**PRI case-control and Family case-only PLINK Datasets**

Summary:

PRI case-control database: 771 subjects, 311 controls and 460 cases

Family case-only database: 345 subjects, 113 fathers, 113 mothers and 119 children

Below describes details and possible alternatives to the way the databases were constructed and number of subjects with PCA codes.

**PRI case-control PLINK Dataset**

There are 1,011 PRI’s in the 5/21/15 database.

460 are cases (dx2 code 2)

311 are controls (dx2 code 5)

240 are neither case nor control (dx2 code 6, 7, 9, 98, plus 4 missing, likely cases)

See below for more information about this dataset.

**Family case-only Database**

There are 113 families that have a mother, father and at least one child who is a case. For this version all subjects who are cases were included, whether they are PRIs or not. This gives 119 children (111 are PRIs, 8 are sibs). Therefore there are 345 subjects in the family case-only database (226 + 119). If only PRIs are used there would be 333 subjects (222 + 111).

Of the 111 PRIs in the database, 5 are not in the PCA “all” database.

See below for more information about this dataset.

***Below is additional information about subjects that were not included in the PRI case-control dataset.***

The 4 missing non-case-controls are the PRIs in the 5/21/15 SNP database that are not in the 8/17/15 demo database. They will probably be coded as cases based on the fact that 1 is coded dx 2 and 3 are coded dx 1 in a 8/29/14 demo database. For now, they are not included in the PRI case-control database.

There are 8 families where an SB1 is the only member of the family who is a case or control (all 8 are cases). 5 of these families have no PRI (assume PRI didn’t pass QC) and 3 of these families have a PRI that is a non-case control (dx2 = 7, 7, 9) and so these PRIs are not included in the case-control PRI database. Possibly these sibs could be used as a “substitute PRI” case-controls (but there are no PCAs for these subjects because PCA was only carried out on PRIs).

The PCA “all” database is made up of 990 PRIs. 18 of the 771 case-control PRIs are not in the PCA “all” database. The 4 PRIs not in the 8/17/15 demo database all have PCA codes.

***Below is additional information about families that was not included in the family case-only dataset.***

Among the 113 families 107 families have only one child (of these 105 are PRI’s and 2 are SB1’s). In the remaining 6 families there are two children, both of which are cases (one is a PRI and one is a sib). Among these 6 families 4 of the families appear to be cases where the two children in the family are twins (the children are reported with the same age at consent).

There are 6 additional siblings (non-PRIs) in these 113 families who are not cases (i.e. not coded with dx2 = 2, they are coded dx2 = 7, 7, 9, 9, 9, 98). They are members of the 105 families with one PRI.

There are 124 families with a mother and a father. Therefore there are 11 families that have a mother and father, but no children who are cases. Of these 11 families 2 of them have one PRI who is currently coded missing because it has no dx2 code (they aren’t in the 8/17/15 demo file), but they will probably both be coded as cases. When they are coded as cases these 2 trios could be added into the family case-only database. Of the remaining 9 families 7 families have one PRI with non-case-control dx2 codes: five 9’s and two 98’s. The remaining 2 families have no children (assume none made it through QC).

There are 11 mothers with no corresponding father. Each family has one child. Ten are PRIs (8 cases, one neither case nor control (dx2 = 9) and one is missing from the 8/17/15 demo file and is probably a case) and one is an SB1 (a case).

There are 4 fathers with no corresponding mother. One of those families has no children, the other three families have a single PRI, two are cases, one is neither case nor control (dx2 = 9).