

Table S1. Example of cohort-specific information collected, harmonization potential, and processing algorithms for the variable ‘Number of estimated pack years of the participant’. Note that harmonization processing accounted for longitudinal cross-checks for consistency among data collection events. For example, if a participant responded “Yes” to “Ever smoked cigarettes” at baseline or preceding data collection events, then they would be considered as “Yes” in later follow-ups.

DataSchema variable: Number of estimated pack-years of the participant. Format: Continuous (decimal) Unit: pack-years				
# datasets	Cohort-specific variables	Cohort-specific variable format	Harmonization potential	Processing Algorithm
3	1. Smoking pack-years	1. Continuous	Complete	Direct mapping from [Smoking pack years]
1	1. Ever smoked cigarettes 2. Avg #cig/day 3. Total years smoked cigarettes	1. Yes No 2. Continuous 3. Continuous	Complete	Values are coded to 0 if [Ever smoked cigarettes] = No ELSE ([Avg #cig/day]/20)*[Total years smoked cigarettes]
1	1. Ever smoked cigarettes 2. Pack-years	1. Yes No 2. Continuous	Complete	Values are coded to 0 if [Ever smoked cigarettes] = No ELSE Direct mapping from [pack years]
1	1. Ever smoked cigarettes 2. Cigarettes-year, former smoker 3. Cigarettes-year, current smoker	1. Yes No 2. Continuous 3. Continuous	Complete	Values are coded to 0 if [Ever smoked cigarettes] = No ELSE If [Cigarettes-year, former smoker] = Missing • [Cigarettes-year, current smoker]/(20*365) ELSE If [Cigarettes-year, former smoker] ≠ Missing • [Cigarettes-year, former smoker]/(20*365)
1	1. Age of participant 2. Current cigarette smoker 3. Past cigarette smoker 4. Cigarette smoking start age 5. Cigarette smoking stop age 6. Avg # cig/day past smoker 7. Avg # cig/day current smoker	1. Continuous 2. Yes No 3. Yes No 4. Continuous 5. Continuous 6. Continuous 7. Continuous	Complete	Values are coded to 0 if [Current cigarette smoker] = No & [Past cigarette smoker] = No ELSE If [Current cigarette smoker] = No & [Past cigarette smoker] = Yes • ([Cigarette smoking stop age] – [Cigarette smoking start age]) * ([Avg # cig/day past smoker]/20) ELSE If [Current cigarette smoker] = Yes • ([Age of participant] – [Cigarette smoking start age]) * ([Avg # cig/day current smoker]/20)
2	1. Qty cig smoked, when daily 2. Total yrs daily smoker	1. Categorical 2. Continuous	Impossible	N/A
1	Information not available	N/A	Impossible	N/A