

*National Longitudinal Study of
Adolescent Health*

Pairs Code Book



David C. Rowe, University of Arizona

Kristen C. Jacobson, Pennsylvania State University

February 1998

This research was supported by grant P01-HD31921-01 from the National Institute of Child Health and Human Development. Further information may be obtained by contacting Jo Jones, PhD, Project Manager, 919/962-8412 (email: jo_jones@unc.edu) at the Carolina Population Center, CB# 8120 University Square, Chapel Hill, NC 27516-3997.

The Pair Definition File

The file PAIRS.SSD01 defines sibling pairs in the Adolescent Health Study. A row of data in PAIRS.SSD01 contains several identification variables. AID_1 is the unique identifier of sibling 1. In the Add Health Wave I data set, this is the variable AID. AID_2 is the unique identifier of sibling 2. The variable PAIR identifies each pair. It contains a character string, "XXYY###," where XX refers to the type of pair (TW = twin, FS = full sibling, HS = half-sibling, and NR = not related), YY refers to the pair's sex composition (MM = male/male, FF = female/female, and MF = male/female), and ### is a three digit number (e.g., TWFF020). FAMID is the household identification number.

PAIRS.SSD01 contains additional demographic, pair classification, and zygosity variables. For most users, crucial variables are SIBCL1-SIBCL4 that classify all sibling pairs. We recommend the use of SIBCL4, but the different classifications permit a user to select the most appropriate sibling classifications for a study's purpose (i.e., uniting cousins and true not relateds as a *low* related group). SIBCL1 separates the twins as monozygotic (MZ) and dizygotic (DZ) twins. There is an additional class of twins (UD) whose zygosity is as yet undetermined. The other three groups in SIBCL1 come directly from PAIR; they are the full siblings (FS), half-siblings (HS), and biologically not related children reared together (NR). It should be noted that the NR group contains *cousins*, as well as not related sibling pairs, such as step sibs, adopted sibs, boyfriend/girlfriends, etc. The NR group consists of a number of subgroups; thus, we have created variables to identify the type of relationship the sibling pairs have. These variables are described further in the following pages.

Other variables in PAIRS.SSD01 are described later in the codebook. In a pairs data set, variables that have the same value for both pair members have no extension (of _1, 1 or _2, 2), whereas variables specific to each sibling are given an extension.

The Diagnosis of Twin Zygosity

All mixed sex twin pairs were classified as dizygotic (DZ). The majority of same-sex twins were determined to be monozygotic (MZ) or dizygotic (DZ) on the basis of their self-reported confusability of appearance (i.e., look like two peas in a pod as young children, and three items, confused by strangers, teachers, or family members). The zygosity scale (the variable, SIMILAR) was an average of confusability item scores over the reports by both twins. A classification decision made on the basis of a cutoff score where the score distribution seemed to divide naturally. When self-report data on appearance was missing, the classification was made on the basis of the mother's report of confusability of appearance.

Some twin pairs, originally of uncertain status, were classified as MZ or DZ on the basis of molecular genetic markers. The markers were four VNTR loci (D17S30, 3'ApoB, D1S80, D19S20) and three, tetra-nucleotide microsatellite loci (CSF1PO, TH01, PLA2A1). A twin pair was diagnosed as DZ if the pair had at least two loci where they were different. In most cases, the twins were different for more than one loci. The twins were classified as MZ if they matched at five or more loci (approximate error rate of 4/1000 or better). The variable DNA contains the DNA results (coded as MZ or DZ). In SIBCL1, DNA-based classification superseded classification on the basis of confusability of appearance. More details on the genetic markers used is given in Appendix A.

In some pairs, there was a conflict between the twins' self-reports of zygosity and our classification based on confusability of appearance. Both twins believed they were fraternal, but by our classification they were MZ with a similarity score less than 100 (i.e., the maximum possible). These twin pairs were designated UD, for undecided. It should be noted that because these pairs contain the most physically similar true DZ pairs, one way to handle this group is to add them to the DZ group. Another way to handle this group is to simply delete them. To the extent that diagnostic errors are made, the true MZ pairs in this group exert a bias towards lowering estimates of heritability, which may be more acceptable than overestimates to many

researchers.

Frequency	Code	Response	Variable Name	Type/Length
The unique pair identification number for each pair.			PAIR	char 7
368	FSFFxxx	full sibling: female/female		
541	FSMFxxx	full sibling: male/female		
342	FSMMxxx	full sibling: male/male		
117	HSFFxxx	half-sibling: female/female		
210	HSMFxxx	half-sibling: male/female		
115	HSMMxxx	half-sibling: male/male		
190	NRFFxxx	not related: female/female		
302	NRMFxxx	not related: male/female		
170	NRMMxxx	not related: male/male		
288	TWFFxxx	twin: female/female		
204	TWMFxxx	twin: male/female		
292	TWMMxxx	twin: male/male		
The identification number of the first adolescent.			AID_1	char 8
The identification number of the second adolescent.			AID_2	char 8
The household identification number. Some households have more than 1 pair of adolescents. This number is the same for all the pairs within a single household.			FAMID	num 4
3139		range 1001 to 3785		
Zygosity as determined by genotypes.			DNA	char 2
47	DZ	dizygotic		
42	MZ	monozygotic		
3050		zygosity not measured		
The original sibling classification variable.			SIBCL1	char 2
452	DZ	dizygotic twins		
1251	FS	full sibling		
442	HS	half-sibling		
289	MZ	monozygotic twins		
662	NR	not related		

Frequency	Code	Response	Variable Name	Type/Length
43	UD	twin pairs, uncertain zygosity		
The second and most differentiated sibling classification variable. Assigns relationships to the pairs in the not related group, (as well as the twin, full sib, and half-sib pairs) based on dummy variables.			SIBCL2	char 2
31	AA	adopted/adopted; both kids adopted		
49	AB	adopted/bio; one kid adopted, the other kid biological		
18	AU	a pair of aunt/uncle and nephew/niece		
201	CO	cousin pairs		
452	DZ	DZ (fraternal) twins		
27	FO	foster children (either one or both respondents are foster children)		
1251	FS	full sibling pairs		
7	GH	not related pairs living in a group home		
442	HS	half-sibling pairs		
12	IL	pairs who are in-laws (i.e., living with a sister/brother's spouse or boy/girlfriend)		
289	MZ	MZ (identical) twins		
151	NR	not related pairs who are NOT step sibs, adopted sibs, cousins, etc.		
16	SP	spousal (or boy/girlfriend) pairs		
150	SS	step sibling pairs (includes blended families whose parents are not married, but are living together)		
43	UD	twin pairs, uncertain zygosity		
A third pair classification variable. Collapses most of the unrelated pair types into the NR group but does separate out the cousins. Codes the aunt/uncle, niece/nephew pairs and the group home pairs as missing.			SIBCL3	char 2
201	CO	cousin pairs		
452	DZ	DZ (fraternal) twins		
1251	FS	full sibling pairs		
442	HS	half-sibling pairs		
289	MZ	MZ (identical) twins		
436	NR	not related pairs who are NOT aunt/uncle, niece/nephew pairs		
43	UD	twin pairs, uncertain zygosity		

Frequency	Code	Response	Variable Name	Type/Length
25		not related pairs who are aunt/uncle, niece/nephew pairs		
A fourth pair classification variable. Collapses most of the unrelated pair types into the NR group but does separate out the cousins. Codes the aunt/uncle, niece/nephew pairs, the group home pairs, the spouse pairs, and the in-law pairs as missing.			SIBCL4	char 2
201	CO	cousin pairs		
452	DZ	DZ (fraternal) twins		
1251	FS	full sibling pairs		
442	HS	half-sibling pairs		
289	MZ	MZ (identical) twins		
408	NR	not related pairs who are NOT aunt/uncle, niece/nephew pairs, the group home pairs, the spouse pairs, and the in-law pairs		
43	UD	twin pairs, uncertain zygosity		
53		not related pairs who are aunt/uncle, niece/nephew pairs, the group home pairs, the spouse pairs, and the in-law pairs		
A continuous score of twins' self-report of confusability of appearance, in most cases used to estimate twin zygosity. See Appendix B for SAS code.			SIMILAR	num 8
293	0			
57	33.34			
44	50			
23	60			
21	66.66			
15	71.42			
59	75			
14	80			
13	83.34			
10	85.72			
66	87.5			
119	100			
2405	!	missing		
Twins self-report of zygosity.			SZYGOS	num 3

Pairs Code Book

Frequency	Code	Response	Variable Name	Type/ Length
472	0	both twins report DZ		
23	0.5	twins disagree		
238	1	both twins report MZ		
2406	!	missing		

Appendix A: Molecular Genetic Markers Used to Classify Twin Pairs

D17S30

repeat = 70 bp
heterogeneity = .75
refs. 1

A1 .30
A2 .17
A3 .17
A4 .07
A5 .05

3'ApoB

repeat 14/16 bp, 32 bp because 2 repeats occur together
heterogeneity = .80 (heterogeneity in our sample, .86)
refs. 1, 2

A1 .27
A2 .16
A3 .15
A4 .14
A5 .10

D1S80

repeat = 16 bp
heterogeneity = (heterogeneity in our sample, .82)
refs. 1, 2

A1 .40
A2 .27
A3 .07
A4 .07
A5 .07

D19S20

repeat = 33 bp
heterogeneity = (heterogeneity in our sample, .77)
refs. 3

A1 .41
A2 .25
A3 .10
A4 .10
A5 .05

CSF1PO

HUMCSF1PO[AGAT]_n/CSF1R
5q 295/327 pb
heterogeneity = .74 in Caucasians (heterogeneity in our sample, .91)
refs. 4, 5

8 .3
 9 .4
 10 26.4
 11 26.7
 12 33.4
 13 7.0
 14 2.0
 15 .6

PLA2A1

HUMPLA2A1[AAT]/PLA2A

12q 118/139

heterogeneity. = .72 in Caucasians (heterogeneity in our sample, .87)

refs. 4, 5

10 1.9
 11 46.2
 12 13.00
 13 1.9
 14 2.0
 15 .6

TH01

HUMTH01[AATG]n/TH

11p 183/207 bp

heterogeneity = .773 in Caucasians (heterogeneity in our sample, .87)

refs. 4, 5

5 .5
 6 22.6
 7 15.9
 8 11.0
 9 14.3
 10 35.3
 11 .5
 12 0

References

1. Lambropoulous, A. F., Frangoulides, E., Kotsis, A., & Dozi-Vassiliades, I. 1995. Rapid typing of 4 NVTR loci, A'poB, MCT118, St14, and YNZ22 by the polymerase chain reaction of a Greek sample. *Cell Molecular and Molecular Biology* 41: 699-702.
2. Helminen, P., Sajantila, A., Johnsson, V., Lukka, M., Ehnholm, C., & Peltonen, L. 1992 . Amplification of three hypervariable DNA regions by polymerase chain reaction for paternity determinations: comparison with conventional methods and DNA fingerprinting 6:21-26.
3. Tully, G., Sullivan, K. M., & Gill, P. 1993. Analysis of 6 VNTR loci by 'multiplex' PCR and automated fluorescent detection. *Human Genetics* 92:554-562.
4. Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., & Chakraborty, R. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics* 55:175-189.
5. Alford, R. L., Hammond, H. A., Coto, I., & Caskey, C. T. 1994. Rapid and efficient resolution of parentage by amplification of short tandem repeats. *American Journal of Human Genetics* 55:190-195.

Appendix B: SIMILAR and SZYGOS

```

/* the following code recodes and sums up the twin similarity items for each twin */

/* are you and your twin identical or fraternal */

/* twin one */

/* identical */
if hlws7a_1=1 or hlws7b_1=1 or hlws7c_1=1 or hlws7d_1=1 or hlws7e_1=1 then
    zgostw1=1;

/* fraternal */
if hlws7a_1=2 or hlws7b_1=2 or hlws7c_1=2 or hlws7d_1=2 or hlws7e_1=2 then
    zgostw1=0;

/* twin two */

/* identical */
if hlws7a_2=1 or hlws7b_2=1 or hlws7c_2=1 or hlws7d_2=1 or hlws7e_2=1 then
    zgostw2=1;

/* fraternal */
if hlws7a_2=2 or hlws7b_2=2 or hlws7c_2=2 or hlws7d_2=2 or hlws7e_2=2 then
    zgostw2=0;

/* growing up, how alike did you and your twin look */

/* twin one */

/* two peas in a pod */
if hlws8a_1=1 or hlws8b_1=1 or hlws8c_1=1 or hlws8d_1=1 or hlws8f_1=1 then
    simtw1=1;

/* like same family */
if hlws8a_1=2 or hlws8b_1=2 or hlws8c_1=2 or hlws8d_1=2 or hlws8f_1=2 then
    simtw1=0;

/* twin two */

/* two peas in a pod */
if hlws8a_2=1 or hlws8b_2=1 or hlws8c_2=1 or hlws8d_2=1 or hlws8f_2=1 then
    simtw2=1;

/* like same family */
if hlws8a_2=2 or hlws8b_2=2 or hlws8c_2=2 or hlws8d_2=2 or hlws8e_2=2 then
    simtw2=0;

/* did you and your twin ever confuse strangers */

```

```

/* twin one */

/* yes */
if hlws9a_1=1 or hlws9b_1=1 or hlws9c_1=1 or hlws9d_1=1 or hlws9e_1=1 then
    cstrtw1=1;

/* no */
if hlws9a_1=0 or hlws9b_1=0 or hlws9c_1=0 or hlws9d_1=0 or hlws9e_1=0 then
    cstrtw1=0;

/* twin two */

/* yes */
if hlws9a_2=1 or hlws9b_2=1 or hlws9c_2=1 or hlws9d_2=1 or hlws9e_2=1 then
    cstrtw2=1;

/* no */
if hlws9a_2=0 or hlws9b_2=0 or hlws9c_2=0 or hlws9d_2=0 or hlws9e_2=0 then
    cstrtw2=0;

/* did you and your twin ever confuse teachers */

/* twin one */

/* yes */
if hlws10a_1=1 or hlws10b_1=1 or hlws10c_1=1 or hlws10d_1=1 or hlws10e_1=1
    then ctchtw1=1;

/* no */
if hlws10a_1=0 or hlws10b_1=0 or hlws10c_1=0 or hlws10d_1=0 or hlws10e_1=0
    then ctchtw1=0;

/* twin two */

/* yes */
if hlws10a_2=1 or hlws10b_2=1 or hlws10c_2=1 or hlws10d_2=1 or hlws10e_2=1
    then ctchtw2=1;

/* no */
if hlws10a_2=0 or hlws10b_2=0 or hlws10c_2=0 or hlws10d_2=0 or hlws10e_2=0
    then ctchtw2=0;

/* did you and your twin ever confuse family members */

/* twin one */

/* yes */
if hlws11a_1=1 or hlws11b_1=1 or hlws11c_1=1 or hlws11d_1=1 or hlws11e_1=1
    then cfamtw1=1;

/* no */
if hlws11a_1=0 or hlws11b_1=0 or hlws11c_1=0 or hlws11d_1=0 or hlws11e_1=0

```

```

    then cfamtw1=0;

/* twin two */

/* yes */
if hlws11a_2=1 or hlws11b_2=1 or hlws11c_2=1 or hlws11d_2=1 or hlws11e_2=1
    then cfamtw2=1;

/* no */
if hlws11a_2=0 or hlws11b_2=0 or hlws11c_2=0 or hlws11d_2=0 or hlws11e_2=0
    then cfamtw2=0;

/* the following creates a similarity scale score for each twin pair by averaging the perceived
similarity question with the confuse others questions. Range is in percent, from 0-100 */

similar=mean(of simtw1 cstrtw1 ctchtw1 cfamtw1 simtw2 cstrtw2 ctchtw2
              cfamtw2)*100;

/* the following is the zygoty rating from the twins self-report. It is the average of the two twins'
reports of zygoty, with 1=both twins report identical, 0=both twins report fraternal, and 0.5=one
twin says identical and one twin says fraternal */

szygos=mean(of zgostw1 zgostw2);

```