### Partial list of potential topics for BG meeting Oct 10, 2012

1. Outcome variables to include in package now for simple ACE. I’d like it to support an MDAN without extra work.
   1. Height
      1. Matching Gen1 & Gen2 procedures
      2. Clipping & standardizing
      3. Concerns about being scooped?
   2. Weight
      1. How is pregnancy accounted for?
   3. ASFT/IQ –ACE stinks for Gen1, good for Gen2
   4. Eye color & hair color
   5. Afm
   6. Afi
   7. Others?
2. Gen2 Height

Two pieces of good news:

(1) the ACE estimates are very consistent with both version of standardized height, and

(2) the Npairs jumped from 3,487 to 5,884.  I changed some things how the standardized heights were calculated.

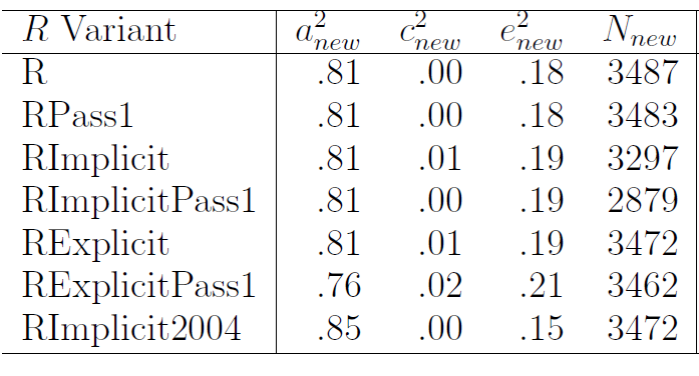
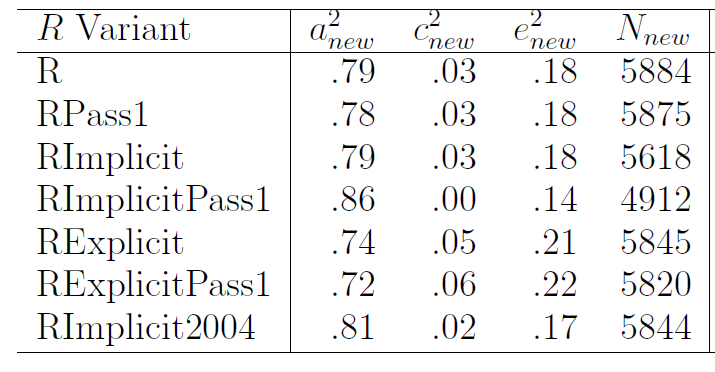
The two big changes were that I (a) cast a larger net for the age range, and (b) I standardized *all* heights, regardless if it wasn’t the value we were going to retain.  My hope was that it would make the heights at the younger ages more stable (because there are more scores to estimate the mean & SD of the Age\*Gender cell).

When the R=.375 group was excluded, the ACE estimates didn’t budge.

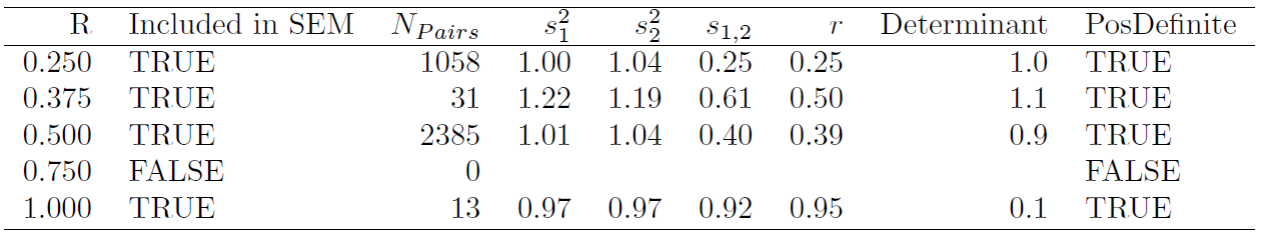
I’m in the middle of upgrading’s Gen1 status to a full member of the R package.  Before, the package didn’t have anything like height.  There was a slight problem, because the Gen2 heights (that Kelly did last year) weren’t aligned with Gen1 height.

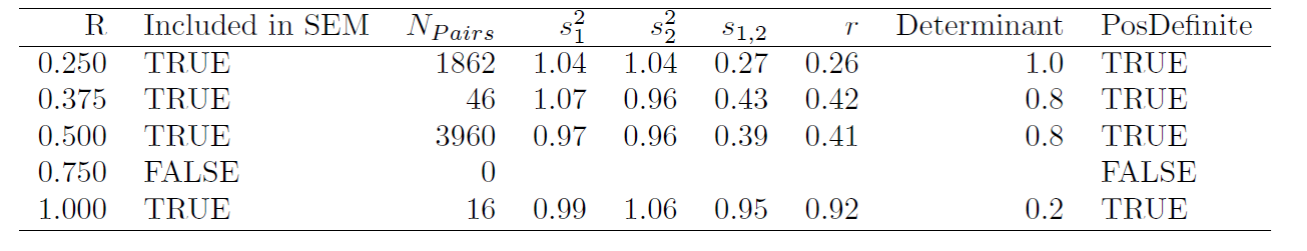
Gen1 was surveyed in a more restrictive way (height was asked in 1981, 1982, 1985, and not another one until 2006).  So I went backwards, re-extracted Gen2 heights, and realigned it with the Gen1 age window in 1981 (which are ages 16-24).

Older (left) and newer (right) Heights

Older (top) and newer (bottom) subgroup correlations





1. Geocode Variables
   1. The program’s ready to go. Karima will run it 9:00am Oct 25, with me on the phone. Check me on my logic/flow. See if I’m overlooking something, or if I can squeeze out something more.
   2. The DOBs are primarily for ambiguous twins; the parents’ birth locations are for ambiguous sibs.
   3. Input Variable List:

"R00001.00"="CaseID",

"R00003.00"="DobMonth1979",

"R00002.00"="DobDay1979",

"R00005.00"="DobYear1979",

"R04101.00"="DobMonth1981",

"R04102.00"="DobDay1981", (more accurate than Dob1979, but 491 are missing)

"R04103.00"="DobYear1981",

"R00009.00"="BirthSubjectCounty", (these are in FIPS codes)

"R00010.00"="BirthSubjectState",

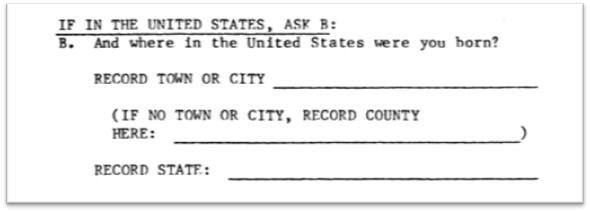
"R02191.14"="BirthSubjectCountry",

"R00062.00"="BirthMotherState",

"R00063.00"="BirthMotherCountry",

"R00074.00"="BirthFatherState",

"R00075.00"="BirthFatherCountry"

* 1. Sanitized Output Variable list:  
     DobDifferenceInDays1979V1979, DobDifferenceInDays1979V1981, DobDifferenceInDays1981V1979, DobDifferenceInDays1981V1981,   
     DobDayIsMissing1979\_1, DobDayIsMissing1979\_2,   
     DobDayIsMissing1981\_1, DobDayIsMissing1981\_2,   
       
     BirthSubjectCountyMissing\_1, BirthSubjectCountyMissing\_2, BirthSubjectCountyEqual, BirthSubjectStateMissing\_1, BirthSubjectStateMissing\_2, BirthSubjectStateEqual, BirthSubjectCountryMissing\_1, BirthSubjectCountryMissing\_2, BirthSubjectCountryEqual,   
       
     BirthMotherStateMissing\_1, BirthMotherStateMissing\_2, BirthMotherStateEqual, BirthMotherCountryMissing\_1, BirthMotherCountryMissing\_2, BirthMotherCountryEqual,   
       
     BirthFatherStateMissing\_1, BirthFatherStateMissing\_2, BirthFatherStateEqual, BirthFatherCountryMissing\_1, BirthFatherCountryMissing\_2, BirthFatherCountryEqual
  2. 
  3. Will: I have a question about the county/state of birth.  Does that refer to the hospital (or whatever location they were at the moment of birth)?  Or does that refer to their house/apartment at that time?  Is there a reasonable chance that some twin members interpret that item differently than each other?  I’m asking because I’m remembering our birthing classes a few years ago.  At least 20% of the parents lived in rural counties 30+ minutes from Norman.
  4. Karima: I'll see what I can find on the coding of the item below or any interviewer instructions on how to answer questions.  I would interpret that to be asking for the actual place of birth (the location of the hospital or other place), but I'm not at all sure that would be a consistent interpretation.  In fact, I suspect it wouldn't be.
  5. Joe: You do bring up the possibility that twins would respond to this differently (e.g., if they were born in OkC at Baptist, but resided in Norman) …But I'd be really surprised if there's any way in the NLSY79 information to tease this apart.  I would expect that virtually everyone would list the location of the hospital
  6. Will: With the full FIPS codes, we could calculate a distance, and make a guess if the hospital was in a neighboring county and the pair interpreted the item differently (eg, one pair responded St. Louis and the other responded East St. Louis), or if they were across the state (eg, Sacramento vs San Diego). I’m willing to let it slide.  The extra info probably isn’t worth the tons of extra legal work and extra spatial calculation work.
  7. Will: I guess we can consider asking them in a few months to run a second R program that calculates and spits out the distances (capped at 60 miles or something, so that we couldn’t deduce that one was born in Hawaii).  Don’t bother responding to this email, if you want to discuss in Friday instead.  If this first run doesn’t go smoothly, I’ll definitely be opposed.
  8. Will: This won’t be a straight-forward routine in R –at least 30 lines of code. The FIPS codes are in a different data.frame, and aren’t used directly by the polygons of ‘maps’ package. The vertices are relatively easily exposed. I could find the minimum distances between the two counties’ vertices. Or the distance from their center points. Hints for FIPS and the maps package: <http://blog.revolutionanalytics.com/2009/11/choropleth-challenge-result.html>

1. Outcome variables to include in package for growth curves

* See longitudinal graphs on pages 3 & 4 of the Gen2 markdown
* Earlier this week Joe and I met and discussed, among other things, some ideas for growth curves on the NLSY-79 Children cognitive variables (PPVT, reading recognition, reading comprehension, digit span, math).  These variables are measured every other year from 1986 to 2010 (as long as the child is within some age range).  We discussed three options.
  + Option 1: Make a separate growth curve for each variable.
  + Option 2: Make a factor for each time point consisting of the five cognitive variables and make a growth curve on the factors.
  + Option 3: Make a factor for each time point consisting of the five cognitive variables and make an autoregressive (moving average too?) dynamic factor analysis on the factors.  With certain constraints, you should be able to make the dynamic factor analysis give you the same information as the growth curve of factors.
* We also discussed the relative merits of using AGE versus YEAR as the time variable, investigating intraindividual effects versus cohort effects.  Finally, we considered adding genetic relatedness (e.g. ACE  models) to this and mother's IQ as measured by ASVAB AFQT.
* So far, we settled on the following.  Extract a suitable 100-300 participants from the 11,000 possible and do some exploratory analyses on these data.  Later we will test hypotheses on the remaining 10,700-10,900 participants.  The exploratory analyses will include the growth curves discussed above, along with plenty of plots, and whatever else the data suggest.

1. Grant Proposal Timeline (again)

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. Pitching similarities & differences between current and older links (again)
   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 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.
2. Priority between 1979 Roster, Implicits, 1994 Explicit twins, and 2006 Explicits biomom/dad  
   -What did we say about this in September?
3. Nancy Segal and virtual twins.