1. Subject refused to answer or don’t know treated as missing

2. For Annual Family Income, over or under $2000 treated as missing

3. For Annual Family Income, create 4 categories: $0-$14,999; $15000-$34999; $35000-$64999; Over $65000.

4. For question ALQ120Q - How often drink alcohol over past 12 mos, people can answer 0 which means they never drink alcohol, and their corresponding information in ALQ130: Avg # alcoholic drinks/day -past 12 mos will be missing. We give these missing values a 0 in ALQ130.

5. Systolic Blood pressure have 4 readings; we use the average of these four readings.

6. For depression, DPQ020, we define subject that answers several days or more than half the days or nearly every day as having depression.

7. History of cardiovascular disease is defined as participants has history of coronary heart disease or history of stroke, if one of them is missing, its value defined as the un-missing value, if both are missing, then its value is missing.

8. Menopausal status: if RHQ060 - Age at last menstrual period has a value, then define it as Postmenopausal, if RHQ020 - Age range at first menstrual period has a value and RHQ060 - Age at last menstrual period is missing, then define it as premenopausal.

9. We searched records of prescription medication use and identified users of aspirin, atovarstatin, ibuprofen, opium, statin, valsartan users. If we did not find any record for a medication for a person, then it is treated as non-user.

10. For total nutrition intake variables (TCAL, TCARB ….), we averaged day 1 and day 2. If one is missing, then the un-missing value is used. If both missing, then missing.

11. For dietary supplement use, if either day 1 or day 2 they used dietary supplement, then they are using dietary supplement overall. If neither day 1 and day 2 they used dietary supplement, they are not using dietary supplement. If one is missing, the un-missing value will be used. If both missing, then missing.

12. If participants answered OCD150 - Type of work done last week to be not working at a job or business, their values in OCQ180 - Hours worked last week at all jobs will be missing. In this case, we give these missing values a 0 in OCQ180.

13. For occupation, we categorized participants into 3 categories: non-worker (0 hours a week), Part time worker (1-30 hours a week), Full time worker (>=31 hours a week)

14. For physical activity, if the participants either do a vigorous recreational activity or a moderate recreational activity, then they are defined as they do physical activity. If they neither do a vigorous recreational activity and a moderate recreational activity, then they are defined as they don’t do physical activity. If one is missing, the un-missing value will be used. If both missing, then missing.

15. People whose sedentary lifestyle minutes answered >=1000 is treated as missing because they are probably including the time they sleep and that is not plausible.

16. People who answered they never smoked cigarettes regularly or not smoked at least 100 cigarettes in life or smoked at least 100 cigarettes in life but quit are categorized as non or light smoker. People who answered they smoked at least 100 cigarettes in life and continue smoking are categorized as follows: if >67 cigarettes (40 pack year) then heavily smoker, if <=67 cigarettes (40 pack year) then moderate smoker.

17. For individual food components, we averaged day 1 and day 2. If one is missing, then the un-missing value is used. If both missing, then missing.

18. Fruit consumption is defined as CITMLB+OTHER fruit, excluding juice.

19. Sea food consumption is defined as Seafood high in n-3 fatty acid + Seafood low in n-3 fatty acid.