**NLSY Kinship Links: Creating Biometrical Design Structures from Cross-Generational Data**

**William Howard Beasley, David Bard, Kelly Meredith, Mike Hunter, Joe Rodgers**

In the fall of 2012, our research team released updated version of kinship links for the two primary National Longitudinal Survey of Youth datasets, the NLSY-Children (NLSYC) and the original NLSY79 surveys. These new kinship links reflect improvements in an already widely used and valuable biometrical resource. The original 12,600 NLSY79 respondents are a U.S. household probability sample; thousands of kinship pairs – a few dozen MZ and DZ twins, thousands of full and half siblings, and hundreds of cousins – reflect population prevalences. Further, the data from the NLSYC, the 11,500 biological children of the NLSY79 females (whose childbearing is now completed) also contain 5,000 kinship pairs, including full and half siblings, and MZ and DZ twins.

Using these kinship links, the value of hundreds of NLSY phenotypic variables – from health, cognition, marriage/fertility, labor force, antisocial and delinquent behavior, and many other domains – can be studied biometrically, and generalized to relevant U.S. populations. More than 40 research articles (and 6 theses/dissertations) have been published using earlier versions of these kinship links, applied to standard NLSY design structures. The current links are freely available to researchers through several sources, including the NlsyLinks R package.

In this paper, we present innovative NLSY designs. We begin with a review of the Mother-Daughter-Aunt-Niece (MDAN) design (Rodgers et al, 2008) and expand this to include other relationships simultaneously, including the 5,000 NLSYC first cousins. Following we discuss the potential for limited three-generational designs using the available information about the parents of the original NLSY79 respondents. Finally, we discuss how incorporating a third dataset, (the NLSY97) provides a “phantom mother” design, developed by (age, SES, family, etc) matching of the NLSYC to the NLSY97 respondents, and assigning NLSY79 mothers to NLSY97 respondents across these matches.

----Not submitted below this line----

* Gen1 shown above a timeline
  + DOBs bunched between 19xx and 19xx –
  + Complete Fertility spread out between xx-xx
  + Outcome measures spread out between 1979-2012
* Gen2 shown below a timeline
  + DOBs spread out
  + Incomplete Fertility
  + Outcome measures start at 19xx.
* Combine the two timelines on the same slide
  + Argue that if the two generations are combined, their sum could be more valuable to many research questions than the sum of two independent cohorts.