

Forms of WLS Data

The Wisconsin Longitudinal Study (WLS) created a long-form of its popular wide-form data set. 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. We launched a second panel (the “sibs”) in 1977 when we attempted interviews with 2,000 of the graduates’ siblings. We expanded the second panel in 1993 when we attempted to interview a sibling for every graduate. Until now, the information from both members of the sibling pair (the graduate and the sibling) were stored in the same record, commonly called the wide form.

Wide Data Structure										
One id for Each Family	Graduate Measures Across Waves					Sibling Measures Across Waves				
idpub	brdxdy	sexrsp	mx001rer	ix001rer	jx001rer	xbrdxdy	ssbsex	nx001rer	dx001rer	px001rer
900018	-3	1				-2	-3			
900021	38	1	4	4	4	20	2			
900024	39	2				38	2			
900026	39	1	5	4	3	33	1	5		4
900034	38	1			4	37	2			
900035	39	1		-4		42	1			
900042	39	1	5	5	4	44	1			
900043	38	1		5	5	41	1	5	4	4
900054	39	1	5	4		41	2			
900069	39	1	5	5	4	-2	-2			
900071	-3	1				-2	-3			

We reshaped the data so that now graduates and siblings from the same family each have their own record with a common family id, familypub. The long form of the data will be useful to researchers who work with the genetic data and those who want to pool graduates and siblings into the same analyses. From now on, we will distribute both the wide and long forms of the data.

Long Data Structure										
One id for Each Person (Two measures)		Unique personids Within Common familypub		Subject of questions regarding paired sibling		Measures Across Waves				
idpub	rtype	familypub	personid	selsibidpub	selsibtype	z_brdxdy	z_sexrsp	z_mx001rer	z_ix001rer	z_jx001rer
900018	g	1201627	1			-3	1			
900021	g	1205260	1	900021	s	38	1	4	4	4
900021	s	1205260	2	900021	g	20	2			
900024	g	1208924	1	900024	s	39	2			
900024	s	1208924	2	900024	g	38	2			
900026	g	1203989	1	900026	s	39	1	5	4	3
900026	s	1203989	2	900026	g	33	1	5		4
900034	g	1202617	1	900034	s	38	1			4
900034	s	1202617	2	900034	g	37	2			

The naming convention for the variables in the long form is as follows. Measures that are judged identical, or nearly identical, for the graduate and sibling participants are stacked and renamed with the prefix “z_” followed by the name of the original graduate variable. Measures that were asked of only one participant retain their original wide name and are coded system missing for the participant who was not asked. There are roughly 200 measures that are similar between the graduate and sibling but

are not stackable in their current form. In a future release of the long file WLS staff will harmonize these measures to create a stacked variable for the convenience of data users.

Experienced WLS data users are likely familiar with the online codebooks of the wide data. As of this writing, we do not have online codebooks that match the long structure of the data. We do have two pieces of documentation for long-form users to reference. The first is a text file with a stata-generated codebook that includes the frequency distributions of all the variables. The second is an excel spreadsheet that lists all measures by Round and documents the original wide source of the stacked long-form pairs. It also includes a link to the online documentation of the original wide variable.

Family Size in the WLS Long-Form Data		
Number of Siblings	Individuals	Families
1	1,536	1,536
2	17,468	8,734
3	30	10
4	16	4
Total	19,050	10,284

The long-form data makes it much easier to identify families that contain more than two people in the study. This happened when a family had two 1957 high-school graduates, usually twins. Most often, these two graduates served as each other's sibling. However, in some cases the two graduates were paired with a third, or sometimes a third and fourth, sibling. In the wide form of the data this issue was addressed by duplicating the data for cases where each graduate in the family was paired with the same third family member. Each graduate had their own record in the data with the paired sibling data attached to both records. Thus, the sibling was in the data twice. The long form of the data removes the duplicated sibling cases.

The creation of the long form data allowed for new measures that capture the self-reported relationship among the siblings. These measures (srel#) came from the early waves of the data and interviewer notes from the original instruments. Used in combination with personid the srel# series of variables shows the relationship between the person in the record and the other members of the family who are also in the sample. For example when personid equals one, srel1 will always equal zero, or "self," as that is the relationship between the person in the record and the first person in the family. However when personid=2, srel1 will describe the relationship between the person in the record and the person in the family with personid=1. As some families have as many as four people in the study there are up to four valid values of the srel series.

Types of Siblings									
One id for Each Person (Two measures)		Unique personids Within Common familypub and Number of Siblings in Sample			Self-reported Relationship with other sibling(s)				Key
idpub	rtype	familypub	personid	sibcount	srel1	srel2	srel3	srel4	
900018	g	1201627	1	1	0				-4 Missing
900021	g	1205260	1	2	0	6			0 Self
900021	s	1205260	2	2	6	0			1 Full Sibling
905087	g	1200781	1	3	0	-4	-4		2 Identical Twin
923670	g	1200781	2	3	-4	0	-4		3 Fraternal Twin
923670	s	1200781	3	3	-4	-4	0		4 Common Dad
916370	s	1203664	1	4	0	3	1	1	5 Common Mom
916370	g	1203664	2	4	3	0	1	1	6 Half (Common Parent Not Specified)
904029	s	1203664	3	4	1	1	0	1	7 Adopted
904029	g	1203664	4	4	1	1	1	0	8 Step

During each round of the survey participants were asked questions about their sibling who was also in the study. As noted above, most families in the study are only represented by two siblings. In these families these siblings are asked about each other. For families with more than two members in the

study, researchers who want to use any of the graduates' reports about their siblings, or vice versa, can use two measures, selsibidpub and selsibtype, to learn which participant was the subject of the questions.

Type of Participant	
g	10,317
s	8,733
Total	19,050

Along with designating whether or not a participant was in the graduate panel or the sibling panel the measures rtype and selsibtype also tell the researcher what instruments the participant received across time. These measures are stored as strings where "g" represents graduate and "s" represents sibling.

Most families have one graduate and one sibling but some have more than one graduate. In the most recent round of surveys, the sixth round, graduates and siblings had identical instruments for all modes of the survey and the panels were fielded concurrently. Prior to that, the instruments for the siblings collected less information than the instruments for the graduates. Rounds three through five were fielded sequentially. Interviews with the sibling panel were not initiated until the interviews with the graduate panel were complete or nearly complete.

Round	Year and Participant	
R1	1957	Graduates Only
R2	1964	Parent informant reporting about Graduate
R3	197X	Graduates interviewed in 1975 Siblings interviewed in 1977
R4	199X	Graduates interviewed in Calendar years 1992-1994 Siblings interviewed in Calendar years 1993-1994
R5	200X	Graduates interviewed in Calendar years 2003-2005 Siblings interviewed in Calendar years 2004-2006 Graduates' Spouses interviewed in Calendar years 2004-2005 Siblings' Spouses interviewed in Calendar years 2005-2007
R6	2011	Graduates and Siblings interviewed in Calendar years 2010-2012

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