Wave 2 Changes between V3.1 and V4.0.0

This document highlights the final numbers for Wave 2 V4.0.0 compared to Wave 2 V3.1 (previous public release dataset) and also documents some useful information for users and major changes in variable names.

i)   Change in numbers between releases

The below table shows the number of individuals and households in each data file.

Dataset	Wave 2 V3.1	Wave 2 V4.0.0	Change from V3.1 to V4.0.0
Individual Questionnaires			
Adult	21880	21874	-6
Child	11081	11293	212
Proxy	1124	1124	0
indderived	34085	34291	206
Household Questionnaires			
hhderived	9127	9125	-2
HHQuestionnaire	9127	9125	-2
Household Roster Breakdown			
CSM	30159	30373	214
TSM	5997	5994	-3
Total	36156	36367	211
Resident	34085	34291	206
Non-Resident	1054	1059	5
Deceased	1017	1017	0
Total	36156	36367	211
Link File Breakdown			
CSM	29222	29431	209
TSM	5739	5736	-3
Total	34961	35167	206
Deceased	876	876	0
Not Deceased	34085	34291	206
Total	34961	35167	206

Note: Changes in numbers in data files are largely driven by previously incorrect classification of TSM/CSM status, duplicate interviews, as well as respondents incorrectly marked as deceased in previous waves and additional baby CSMs not captured in a previous wave.

ii) New variables and useful information

1. Top-up sample variable

In Wave 5 (2017) a sample top-up was undertaken. To identify individuals that were added in the 2017 top-up, a sample indicator variable was added to all the Wave 5 data files. The variable *sample* was also included in previous waves' Link Files. This variable identifies which sample households and individual respondents originated from. It takes on the value 1 for "2008 sample" and 2 for "2017 sample".

2. Education subject/program

In the Proxy data file, the variable *w2_p_ed10cursub* had a change of coding: "Veterinary" was recoded to 15, "Tourism" to 16 and "Beauty" to 17.

3. Other description variables added

The variable *w2_h_aglndacc_o* was added to the Household data file. The variables *w2_a_unems_o* and *w2_a_intlng_o* were added to the Adult data file.

4. Refusal gender and age

The variables *w2_*_refgen* and *w2_*_refage*, which capture the gender and the age of the person who refused respectively, were added to the Adult and Child data file for consistency (where * represents "a" for the Adult data and "c" for the Child data).

5. Weights

Changes have been made to the design of the weights. Please refer to the draft technical paper *Branson, N & Wittenberg, M (2018). Longitudinal and Cross sectional Weights in the NIDS data 1-5. NIDS Technical Paper 8* for further information about the weights. We don't anticipate that these changes will result in substantive changes to your analysis, therefore if anything seems incorrect or strange, please let us know. Given these changes, please ensure that you use the complete suite of data in the2018 release (i.e. do not use data files from an older NIDS data release together with some of the Release 2018: Version 1 data files).

6. Addition of newly identified CSM babies to prior and current waves.

During Wave 5 data production a new variable called Post Field Respondent ($w`x'\_c\_pfr$) was added to the Child data files in Waves 2, 3, 4 and 5 (where x represents the relevant wave number). This variable was included to indicate CSM babies who were added to a given wave after the conclusion of that wave's fieldwork.

These Post Field Respondent CSMs were identified after CSM mothers confirmed that they had neglected to include these newborn children in their birth histories in prior and current waves. As these children were born to CSMs after the Wave 1 baseline, the children themselves are CSMs and thus form part of the NIDS sample. These children have been added to the prior waves retrospectively with "Not Tracked" interview outcomes. At Wave 5, the total number of CSM babies added across the panel was 354.

The number of children added to the data in each wave is represented in the below table*.

Wave	Number of CSM Children Added
2	204 not tracked
3	165 not tracked, 2 deceased
4	58 not tracked, 1 deceased
5	15 not tracked, 2 deceased

It is important to note that the above totals refer to the numbers of CSM babies added to a particular wave, in most cases the same CSM baby needed to be added to multiple waves. Thus, the above table represents the total number of additions to each wave not the unique number of CSM babies.

Users may be concerned about the increase in household size after the addition of CSM babies. The below table illustrates that there is zero difference for weighted average household size and a small difference for unweighted average household size for each of the affected waves.

Wave	Weighted		Unweighted	
	Average HH Size BEFORE Addition	Average HH Size AFTER Addition	Average HH Size BEFORE Addition	Average HH Size AFTER Addition
2	3.741	3.741	3.899	3.922
3	3.689	3.689	4.095	4.112
4	3.582	3.582	4.138	4.143
5	3.322	3.322	3.804	3.805

iii) Variables renamed

The below table shows all the variables that have been renamed in V4.0.0 data.

Questionnaire	Section	Variable name	New variable name
Adult	Section aa	w2_a_intvr_c	w2_a_intrv_c
Adult[1]	Labour market participation	w2_a_unemwnt_dy	w2_a_unemwnt_v
Adult[1]	Labour market participation	w2_a_unemwnt_dyunit	w2_a_unemwnt_u
Adult	Personal ownership and debt	w2_a_ownvehpr_v	w2_a_ownvehpri_v
Adult	Health	w2_a_hlmedpid	w2_a_hlmedpaypid
Adult	Health	w2_a_hl#	w2_a_hlser#
Adult	Health	w2_a_hlo_o	w2_a_hlser_o
Child	Section aa	w2_c_intvr_c	w2_c_intrv_c
Child	Health	w2_c_hlo_o	w2_c_hl_o
Proxy	Section aa	w2_p_intvr_c	w2_p_intrv_c
Household Questionnaire	Section aa	w2_h_intvr_c	w2_h_intrv_c

Note: # denotes a number.

---

[1] This change was made due to consistency due to a structural change in the questionnaire from Wave 2.