The file S1\_Dataset\_Model.csv contains

link | relative\_date | steps | arm | month | minutes

link is variable for linking the two data sets. It is a randomly generated 3-digit number.

relative\_date is the number of days since the subject started the study for the steps described in that row

steps is the total number of steps taken on a given day

arm is whether the subject was in FB, FB+G or FB+R

month is the month of year (used in our model to adjust for seasonal variances in activity)

minutes is the number of minutes of activity on relative\_date. In our analysis, we required that days have at least 20 minutes of activity but I did not apply that filter to this data set in case someone wants to explore the effect of that rule on our results.

The file S2\_Dataset\_Clinical\_Data.csv contains the clinical variables at enrollment, visits 2 and 3. The two datasets can be linked by the link variable.