**Stata/R script for a between/within twin analysis as outlined in:**

Sjölander, Arvid, Frisell, Thomas and Öberg, Sara. "Causal Interpretation of Between-Within Models for Twin Research" *Epidemiologic Methods*, vol. 1, no. 1, 2012, pp. 217-237. <https://doi.org/10.1515/2161-962X.1015>

This script is solely to create the variables needed for a between/within analysis. Modeling decisions and relevant scripts will vary by analysis:

Stata Script:

\*Pair mean and individual departure from pair mean for any variable (var) of choice

bysort pairid: egen pair\_mean\_var=mean(var)

bysort pairid: egen pair\_n\_var=count(var)

gen dep\_fr\_pair\_mean\_var=var-pair\_mean\_var

R Script:

##### Generate pair means and departure from pair means for between-within analyses (data frame named d1)

d1$pair\_mean\_var <- ave(d1$var,  d1$pairid), FUN=mean)

d1$ dep\_fr\_pair\_mean\_var <- d1$var -d1$ pair\_mean\_var

library(plyr)

d1 <- ddply(d1,.(pairid),transform, pair\_n\_var = NROW(piece))

Use both predictors in the same model to obtain the between (pair\_mean\_var) and the within (dep\_fr\_pair\_mean\_var) estimates. The between estimate should closely resemble the population-based estimate (from the corresponding model that does not specifically model the twin structure of the data), although it will be a slight under-estimate of the population-based effect. For comparison of the reduction in effect size, the population-based effect (ignoring twin structure) should be compared to the within estimate from the twin model.

Write-up by: Malin Ericcson, Ida Karlsson, and Brian Finch