**Mike Allerhand’s guide to deriving attrition weights**

1. Create a dummy variable for response/nonresponse.

2. Fit a logistic regression of that dummy variable predicted by covariates you have available to explain drop-out. If any of the covariates have missing values, impute a value using the mean of that variable so that your logistic regression will be based on the largest possible number of cases.

3. Extract that model's "fitted values" which will be probabilities. In other words use the model to predict the probabilities.

That's how to calculate the p. Then to apply these as weights I'd do:

4. Derive a variable of weights that contains 1/p for each responder and 1/(1-p) for each non-responder.

5. Finally when you come to fit your multiple regression model use that variable as regression weights.

**Here’s an example using data from the British Cohort 1970 following these guidelines. I used a logistic regression model of covariates at age 5 to predict non-response at age 26. What follows is STATA syntax but is hopefully understandable.**

\*complete variables (ie no missing values)

d003 e038

\*where covariates have missing values, these imputed using mean

gen dadclass\_imp=BD2SOC

recode dadclass\_imp (.=3.44)

gen rutter5\_imp=BD2RUTT

recode rutter5\_imp (.=7.80)

gen birthweight\_imp=e012a

recode birthweight\_imp (.=116.5)

\*create non response variable

gen nonresponse26=.

replace nonresponse26=1 if d003<3 & sex==.

replace nonresponse26=0 if sex<3

\*logistic model of age 5 predictors of attrition

logistic nonresponse26 d003 e038 FAC1\_1 dadclass\_imp rutter5\_imp birthweight\_imp

predict fitted

\*deriving an attrition weight

gen attritionweight=.

replace attritionweight =1/fitted if nonresponse26==0

replace attritionweight=1/(1-fitted) if nonresponse26==1