

Changes to Wave 2 Data from V1.0 to V2.0

Changes were made to both the Wave 1 and Wave 2 datasets during the production of Wave 3. Wave 1 and Wave 2 cleaning of the data has been on-going since the previous release of the data. We paid special attention to key variables, such date of birth, gender, population group and highest education, as well as the birth history section and questions about parent vital status (are they alive or dead) in Wave 3. Where there was a discrepancy across waves we did callbacks to confirm what the correct answer is. Where we have learned that either Wave 1 or Wave 2 was incorrect we updated the data in those datasets to reflect the correct information. In some cases this has resulted in questionnaire type changes due to changes in dates of birth. In other cases we have had to update the CSM/TSM tag for new children as we learned either their real date of birth or where we learned that their mother was interviewed as a TSM in Wave 2.

During Wave 3 fieldwork we also learned about more cases where we had duplicate households from Wave 1 or where people were interviewed in more than one household and incorrectly presented as two separate records. The second scenario was especially prevalent in Wave 2. In extreme cases we also learned of people being incorrectly recorded as deceased in Wave 2 when they are still alive.

As a result there is a change in the number of individuals and households in the Wave 2 sample.

Sample size changes

The table below shows the changes in the number of individuals and number of households between Version 1 and Version 2.

Questionnaire	W2_Version 2.0	W2 Version 1.0	
<i>Adult</i>	21 880	21 955	(75)
<i>Child</i>	11 104	11 092	12
<i>Proxy</i>	1 121	1 134	(10)
<i>indderived</i>	34 098	34 181	(83)
<i>hhderived</i>	9 134	9 170	(36)
<i>HHQuestionnaire</i>	9 134	9 170	(36)

As a result of the above changes the number of CSMs increased by 8 and the number of TSMs decreased by 99.

	W2_Version 2.0	W2 Version 1.0	Differences
CSM	29 222	29 214	8
TSM	5 757	5 856	(99)

Variable Changes

New variables

Additional work was done to identify mothers and father in the NIDS panel even when they are not co-resident with their children or had passed away in previous waves. The new variables are *w2_best_mthpid* and *w2_best_fthpid* and can be found in the indderived file.

Other new variables:

w2_c_care1_o	w2_h_intnohwk
w2_c_ede09wdexp	w2_h_intnon
w2_c_edtrn7	w2_h_intnotres
w2_c_hlo_o	w2_h_intsecdr
w2_c_intfillout	w2_h_nego_o
w2_c_mvsuby	w2_h_ownoth_o
w2_c_mvtwny	w2_h_refage
w2_c_refexpl_o	w2_h_toi_o
w2_h_dwltyp_o	w2_r_phase
w2_h_fdo_o	w2_r_popgrp_o
w2_h_hh_head	w2_best_fthpid
w2_h_intbbar	w2_best_mthpid
w2_h_intbwdog	w2_lo_edu
w2_h_intdog	w2_p_edterlev_o
w2_h_intgrd	w2_p_mvsuby
w2_h_intlkint	w2_p_mvtwny
w2_h_intlkoint	w2_p_resrel
w2_h_intnbhd	

Renamed variables

The following variables have been renamed in the *Child* questionnaire. The changes were made to ensure consistency in the variable names in the *Adult*, *Child* and *Proxy* files.

Old	New
w2_a_decdpid	w2_a_decd
w2_a_decdpid2	w2_a_decd2
w2_a_declrgpid	w2_a_declrg
w2_a_declrgpid2	w2_a_declrg2
w2_a_declvpid	w2_a_declv
w2_a_declvpid2	w2_a_declv2
w2_a_decmempid	w2_a_decmem
w2_a_decmempid2	w2_a_decmem2
w2_a_decschpid	w2_a_decsch
w2_a_decschpid2	w2_a_decsch2
w2_c_ed08wdexp	w2_c_ede08wdexp
w2_c_ed09exp	w2_c_ed09ex
w2_c_ed09exp_o	w2_c_ed09ex_o
w2_c_ed09lev_o	w2_c_edenrol09_o

w2_c_ednoenrolexp_o	w2_c_ed09wdexp_o
w2_c_meas_c	w2_c_chldagre

Dropped variables

Most of the variables dropped were empty variables and the change was done to streamline the dataset. The w`x'_r_age variable has been drop for the *HouseholdRoster* file in all Waves. We encourage users to use the best_age variable in the individual derived. The dropped variables are:

w2_c_ednoenrolexp
w2_c_hl1_o
w2_r_age
w2_r_lng_o
w2_mthpid

Weights

All weights were recalculated in 2013 release. There is a very informative description in the Wave 3 User Manual that explains how the calculation was done and what the relationship is between the different weights. Weights were also calibrated to the latest mid-year population estimates.