

## Documentation of GEDI.DAT

File contains information on family structure, covariates, and phenotypes for initial GWAS analysis in GEDI. There are 9882 records in the file, with one record per subject. Of these, 7278 were genotyped and 1128 were MZ cotwins of someone genotyped. Consequently, the sample for which we have genotyped data has N=8406.

The file is structure such that for each of the 5 family types (see FTYPE variable) there are four consecutive records – offspring1, offspring2, mother, father (see IND variable). For the first 4 family types, the offspring are exchangeable. For the 5<sup>th</sup> family type (MIXED) , the offspring are ordered such that the biological offspring is first of the four and the adopted offspring is second. The 6<sup>th</sup> family type (MISC) are single-member families, so all individuals have IND coded as one. The order of the FTYPES in the file is as follows:

FTYPE	# Families in File	# Individuals in File	Comment
MZ	1203	4812	IND=1,2 for offspring, 3=mom, 4=dad
DZ	655	2620	IND=1,2 for offspring, 3=mom, 4=dad
ADOPT	269	1076	IND=1,2 for offspring, 3=mom, 4=dad
BIOL	191	764	IND=1,2 for offspring, 3=mom, 4=dad
MIXED	111	444	IND=1 for bio off, 2 for adopted, 3=mom,4=dad
Total in Fams	2429	9716	This is the first 9716 records in file
MISC	166	166	All have IND = 1
Total		9882	

The file is in fixed field format, so blank is an indication of missing data. There are some other missing value codes though (see below)

Variable	Description	Columns
FAMID	5-digit Family ID #	1-5
SC	2-digit Sib Code	6-7
ID	7-digit unique Individual's ID	1-7
FTYPE	Family Type (1=MZ, 2=DZ, 3=Adopt, 4=Bio, 5=Mixed, 6=Misc)	12
IND	Individual w/i family code (1=first offspring, 2=second offspring 3=mother, 4=father)	14
GSTAT	Genotyping status (2=imputed MZ cotwin, 1=genotyped, 0=not genotyped)	16
SEX	Coded 1 = male, 2=female – this is a covariate	18
COHORT	1=Offspring generation, 2=Parent generation – A covariate	20
ETHNIC	Original ethnicity code – DON'T USE NOW (Miss = 99)	23
WHITE	1 = in White sample for analysis, 2=Not in sample	25
AGE	Age at time of clinical assessment	27-31

**Clinical Phenotypes (all missing = 96):**

NIC_FAC	Nicotine Factor	33-40
CON_FAC	Alcohol consumption factor	42-49
DEP_FAC	Alcohol dependence factor	51-58
DRG_FAC	Drug factor	60-67
BD_FAC	Behavioral Disinhibition factor	69-76
EXT_FAC	Externalizing factor	78-84

**Psychophysiological Phenotype (Missing = 996):**

P3_AMP	P3 Amplitude	86-94
--------	--------------	-------

**Principal Component Covariates:**

PC1	1 <sup>st</sup> Principal component	97-103
PC2	2 <sup>nd</sup> Principal component	105-111
PC3	3 <sup>rd</sup> Principal component	113-119
PC4	4 <sup>th</sup> Principal component	121-127
PC5	5 <sup>th</sup> Principal component	129-135
PC6	6 <sup>th</sup> Principal component	137-143
PC7	7 <sup>th</sup> Principal component	145-151
PC8	8 <sup>th</sup> Principal component	153-159
PC9	9 <sup>th</sup> Principal component	161-167
PC10	10 <sup>th</sup> Principal component	169-175