### Partial list of potential topics for BG meeting Sept 21, 2012

1. Excluding R=0 group
   1. We’re somewhat less certain how much genetic material the pair members share (relative to our certainty about other *R* groups).
   2. We’re **much** less sure how much environmental influences the members share.  This is especially true if they’re not the same generation according to the roster.  But still true if they’re stepsisters.
   3. The Victorian Pasta and other influences are totally different (Mike, don’t worry about understanding this reference).
   4. With all the spouses in the *R*=0 group, there’s problems introduced with selective mating isn’t there?  Could that partially explain the higher than expected correlation within the group?
   5. Why are the within-pair correlations so much higher than we would expect for unrelated pairs?
2. SameGeneration in (0, 255)
3. Classifying ambiguous twins and ambiguous sibs with geodataset
   1. Specific dates of births, marriages, divorces, school attendance, and other events; the NLSY public-use files provide only the months and years of these events.(the DOBs should eliminate a lot of nontwins that are still among the MZs and DZs)
   2. Respondent’s country of birth or U.S. State and county of birth.
   3. U.S. State or country in which respondent’s parents and grandparents were born (NLSY79 only). (the regular dataset only provides their country of birth)
   4. ‘BornWithinThreeDaysOfEachOther’, ‘BornInTheSameStateSubject’, ‘BornInTheSameStateMother’, and ‘BornInTheSameStateFather’
   5. “Why didn’t you take advantage of all the information available to you to distinguish the remaining ambiguous twins and sibs?  What other shortcuts did you take?”
   6. Another advantage of using the geocode is that it would (continue to) show that we go well beyond what other researchers could or would do to get effective links created.
   7. Designating mother ship and IT admin
4. Classifying ambiguous twins eye color & hair color
5. ACE variables
   1. Eye color & hair color –ACE a categorical variable?
   2. Height
   3. Weight
   4. Afm
   5. Afi
   6. Afqt & AfqtGaussified
6. V49 no longer uses RImplicit2004. Lost 344 links. 300 of those had height values:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | R=0 | .125 | .25 | .375 | .5 | .75 | 1 |
| Newer | 270 | 63 | 238 | 45 | 3382 | 13 | 11 |
| Older | 270 | 70 | 248 | 90 | 3620 | 13 | 11 |

1. It's also worth looking at our timeline, which we should discuss at our meeting Friday -- here it is:

Months 1-8.5 (May 16, 2011 to Feb. 28, 2012): Prepare NLSYC kinship pair file

Months 5.5-15 (Nov 1, 2011 to August 14, 2012): Prepare NLSY79 kinship pair file

Months 13-24 (August 15, 2012 to May 15, 2013): Prepare cross-generational files

Months 13-24 (August 15, 2012 to May 15, 2013): Validity studies, kinship pair files

Months 16-24 (May 16, 2012 to Aug 14, 2012): Prepare three NLSY multi-level files

Months 19 & following (Nov 15, 2011 & continuing) Circulate kinship pair/multi-level files

Months 19-24 (Nov 15, 2012 to Dec 31, 2012): Load data online; Prepare SAS,

Access/SQL, R files; Send files to CHRR

Months 24 to 31.5 and on (Aug. 16, 2013 and on) Develop technical user support services"

You'll see that we're a little bit behind on the NLSY79 links, but also quite a bit ahead on the validity analyses.  So bottom line, I'm fine with where we are -- and I hope the most interesting work is ahead of us, some productive science using the links, and creation of the cross-generational files (which is very exciting to me, and I hope will be to others as well -- Brian D'Onofrio already is chomping at the bit to use these).

1. Hand coding twins vs systematic
   1. Documenting process
2. My motivation for closely scrutinizing differences between all *R* variants, including RImplicit2004.
3. Pitching similarities & differences between current and older links
   1. Goal 1: Justify federal money spent on new links
   2. Goal 2: don’t discredit previous linking effort
   3. Goal 3: don’t discredit applied research that used previous links
   4. We’ve built up the knowledge and experience of ~20 years. We’re miles beyond what anyone is likely capable of if they’re fresh to this (especially if they don’t have funding specifically to link pairs).
   5. New software approaches make the MV comparisons much more easy
   6. More Gen2 subjects are available (especially those who’ve aged into the good items)
   7. New variables are available (1994 twins & 2006 biomom/dad)
   8. More variables are available (we have two decades more information than the first time Gen1 was seriously linked)
   9. We had to adapt the links to the newer statistical techniques are available
      1. Addresses the reduced bias of including links among younger sibs (instead of linking only to first born)
      2. Multilevel
      3. Spatially inspired
   10. The trajectory of the newer and older links are similar (especially for Gen2), but the newer links are more complete,
       1. More power
       2. Fewer issues about which ambiguous groups to include/exclude in analyses.
4. Priority between 1979 Roster, Implicits, 1994 Explicit twins, and 2006 Explicits biomom/dad
5. Hand-coding terminology
6. Nancy Segal and virtual twins. Does she just want the numbers identified? Will she use NLSY as a recruiting pool for her stuff, or just use the available data?
7. List of the 14 Ambiguous Gen1 Twins

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject1 Tag | Subject2 Tag | Sibling Count Of Subject1 | Sibling Count Of Subject2 | Kid Count Of Subject1 | Kid Count Of Subject2 |  | Twin Count | Aunt-Niece Count | Potential Cousin Count | Parent- Child Count |
| 110300 | 110400 | 4 | 4 | 1 | 1 |  | 1 | 6 | 1 | 2 |
| 117800 | 117900 | 3 | 3 | 0 | 2 |  | 1 | 4 | 0 | 2 |
| 163500 | 163600 | 3 | 3 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 541300 | 541400 | 3 | 3 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 557000 | 557100 | 2 | 2 | 0 | 2 |  | 1 | 2 | 0 | 2 |
| 572000 | 572100 | 2 | 2 | 2 | 0 |  | 1 | 2 | 0 | 2 |
| 598500 | 598600 | 3 | 3 | 2 | 2 |  | 1 | 8 | 4 | 4 |
| 736500 | 736600 | 3 | 3 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 771700 | 771800 | 5 | 5 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 786000 | 786100 | 4 | 4 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 883600 | 883700 | 3 | 3 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 941000 | 941100 | 2 | 2 | 0 | 0 |  | 1 | 0 | 0 | 0 |
| 1008700 | 1008800 | 2 | 2 | 0 | 1 |  | 1 | 1 | 0 | 1 |
| 1232000 | 1232100 | 3 | 3 | 1 | 0 |  | 1 | 2 | 0 | 1 |
|  |  |  |  |  |  |  | 14 | 25 | 5 | 14 |

SELECT Process.tblRelatedValuesArchive.AlgorithmVersion, Process.tblRelatedStructure.RelationshipPath, Process.tblRelatedValuesArchive.Subject1Tag,

Process.tblRelatedValuesArchive.Subject2Tag, Process.tblRelatedValuesArchive.RImplicitPass1, Process.tblRelatedValuesArchive.RImplicit,

Process.tblRelatedValuesArchive.RImplicitSubject, Process.tblRelatedValuesArchive.RImplicitMother, Process.tblRelatedValuesArchive.RImplicit2004,

Process.tblRelatedValuesArchive.RExplicitPass1, Process.tblRelatedValuesArchive.RExplicit, Process.tblRelatedValuesArchive.RPass1,

Process.tblRelatedValuesArchive.R, Process.tblRelatedValuesArchive.SameGeneration, Process.tblMzManual.Generation, Process.tblMzManual.IsMz,

Process.tblMzManual.Undecided, Process.tblMzManual.Related, Process.tblSubjectDetails.SubjectTag, Process.tblSubjectDetails.KidCountInNls,

Process.tblSubjectDetails.SiblingCountInNls

FROM Process.tblRelatedValuesArchive INNER JOIN

Process.tblRelatedStructure ON Process.tblRelatedValuesArchive.Subject1Tag = Process.tblRelatedStructure.Subject1Tag AND

Process.tblRelatedValuesArchive.Subject2Tag = Process.tblRelatedStructure.Subject2Tag INNER JOIN

Process.tblMzManual ON Process.tblRelatedValuesArchive.Subject1Tag = Process.tblMzManual.Subject1Tag AND

Process.tblRelatedValuesArchive.Subject2Tag = Process.tblMzManual.Subject2Tag INNER JOIN

Process.tblSubjectDetails ON Process.tblRelatedValuesArchive.Subject1Tag = Process.tblSubjectDetails.SubjectTag

WHERE (Process.tblRelatedStructure.RelationshipPath = 1) AND (Process.tblRelatedValuesArchive.AlgorithmVersion IN (49)) AND (Process.tblMzManual.IsMz = 255)