Comparing Old NLSY Sibling Data to New NLSY Sibling Data:
Sexuality and Fertility Patterns in the NLSY

Joseph Lee Rodgers
Department of Psychology
University of Oklahoma
Norman OK 73019
Jrodgers@ou.edu

Many sibling studies have been conducted using the National Longitudinal Survey of Youth data. The NLSY79 refers to the data from the original cohort of 14-22 year old adolescents collected first in 1979, and followed longitudinally to the present into middle adulthood (this cohort is aged 41-49 in the most recent 2006 release). The NLSYC refers to the data collected from all biological children born to females in the NLSY79, collected from 1986 to the present.

Both surveys include all biological children in a sampled household, and thus have been the basis for many sibling studies. In particular, behavior genetic studies have used the NLSY; these use different levels of genetic relatedness (e.g., twins, full siblings, half siblings, cousins) to estimate models that partition genetic and environmental variance into separate components. However, the different levels of relatedness in both the NLSY79 and NLSYC siblings have never been assessed directly, through survey questions. Rather, kinship relatedness has been inferred from kinship algorithms that use external information about living structure (e.g., "When you were 0, 1, 2, … years old, were you living with your biological father/mother?) or distance relationships (e.g., "How far away did your biological father live from you?") to ascertain genetic relatedness between pairs of siblings.

These algorithms have, apparently, provided kinship links of high reliability and validity. There are at least three sources for this statement. First, internal validity analyses of these links have been consistently positive in matching results from meta-analyses (see Rodgers, 1996 for validity analyses of the NLSY79, and Rodgers, Johnson, & Bard, 2005, for validity analyses for the NLSYC). Second, patterns in many research studies have been highly similar to those from other studies using other data, providing concurrent validity to support the value of these kinship links (e.g., see Rodgers et al, 2006, for several examples, using outcome variables including sexuality and fertility, delinquency, birth weight, education, and achievement). Finally, results invoking different levels of assumptions about the nature of genetic relatedness have consistent patterned in reasonable and interpretable ways. For example, kinship correlations for many different outcome variables have been consistently patterned in relation to both genetic and environmental relatedness.

However, researchers have occasionally criticized these results, because most sibling studies use kinship relatedness that emerges from direct survey questions rather than inferences based on kinship algorithms. For example, asking "is your next older sibling a full or half sibling" appears to some to have higher (face) validity than inferring that relatedness because of the living patterns of these two siblings in relation to biological father and mother during a fixed year. In response, for the first time in 2006, the NLSY Survey included questions to directly ascertain sibling relatedness. Questions in the NLSY79 2006 data were asked of all remaining respondents. Questions in the NLSYC data were only asked of NLSYC respondents who were turned 15 years of age or

older in 2006 (and will continue to be asked as the younger NLSYC respondents pass into this age interval).

In the current study, three analyses are presented. First, I present descriptive analyses of the direct survey questions. These are the first analyses of these new questions, indicating sibling relatedness through direct questions for the first time. Second, I present comparisons of the results of those analyses to results obtained from the kinship algorithms. In other words, this is a matching analysis showing similarities and differences in the number of twins, full siblings, half siblings, cousins, and adoptive siblings, using both direct assessment (now available) and inferred assessment (which has been used in several dozen studies published since the mid 1990's). Finally, I will present some analyses of female measures of age at first intercourse and age at menarche. These outcomes have been studied in a number of research projects using the inferred kinship links, and will provide an excellent opportunity to evaluate similarities and differences in model-based results from the two different approaches to ascertaining kinship relatedness.

Several concluding comments: First, many studies besides those based on behavior genetic methodology use sibling status (e.g., full- versus half-siblings), and so these results have implications far outside just the behavior genetic arena (though the many BG-based studies that have been run using these kinship links provides an excellent opportunity for comparison). Second, though the earlier linking algorithms have occasionally been criticized, and though some view survey indications as a kind of "gold standard," we do not necessarily expect the survey results to improve on the algorithm-based results; this is, rather, an empirical question to be addressed through validity

analyses, and the first validity analysis comparing the two approaches it the one presented here. Third, it is likely that each approach can help support the other. For example, there is substantial missing data that emerges from application of the linking algorithm to the NLSY79 (e.g., one of the most often-used algorithms only classifies around 65% of the available kinship pairs); the direct survey responses can help fill in the missing data. Alternatively, if the direct survey questions do show better results in the validity analyses, the algorithm could be used to fill in missing data from that procedure. Finally, it is certain that the two approaches in combination will provide improvement in both reliability and validity compared to using only one approach.

References

Rodgers, J. L. (1996). NLSY-Youth linking algorithm. Unpublished document.

Rodgers, J. L., Bard, D. E., & Miller, W. B. (2007). Multivariate cholesky models of human female fertility patterns in the NLSY. Behavior Genetics, 37, 345-361.

Rodgers, J. L., Johnson, A., & Bard, D. (2005). NLSY-Children/Young Adult (1986-2000) kinship linking algorithm. Unpublished document.