FAQ for the NlsyLinks Package

William Howard Beasley (Howard Live Oak LLC, Norman)
Joseph Lee Rodgers (Vanderbilt University, Nashville)
David Bard (University of Oklahoma Health Sciences Center, OKC)
Kelly Meredith (University of Oklahoma, Norman)
Michael D. Hunter (University of Oklahoma, Norman)

November 14, 2013

Contents

1   Getting Started

Where is a good place to start?

Welcome to the `NlsyLinks` package, which facilitates research with the NLSY. The initial focus of the package was to assist behavior genetics, but it has been expanded to help with NLSY research in general.

The current FAQ and the NLSY ACE vignette provide overview, and the reference manual documents the package's syntax; the documents are available on CRAN. Furthermore, the articles involving the NLSY kinship links provides the theoretical and academic perspective.

2   Ambiguous twins

What are "ambiguous twins"?

MZ twins share all of the genetic information (*i.e.*, $R = 1$), while DZ twins on average share half (*i.e.*, $R = 0.5$). Sometimes a sibling pair doesn't have enough information for us to classify comfortably as either MZ or DZ. We assign these "ambiguous twins" $R = .75$. Currently there are 12 ambiguous twins in the NLSY79C sample.

Of these 13 pairs, all had close birthdays and were the same gender. 12 pairs are ambiguous because the mother didn't complete an NLSY survey since 1993; the first twin items were presented in 1994 (*e.g.*, `R48257.00`, and `R48260.00`). The mother of the 13th pair (*i.e.*, subjects 864902 and 864903) simply avoided responding to the twin survey items.

Occasionally they mother of twins provided conflicting evidence. Fortunately, these mother were consistent among their most recent responses. For instance, Subjects 392401 and 392402 were indicated DZ in 1998, but MZ in 2000, 2002, and 2004. This pair was assigned $R = 1$.

Gen2 ambiguous twins can be viewed with:

```
subset(Links79Pair, RelationshipPath == "Gen2Siblings" & R == 0.75)

##       ExtendedID Subject1Tag Subject2Tag    R RelationshipPath
## 5200        1460      146001      146002 0.75      Gen2Siblings
## 20230       5658      565901      565902 0.75      Gen2Siblings
## 24698       6639      663901      663902 0.75      Gen2Siblings
## 26363       7111      711101      711102 0.75      Gen2Siblings
## 29752       7913      791406      791407 0.75      Gen2Siblings
## 36439       9596      959601      959602 0.75      Gen2Siblings
## 37729      10012     1001201     1001202 0.75      Gen2Siblings
## 39899      11191     1119103     1119104 0.75      Gen2Siblings
## 40067      11486     1148601     1148602 0.75      Gen2Siblings
## 40154      11733     1173301     1173302 0.75      Gen2Siblings
## 40158      11739     1173901     1173902 0.75      Gen2Siblings
## 42728      12574     1257402     1257403 0.75      Gen2Siblings
```

3   Ambiguous siblings

What are "ambiguous siblings"?

Similar to ambiguous twins, ambiguous siblings are sibling pairs that we cannot comfortably classify as either full-siblings ($R = .5$) or half-siblings ($R = .25$). All siblings in the NLSY79-C/YA dataset share the same biological mother, so for these pairs, the problem is reduced to determining if they share the same biological father. There are two typical reasons for classifying siblings as ambiguous: (a) the relevant items are missing responses, or (b) the existing responses conflict with each other.

For instance, there are at least 194 Gen2 pairs where one sibling explicitly reported they shared a biological father, while the other sibling explicitly reported they did not. These subjects can be viewed with:

```
dsLinks <- Links79PairExpanded
isGen2Sib <- dsLinks$RelationshipPath == "Gen2Siblings"

olderFullYoungerHalf <- (dsLinks$RExplicitOlderSibVersion == 0.5 & dsLinks$RExplicitYoungerSibVersion ==
    0.25)

olderHalfYoungerFull <- (dsLinks$RExplicitOlderSibVersion == 0.25 & dsLinks$RExplicitYoungerSibVersion =
    0.5)


dsLinks[isGen2Sib & (olderFullYoungerHalf | olderHalfYoungerFull), ]
```

Another example occurs when a subject reports they are unsure or if their own responses are inconsistent over the years. These 80 Gen2 pairs can be viewed with:

```
dsLinks[isGen2Sib & (dsLinks$RExplicitOlderSibVersion == 0.375 | dsLinks$RExplicitYoungerSibVersion ==
    0.375), ]
```

When the one perspective provided inconclusive evidence of R, we looked at other perspectives to resolve their relationship.

4 Retaining vs. dropping the ambiguous twins and siblings

I am running ACE models with sibling pairs. Do you recommend including the pairs who are classified as $R = .375$ or $R = .75$? Or should I exclude them from the analyses?

This important issue touches Behavior Genetic concepts and modeling pragmatics. However, this issue typically has an easier resolution than it used to. In the links we released 10 years ago, there were 3,053 Gen2 pairs classified as ambiguous;in our current version, this has been reduced to 610. From one perspective, we are more likely to recommend dropping the ambiguous siblings because there are fewer of them (and therefore less potential gain by including them).

Here's another perspective. Usually if they're missing the data necessary to determine the R value, they're also missing the phenotype, so they'd contribute very little to the analysis anyway. If there's only a few in an R group, it may not be worth including them. Virtually none of the ambiguous twins have phenotype values for both Gen2 siblings.

Our advice to include/exclude an R group also depends on the kind of analysis. Some analyses break up the R values into separate categories (like multiple group SEMs). While some analyses treat R like a continuous variable (like DF analysis, or SEMs with constraint/definition variables). If you're running the former, we're more likely to recommend dropping small R groups, because they're more likely to be estimated poorly (eg, the covariance matrix is more likely to misbehave). If you're running the latter, the estimation is more robust. (Though the estimation's robustness is a different issue that if that R group is a good representation).

We don't recommend blindly dropping the ambiguous twins and siblings in every analysis. For each scenario, the group sizes and phenotypic measurement issues should be considered.

We do recommend running a casual sensitivity test, at the very least. Run different models that include and exclude the small R groups. Hopefully the estimates change in expected ways (*e.g.,*, including ambiguous siblings makes only a small difference) and you don't have to dig deeper. For all analyses, inspect each R group's covariance matrix, especially with for the MZs, which typically is the smallest group.

5 Grant support