

N.B.: For binary phenotypes cases are always coded 2 and controls coded 1.

LIST OF PHENOTYPES

1. Cigarettes per day

- a. Average number of cigarettes smoked per day, either as a current smoker or former smoker. Individuals who either never smoked, or on whom there is no available data (e.g., someone was a former smoker but former smoking was never assessed) will be set to missing.
- b. For studies that collect a quantitative measure of CPD, where the respondent is free to provide any integer (e.g., 13 CPD) we will bin responses as follows:

1 = 1-5

2 = 6-15

3 = 16-25

4 = 26-35

5 = 36+

For studies that already have pre-defined bins, which are different from ours, we will prefer their existing bins.

- c. Don't include information about pipes/cigars/chew, or other non-cigarette forms of tobacco use.
- d. Cigarettes per day is almost always measured with a single question:
 - i. *How many cigarettes do you smoke per day?*
 - ii. *How many cigarettes did you smoke per day?*

2. Smoking initiation (Smoker versus nonsmoker)

- a. This is a binary phenotype. Code "2" for everyone in the study who reports ever being a regular smoker in their life (current or former). Code "1" for everyone who denies ever being a regular smoker in their life. This phenotype is not available in studies that only address current smoking and ignore former smoking.
- b. Don't include information about pipes/cigars/chew, or other non-cigarette forms of tobacco use.
- c. This can be measured in a variety of ways:
 - i. *Have you smoked over 100 cigarettes over the course of your life?*
 - ii. *Have you ever smoked every day for at least a month?*
 - iii. *Have you ever smoked regularly?*
 - iv. *Do you smoke?*

3. Smoking Cessation (Current versus former smoker)

- a. Binary phenotype with current smokers coded as "2" and former smokers coded as "1".
- b. Don't include information about pipes/cigars/chew, or other non-cigarette forms of tobacco use.
- c. Usually measured through a combination of questions, including:
 - i. *Do you currently smoke?* and *Have you ever smoked regularly?*
 - ii. *Do you smoke?* and *Have you smoked over 100 cigarettes in your entire life?*

4. Age at which an individual started smoking regularly

- a. The age at which an individual first became a regular smoker
- b. Don't include information about pipes/cigars/chew, or other non-cigarette forms of tobacco use.
- c. This can be measured in a variety of ways:
 - i. *At what age did you begin smoking regularly?*
 - ii. *How long have you smoked?* combined with *What is your current age?*

5. Drinks per week in individuals who are active drinkers

- a. The average number of drinks a subject reports drinking each week, aggregated across all types of alcohol. If a study recorded binned response ranges (e.g., 1-4 drinks per week, 5-10 drinks per week) we will use the midpoint of the range. So if an individual reports 1-5 DPW, we assume they drink 2.5 DPW on average.
- b. This can be measured in a variety of ways:
 - i. *In the past week, how many alcoholic beverages did you have?*
 - ii. *Thinking about the past year, on the average how many drinks did you have each week?*
- c. For studies that collect drinks per week separately for different types of alcoholic beverage (e.g., beer, wine, spirits), please contact Scott for details on how to collapse across beverage types.

6. Drinker versus Non-drinker

- a. If a respondent reports drinking during the timeframe under study (e.g., last week, month, year), then they are coded “2”. If they report that they did not drink, they are coded “1”.
- b. This can be measured in a variety of ways:
 - i. *In the past week (or month, or year) how many drinks did you have on average each week? (Those reporting zero drinks are considered non-drinkers. Those reporting 1 or more are considered drinkers.)*
 - ii. *Do you currently drink alcohol?*
 - iii. *Thinking about the last week, on how many days did you drink alcohol?*

7. Binge Drinking (in everyone)

- a. This is a complex phenotype that may be measured in a wide variety of ways in different studies. The point is to have a phenotype that measures pathological drinking. Pragmatically, we propose a binary variable, with study-specific definitions, such as:
 - i. Consuming 5+/4+ standard drinks in one sitting (males/females)
 - ii. *In the last 4 weeks, did you drink so much that you felt very intoxicated (drunk)?*
- b. Binge drinkers are coded as “2” and non-binge-drinkers are coded as “1”.
- c. Please for this phenotype include all individuals who have responded to the binge drinking-related questions, regardless of lifetime drinking status. We don’t care if an individual is a lifetime teetotaler, as long as we know their answers to the binge drinking-related items.

Only for studies that have measures of lifetime drinking (e.g., “Have you drunk alcohol in your life”)

8. Binge Drinking in lifetime drinkers

- a. Defined in the same way as for “7. Binge Drinking” but any individual who has never used alcohol (aka, a teetotaler) would be coded as missing and excluded from the analysis.

COVARIATES

1. All Phenotypes
 - a. Sex
 - b. Age
 - c. Genetic principal components (not necessary if using linear mixed model with genetic kinship)
 - d. Other study specific covariates (e.g., cohort, case/control status)
2. Additional covariates for specific phenotypes
 - a. Drinks per week & Binge Drinking
 - i. Height and weight
 - b. Cigarettes per day
 - i. Current versus former smoker status

TRANSFORMATIONS

Drinks per week will be left-anchored at 1 and log-transformed (natural log).

***** Considerations affecting phenotype definition that are part of the analysis plan*****

SEX

We will include sex as a covariate in the association models, and not stratify by sex.

ANCESTRY STRATIFICATION

We conduct all association analyses stratified by ancestry.

NON-ADDITIVE MODELS

We will impute recessive and dominance models summary statistics from the additive model results. This can all be done centrally, such that individual studies have only to run an additive model and submit results