### Partial list of topics for BG meeting Dec 1, 2012

1. Updated ACE Vignette
   1. Explanation/recommendation for dropping Gen1 R=0 group. ¿Add any more explanation?
   2. Midstream SAS –Help Mike?
      1. CSVs or sasbdats?
2. GitHub (<https://github.com/wibeasley/NlsyLinksDetermination>)
   1. Do we want to distribute the SAS links & datasets through it?
   2. Can we altogether avoid a page on OU server? Our GitHub site provides a wiki capability that is a decent way to communicate.(eg, <https://github.com/OuhscCcanMiechvEvaluation/MReporting/wiki/Index>)
3. Helping debug ggplot2, plyr, and devtools.
   1. When they’re a few weeks away from releasing a new major version, they send ‘release candidates’ to all the package authors who depend on them.
   2. What are people’s feelings about our role and how much time?
4. Refactoring package structure so it’s more portable
   1. The restructuring will help Mike and others contribute to it, especially after I’m gone.
   2. Difficult b/c it’s so tied to changing data on my SQL Server database of NLSY kinship & outcome variables.
   3. Unless something miraculous happens at Hadley’s R package workshop, I’m afraid it won’t be restructured as nicely as I would want. I don’t want to sink too much time into it, and neglect features that we’ve already committed to (and are arguably more important).
   4. Even without any further restructuring, Mike & others will always be able to create vignettes pretty easily. I’ve managed to get them fairly encapsulated during the build process.
5. Package Developments since last mass email (mostly again)
   1. Gen1 Links (and MDAN links) are included
   2. Gen1 & Gen2 Links and outcome variables have been updated with the 2010 survey wave
   3. Height is available, and is included in vignette
   4. ACE vignette has explanation about excluding Gen1 R=0 group.
   5. SAS midstream vignette example
6. 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

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 kinship algorithm much easier than 10 and 20 years ago.
      1. Multivariate consideration instead of univariate
         1. eg, simultaneously considering multiple markers/variables, instead of one
         2. eg, simultaneously considering all links within a nuclear family, instead of isolated pairs
      2. Multivariate validation instead of univariate, to catch programming mistakes
         1. eg, mother’s report of twins against the kids’ responses
      3. End-to-end reporting (with dynamic tables and graphs) makes it easier to iteratively modify/calibrate the algorithm several times an hour.
   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,

(i) More power and (ii)Fewer issues about which ambiguous groups to include/exclude in analyses.