

Change Notice # 50 - Data Version 2026 v1

Long-Format Only & New Standardized Variable Naming Convention

- Starting with this release WLS will only release data in the long format, where graduate and sibling participants are in separate records. In addition, WLS has updated variable names to make it easier for researchers to work with the longitudinal structure of the data. More information on working with the long data and why we have renamed measures is explained later in this document.
- Mortality Status now in separate file
 - Because we intend to update mortality status more frequently than we update the main data set we are providing a separate download of mortality data. We have removed these measures of mortality status from the main long data file
 - z_livgrad
 - z_deatyr
 - Researchers needing cause of death data should contact wls@ssc.wisc.edu
- Updated Harmonized Marriage Roster
 - The marriage roster (`wls_marriage_public_2026_v1`) now includes all information on marriages through R08. Measures in the main data set, such as start and end years of marriage and truncated rosters have been removed.
 - When you download the marriage roster, there will be one row for every marriage in the dataset. We also provide Stata code (`reshape_marriage_long_to_wide.do`) that will create a file that has one row/record/observation for each WLS participant and information on all marriages.
- Spouse data, collected as part of R05 (2004-2007), is now a separate download from the main data set.
- We are retiring `idpub` and `rtype`, and introducing `respidpub` as our unique id. `Familypub` and `personid` remain the best way to identify families and their members. We provide a crosswalk between the ids for researchers needing to merge older versions of the data to more recent versions. The new measure `sample_type` can be used to distinguish between graduates and siblings.
- New variable measuring number of completed waves
 - We created `sd_z_round_count` (range 0-8) which measures the number of waves each person completed. People who completed 0 waves are all siblings. Despite never completing an interview they are retained in the data because the graduate participant reports about them during their interviews.
- We made new simplified status measures for each wave. (`r01_z_simple_status...r08_z_simple_status`)
 - A value of 0 for any wave means we did not collect any data from or about this participant
 - A value of 1 for any wave means that we did collect data either from the participant or from an informant.

Changes to WLS data structure and variable names

Over the past two years, WLS has been working to improve the organization, documentation, and usability of WLS data and metadata, while preserving continuity with prior releases. One major step in this effort was creating a new online Portal to better document the longitudinal WLS measures and concord similar variables across Rounds.

The newest release introduces two additional changes intended to make the data easier to use, document, and maintain. We recognize that the following changes may require some adjustment for current users, and we have prepared supporting resources to make the transition easier.

First, WLS is transitioning to a single long-format data release, rather than the wide-format structure used in prior releases. In the long-format data, comparable Graduate and Sibling measures are stacked and each observation is represented by its own row. This structure eliminates the need to pair Grad and Sib versions of the same variable, and better supports repeated measures analysis and multi-level modeling commonly applied to longitudinal data designs. As a result, the long dataset contains substantially fewer variables, reducing the total number from more than 27,000 wide-format variables to approximately 14,000 long-format variables. We expect this structure to make the dataset easier for users to navigate and easier for WLS to update and release in the future.

Second, WLS is applying a standard variable naming convention across the full set of WLS variables. Variable names in the earlier rounds of data collection used a mnemonic naming convention that differed from the convention used in later rounds. The new convention applies the following structure to all variable names:

`{round}_{sample}_{variable stem}_{interview mode}`

For example:

'ocsh64' is now 'r02_g_ocsh64_q'	- grad-only variable
'sfu55ree' is now 'r04_s_fu55ree_i'	- sib-only variable
'z_q2a958re' is now 'r08_z_a958re_i'	- grad and sib variable

This structure is intended to make variable names more interpretable and to help users quickly identify variables of interest by parsing the round, sample, measure, and interview mode associated with a variable.

To support users with existing code or workflows, WLS provides a variable crosswalk linking legacy variable names to the new names ([wls_new_vaname_crosswalk.xlsx](#)). To support users with existing WLS files with depreciated unique ids, WLS provides a crosswalk between respidpub and idpub & rtype (insert file name here).

How to use both Graduate and Sibling Data

The WLS includes two related panels of people followed across time. Original panel members (the “grads”) are a one-third random sample of Wisconsin high school seniors in 1957 (R01). We launched a second panel (the “sibs”) in 1977 (R03) when we attempted interviews with 2,000 of the graduates’ siblings. We expanded the sibling panel in 1993 (R04) when we attempted to interview a sibling for every graduate. Legacy copies of the WLS data were shared in wide format where measures from graduate participants and their selected sibling were released on one record. Currently we release measures from graduates and siblings on separate rows. **Researchers using both the graduate and sibling panels in their work should remember that the observations in the data are not independent.**

WLS Participants within the same Families		
Number of Participants	Individuals	Families
1	1,526	1,526
2	17,434	8,717
3	66	22
4	20	5
Total	19,046	10,270

Each record includes a family level id (familypub) making it easy to identify WLS participants within the same families. Most families in the WLS are represented by two people, but there are families represented by three or four participants. This happened when a family had two 1957 high-school graduates, usually twins. Most often, these two graduates served as each other’s selected sibling. However, in some cases the two graduates were paired with a third, or sometimes a third and fourth, sibling. In multiple waves of the study graduate participants are asked about their selected sibling and if that sibling completes an interview, they are asked about the graduate participant. Use the measure ssibidpub to connect these reciprocal

pairs of participants.

In earlier waves of the study (R03, R04, & R05) we used different instruments for the graduate and sibling participants. The variable sample_type shows which instrument was used for participants during these waves.

Round	Year	Sample Type
R01	1957	Graduates Only
R02	1964	Parent informant reporting about Graduate
R03	1975	Graduate Panel Interviewed
	1977	Sibling Panel Interviewed
R04	1992-1994	Graduate Panel Interviewed
	1993-1994	Sibling Panel Interviewed
R05	2003-2005	Graduate Panel Interviewed
	2004-2006	Sibling Panel Interviewed
	2004-2005	Graduates' Spoused interviewed
	2005-2007	Siblings' Spouses interviewed
R06	2010-2013	Graduates and Siblings interviewed
R07	2019-2022	Graduates and Siblings interviewed
R08	2021-2024	Graduates and Siblings interviewed
R09	2024-2026	Graduates and Siblings interviewed
R10		Graduates and Siblings interviewed

Users with any questions about the WLS data are always welcome to ask for help at wls@ssc.wisc.edu.