

Wave 4 Changes between V1.1 and V2.0.0

This document highlights the final numbers for Wave 4 V2.0.0 compared to Wave 4 V1.1 (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 4 V1.1	Wave 4 V2.0.0	Change from V1.1 to V2.0.0
Individual Questionnaires			
Adult	26819	26804	-15
Child	13918	13971	53
Proxy	1600	1597	-3
indderived	42337	42372	35
Household Questionnaires			
hhderived	11895	11889	-6
HHQuestionnaire	11895	11889	-6
Household Roster Breakdown			
CSM	32850	32949	99
TSM	16690	16583	-107
Total	49540	49532	-8
Resident	42337	42372	35
Non-Resident	6304	6261	-43
Deceased	899	899	0
Total	49540	49532	-8
Link File Breakdown			
CSM	31967	32056	89
TSM	18412	18313	-99
Total	50379	50369	-10
Deceased	2467	2465	-2
Not Deceased	47912	47904	-8
Total	50379	50369	-10

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. 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 the 2018: Version 1 release (i.e. do not use data files from an older NIDS data version together with some V2.0.0 data files).

3. Variables added

The other description variables *w4_a_emact1yrago_o* and *w4_a_ed08lev_o* were added to the Adult data file as well as the variable *w4_a_hlvistest*.

4. 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) 2018 Correction of Wave 4 Wealth Calculation

During 2018 data production an error in the Wave 4 wealth calculation was identified and corrected. The error impacted the wealth component household possessions assets (p_ass) and subsequently the calculation of net worth in Wave 4 V1.0 and V1.1. However, this error has been corrected for Wave 4 V2.0.0 onwards.

iv) Variables renamed

The below table shows all the variables that have been renamed in V2.0.0 data.

Questionnaire	Section	Variable name	New variable name
Adult	Section aa	w4_a_intvr_c	w4_a_intrv_c
Adult ¹	Labour market participation	w4_a_unemwnt_dy	w4_a_unemwnt_u
Adult	Individual income from non-employment sources	w4_a_inc_o	w4_a_inco_o
Adult ²	Personal ownership and debt	w4_a_ownvehpri	w4_a_ownveh
Adult ²	Personal ownership and debt	w4_a_ownvehpr_v	w4_a_ownveh_v
Adult	Education	w4_a_ed15cur_any	w4_a_ed15curatt
Adult	Education	w4_a_ed14cur_any	w4_a_ed14curatt
Adult	Health	w4_a_hlmedpaypid	w4_a_hlmedpid
Adult	Health	w4_a_hlmedpayr	w4_a_hlmedr
Adult	Health	w4_a_hl#	w4_a_hlser#
Adult	Health	w4_a_hl8_o	w4_a_hlser_o
Child	Section aa	w4_c_intvr_c	w4_c_intrv_c
Child	Child's health	w4_c_hlmedpaypid	w4_c_hlmedpid
Child	Child's health	w4_c_hlmedpayr	w4_c_hlmedr
Child	Education	w4_c_ed15cur_any	w4_c_ed15curatt
Child	Education	w4_c_ed14cur_any	w4_c_ed14curatt
Proxy	Section aa	w4_p_intvr_c	w4_p_intrv_c
Proxy	Education	w4_p_ed15cur_any	w4_p_ed15curatt
Proxy	Education	w4_p_ed14cur_any	w4_p_ed14curatt
Proxy	Health	w4_p_hl#	w4_p_hlser#
Proxy	Health	w4_p_hl8_o	w4_p_hlser_o
Household Questionnaire	Section aa	w4_h_intvr_c	w4_h_intrv_c
Household Questionnaire ³	Agriculture	w4_h_agcrveg	w4_h_agcroth
Household Questionnaire ³	Agriculture	w4_h_agcrveg_o	w4_h_agcroth_o
Household Questionnaire ³	Agriculture	w4_h_agcrveghkg	w4_h_agcrothhkg
Household Questionnaire ³	Agriculture	w4_h_agcrvegsllyn	w4_h_agcrothsllyn
Household Questionnaire ³	Agriculture	w4_h_agcrvegskg	w4_h_agcrothskg
Household Questionnaire ³	Agriculture	w4_h_agcrvegval	w4_h_agcrothval

Notes: # denotes a number.

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

² This change was made because from Wave 4, the word "private" was removed from this question.

³ The reason for these changes is that the variable *w1_h_agcrveg* was used in Wave 1 to ask whether the respondent had grown "Other vegetables" but from Wave 4, the relevant question in the questionnaire just asked whether they had grown "Other".