Total	31 144	31 144	0

ii) New Variables/Info

Migration Variables

In the Adult, Child and Proxy questionnaires we gather information regarding all the locations in which our respondents have lived in the past (Questions b10 – b16 in the Adult). This set of questions are collectively known as migration questions. In previous releases of the data we have only released these descriptions coded using 2001 Census data to district municipality level (DC). In the latest release of our data we have coded these descriptions to both the 2001 and 2011 Census data and now release both versions of the district municipality codes.

New variables for migration have the suffix *dc_2001* and *dc_2011* for descriptions coded to the 2001 and 2011 Census data respectively.

Birth History Section

We embarked on an exercise to identify and match all the children across Wave 1 – Wave 4 on the Birth History (BH) section. In cleaning this section, we made calls to confirm the number of children

the mother had given birth to. Therefore there are a lot of changes to this section because some children were either added or dropped off in the mother's birth history. An additional gain from this exercise is that all the children on the BH section now have a PID to identify them.

Negative and Positive events Variables

The following "other" variables in the household questionnaire is now available in the *HHQuestionnaire* data file: w1_h_nego_o, w1_h_poso1_o & w1_h_poso2_o.

Employment Codes

Employment codes for Wave 1 data were previously created using the South African Standard Classification of Occupations (SASCO) codes whereas subsequent waves used the International Standard Classification of Occupations (ISCO) codes. In order to make Wave 1 consistent with the other waves, all Wave 1 employment description have now been coded using the ISCO codes. The variable names for the codes created have been changed as shown in Table 2 below. Disaggregated ISCO codes up to the 5-digit level are available in secure data while the one-digit level codes are included in the Public Release data.

Upon inspection, the NIDS team noticed that some respondents in the Adult dataset had answered that they had other self-employment activities but the occupational codes for the other self-employment activities did not exist in data. This new variable has been added: w1_a_emsotatc_isco_c, which has the occupational codes of respondents with other self-employment activities.

Police District data

Police district data has now been included as part of the Admin data file. Variables include distance to the nearest police station and distance to the police station in the district in which the household is located. Only categorical distances have been included in the public release version of the data. Actual distances can be found in the secure release version of the data.

Interviewer Evaluation

Substantive cleaning was done on the variables of which household member helped to complete the questionnaire. This resulted in a fourth person being created for the Proxy questionnaire. The variable added is w1_p_intrespid4.

Parental data

An exercise to reduce inconsistencies in the parental information was carried out for all individuals across all waves. The cases with problems were identified by comparing parental information across waves. In cases where the information varied across waves, calls were made to verify this information. Information obtained from the calls was used to correct the inconsistent parental data. Where we failed to contact the respondents, the data remained unchanged.

Pcode variables

The pcode variables have now been entirely dropped from Wave 1 data. This was done for the following reasons:

- All non-resident individuals now have a pid, thus the pcode becomes a duplicate identifier
- The task of cleaning the identifiers was becoming cumbersome as every pid adjustment required a pcode adjustment
- The pcode did not exist in any of the later datasets therefore we dropped these variable in order to have consistency in the panel.

iii) Variables Re-Named

Table 2 below shows all the variables that have been renamed in V6.0 data.

Table 2: Variable name changes

Questionnaire	Section	Old Variable name	New Variable Name
Adult	Demographics	w1_a_movy	w1_a_moveyr
Adult	Demographics	w1_a_brndc	w1_a_brndc_2001
Adult	Demographics	w1_a_lvbfdc	w1_a_lvbfdc_2001
Adult	Demographics	w1_a_lv94dc	w1_a_lv94dc_2001
Adult	Demographics	w1_a_lv06dc	w1_a_lv06dc_2001
Adult	Labour Market Participation	w1_a_em1occ_c	w1_a_em1occ_isco_c
Adult	Labour Market Participation	w1_a_em2occ_c	w1_a_em2occ_isco_c
Adult	Labour Market Participation	w1_a_emsatc_c	w1_a_emsatc_isco_c
Adult	Labour Market Participation	w1_a_emsoth_c	w1_a_emsothatc_isco_c
Adult	Labour Market Participation	w1_a_emctype_c	w1_a_emctype_isco_c
Adult	Labour Market Participation	w1_a_emhtsk_c	w1_a_emhtsk_isco_c
Adult	Labour Market Participation	w1_a_unemtyp_c	w1_a_unemtyp_isco_c
Adult	Parents' vital status	w1_a_mthwrk_c	w1_a_mthwrk_isco_c
Adult	Parents' vital status	w1_a_fthwrk_c	w1_a_fthwrk_isco_c
Adult	Labour Market Participation	w1_a_eminc	w1_a_emcinc
Adult	Labour Market Participation	w1_a_em1inc_s	w1_a_em1inc_sh
Adult	Labour Market Participation	w1_a_eminc_sh	w1_a_emcinc_sh
Adult	Education	w1_a_ed07payr1	w1_a_ed07paypr1
Adult	Education	w1_a_ed07payr2	w1_a_ed07paypr2
Adult	Education	w1_a_ed07payr3	w1_a_ed07paypr3
Child	Demographics	w1_c_movy	w1_c_moveyr
Child	Demographics	w1_c_brndc	w1_c_brndc_2001
Child	Demographics	w1_c_lv06dc	w1_c_lv06dc_2001
Child	Demographics	w1_c_lvbfdc	w1_c_lvbfdc_2001
Child	Parents' vital status	w1_c_mthwrk_c	w1_c_mthwrk_isco_c
Child	Parents' vital status	w1_c_fthwrk_c	w1_c_fthwrk_isco_c
Child	Education	w1_c_ed07wdexp	w1_c_ed07wdex
Child	Education	w1_c_fthedlev	w1_c_fthtert
Child	Education	w1_c_mthedlev	w1_c_mthtert
Child	Education	w1_c_ed07payr1	w1_c_ed07paypr1
Child	Education	w1_c_ed07payr2	w1_c_ed07paypr2
Child	Education	w1_c_ed07payr3	w1_c_ed07paypr3
Child	Parents and family support	w1_c_fththa	w1_c_fthdtha

Child	Grants	w1_c_grcurecr	w1_c_grcurecrel
Child	Child's Health	w1_c_hlthdes	w1_c_hldes
Child	Parents' vital status	w1_c_mthtrt	w1_c_mthtertyn
Child	Parents' vital status	w1_c_fthtrt	w1_c_fthtertyn
hhderived		w1_hhcluster	w1_cluster
hhderived	-	w1_hhdc2001	w1_dc2001
hhderived	-	w1_hhdc2011	w1_dc2011
hhderived	-	w1_hhgeo2001	w1_geo2001
hhderived	-	w1_hhgeo2011	w1_geo2011
hhderived	-	w1_hhprov2001	w1_prov2001
hhderived	-	w1_hhprov2011	w1_prov2011
hhderived	-	w1_hhmdbdc2011	w1_mdbdc2011
HHQ	A	w1_h_pcode_pid	w1_h_respondent
HHQ	Negative Events	w1_h_negdthfrin	w1_h_negdthfrinc
HHQ	Food Spending and Consumption	w1_h_*prd	w1_h_*prdss
Proxy	Demographics	w1_p_movy	w1_p_moveyr
Proxy	Demographics	w1_p_brndc	w1_p_brndc_2001
Proxy	Demographics	w1_p_lv06dc	w1_p_lv06dc_2001
Proxy	Demographics	w1_p_lv94dc	w1_p_lv06dc_2011
Proxy	Demographics	w1_p_lvbfdc	w1_p_lv94dc_2001
Proxy	Labour Market Participation	w1_p_emp	w1_p_emactcur_u
Proxy	Labour Market Participation	w1_p_empinc	w1_p_em1inc_sh
Proxy	Labour Market Participation	w1_p_empocc_c	w1_p_em1occ_isco_c
Proxy	Labour Market Participation	w1_p_empprod_c	w1_p_em1prod_c