Questions for June 29 meeting:

1. Definition of red meat and processed meat

Red meat: mammalian meat

Processed meat: white or red meat preserved by smoking, curing, or adding chemical compounds (looks like white meat is included)

My question is, because the systematic review conducted separate analyses for **unprocessed red meat**, processed meat, and mixed unprocessed red and processed meat, which are we focusing for our research?

1. Outcome of interest: all-cause mortality, cardiovascular mortality, cardiovascular disease, stroke, MI, type 2 diabetes, anemia, quality of life, satisfaction with diet.

Question: Are we just focusing on **all-cause mortality**?

1. Can you give me a template which include what kind of information I should pay attention to when reading the papers? Like a very general point by point guidelines like formula form or adjusting variables**? How much detail I should pay attention to?**
2. What’s our next steps?

Summarize all the information, find variables in NHANES.

Questions for July 20th meetings**:**

**The NHANES I clicked is 2009-2010, because mortality data linked from CDC is up until 31 December 2019, we have 9 years of follow up.**

21. Could not find time to event information? I mean 0, 1 death event and follow-up time for each person in NHANES.

Solved: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm

23. Does accident constitute all-cause mortality?

YES

24. How many participants and death are we going to examine? The population size from literature examining this relationship using NHANES range from 9000-20000. For example, for NHANES 2009-2010, there are 10537 participants, among them 4027 are ineligible because age<18, leaving a total of 6510 eligible, and among them 861 death occurred. If these participants and death is not enough, then should we consider 2009-2010 NHANES and 2010-2011 NHANES and more NHANES combined?

**2000-2015. stay to the same format. 65000, 8000 death.**

1. Which age group are we focusing on? For example, if we are focusing on children 6-19, then education variables DMDEDUC3 will be used, if we are focusing on adults 20+, then DMDEDUC2 will be used. This also determines marital status because children don’t marry and also determine other things that children do not do. Also, mortality data for less than 18 is ineligible.

Adults

2. For income, what kind of income do we use? Annual household income or Annual family income, or monthly?

Annual family

3. For systolic blood pressure, which reading are we using? 1st, 2nd, 3rd, 4th?

Use 1st one and replace 2nd or 3rd or 4th if missing.

4. Can we discuss to determine the variables for alcohol drinking? I used Avg # alcoholic drinks/day -past 12 months

Use this variable.

5. Can we discuss to determine the variables for smoking?

https://shouldiscreen.com/English/pack-year-calculator

6. Can we discuss to determine occupation variables?

OCD241 - Occupation group code: current job

7. Can we discuss to determine history of cardiovascular disease? I only used coronary heart disease but I think this is only part of cardiovascular disease.

Combine coronary heart disease and stroke

8. Can we discuss to determine history of depression in NHANES? They just asked about lower mood…..

PQ020 - Feeling down, depressed, or hopeless

9. There are two dietary 24-h recalls in NHANES, which one do we use? Day 1 or Day 2?

Use the average. Combine Day 1 and Day 2.

10. Do you already have code for manipulating NHANES data, I mean merging combining dataset together, and most importantly, any analytical requirement that need to be paid attention to, like is sample weights needed to be included in cox regression? need a template code for this? Which software do we use? Need more information on this? SAS or R? I prefer R.

**Chirag have some code for this. We don’t need sample weights. We prefer R.**

11. It looks like we need to link food codes to individual food and create variables such as vegetables intake (vegetables intake), fruit intake, red meat intake (red meat includes many sources) so we need to determine on the exact definition of these things. The problem here is that food is mixed.

Consider variation in definition. Dena will find how to define and combine food in literature.

12. There are so many eggs? Which variable do we need to combine? Need help with this. Exact definition.

13. It looks like we need to link prescription medications with drug information to help find drug use.

We do this.

14. Could not decide menopausal status (premenopausal, postmenopausal)? Only have information about Age at last menstrual period

RHQ060 - Age at last menstrual period 1 year after current age then postmenopausal.

15. Could not decide hormone therapy users? In literature: in postmenopausal women (not taking hormone therapy, taking hormone therapy. But in NHANES, {Have you/Has SP} ever used female hormones such as estrogen and progesterone? Please include any forms of female hormones, such as pills, cream, patch, and injectables, but do not include birth control methods or use for infertility.

This is good.

16. For socioeconomic status, how to determine poor, middle income, high criteria using PIR.

We discuss later.

17. Wealth score is created by Using multiple correspondence analysis, the wealth score was created based on occupation, house ownership, house structure, house size, having a bath in the residence, as well as a personal car, motorbike, black/white TV, color TV, refrigerator, freezer, vacuum, and washing machine. Can we remove wealth score? Only in one literature.

Delete it.

18. NO find for Consumption of trans-fat, family history of cancer, could not find rural or urban residence?

Dena will find trans-fat, we don’t cancer information, we don’t need urban resident.

19. In an effort to control for severity of illness, we included self-reported health status as well as comorbid conditions. A variety of conditions were assessed in the NHANES II. Comorbidities were positive responses in the baseline interview to questions regarding whether a physician ever told the patient that he or she had each of the following conditions: cirrhosis, diabetes, high blood pressure, heart failure, heart attack, stroke, hardening of the arteries, rheumatic fever, rheumatic heart disease, heart murmur, ulcer (peptic, stomach, duodenal), chronic enteritis, ulcerative colitis, spastic colon or mucous colitis, gallstones, hepatitis, yellow jaundice, pleurisy, low blood pressure, cataracts, glaucoma, thyroid disease, polio or paralysis, goiter, hiatus hernia of the diaphragm, cancer, benign tumor, trouble with blood not clotting properly, loss of blood from stomach or bowels, nervous breakdown, neck injury, back injury, anemia, arthritis, gout, asthma, chronic bronchitis, emphysema, tuberculosis, and kidney problems. The Charlson Comorbidity Index was calculated from the responses to these questions. **This is very complex; can we remove this comorbidity index?**

**Remove**

20. Years of entering cohort is based on year of NHANES?

Yes.2009-2010 record it as 2009.

22. How we define unprocessed red meat?

Questions for July 27th meeting:

(1) **Smoking:** Could not decide the variable for **smoking:** only have

Age started smoking cigarettes regularly

Age last smoked cigarettes regularly

Avg # cigarettes/day during past 30 days

During the past 5 days, on the days {you/he/she} smoked, how many cigarettes did {you/he/she} smoke each day?

<https://shouldiscreen.com/English/pack-year-calculator>

Combine SMQ040, SMQ050Q, SMQSMD650

Current non-smoker, current smoker with categories, people who quit smoking

(2) **Occupation:** NHANES 2015-2016 does not have OCD241 - Occupation group code: current job

Remove 2015-2016

(3) **Sleep:** NHANES Sleep hours format for 2015-2016 is different than NHANES 2007-2014, so I change the format for 2015-2016 and using 2007-2014 format.

No issue anymore

(4**) Total nutritional intake:** Someone have dietary day one total energy but not dietary day two, if they have both, then average used, if they have only one, then that value is used.

Yes we do it

(5) **Physical Activity:** NHANES 2011+ does not have: Hours watch TV or videos past 30 days, 0:less than 1 hour 1: 1 hour…..

Remove those variables

(6) **Dietary supplement intake:** Because we have day 1 and day 2 dietary recalls, I combined this information, if any day they have a dietary supplement intake no matter day 1 or day 2, then they are categorized as having dietary supplement intake overall.

Yes

(7) **Multivitamin user:** Because we have day 1 and day 2 dietary recalls, I combined this information, if any day they have a multivitamin intake no matter day 1 or day 2, then they are categorized as having multivitamin intake overall. Query vocabulary: \*VITAMIN\*

Yes

(8) Unit of every variable

(9) Systolic blood pressure has 4 readings, I used average of four readings.

Yes

(10) I think right now the most important question is how to define:

Consumption of fruits

Consumption of Cruciferous vegetables

Consumption of fish

Consumption of unprocessed poultry

Consumption of unprocessed red meat

Consumption of legumes

Consumption of total dairy

Consumption of nuts and

seeds

Consumption of eggs

Consumption of processed meat

Consumption of whole grain

Consumption of whole bread

Consumption of total meat

Consumption of coffee

Consumption of desert/sweet

Consumption of cheese

Consumption of beef

Consumption of methionine

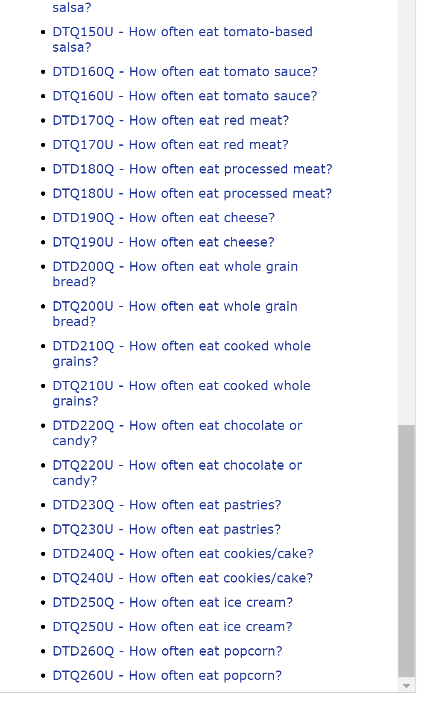
Consumption of berries

Consumption of blood dishes

In terms of short food code description (DRXFCSD):



We also have how often information:



But we will not have gram or kcal information by using food code.

Questions for August 10th meeting:

**Our focus will be age 20-79, this is because some variables only have information for 20+, and age is top coded at 80, which means >= 80 years of age is coded as 80. Also, some people are just ineligible for follow up because of death linkage issues, so there are 21852 participants and 2022 death without considering missing data.**

**First, the problem is that we don’t have food frequency information (except for fish only) because food frequency is only available for NHANES 2009-2010 but not for other years. This means that we can only use Food Patterns Equivalent Database. Let’s discuss food variable first.**

(1) Consumption of fish

In NHANES, we only have frequency information about consumption of fish (that means we don’t have frequency information about consumption of other foods)

In NHANES, we have Shellfish: Clams, crabs, crayfish, lobsters, mussels, oysters, scallops, shrimp, other shellfish, other unknown shellfish eaten

Fish: breaded fish products eaten, tuna, bass, catfish, cod, flatfish, haddock, mackerel, perch, pike, pollock, porgy, salmon, sardines, sea bass, shark, swordfish, trout, walleye, other fish, other unknown fish eaten

On the other hand, in Food Patterns Equivalent Database, we have ounce equivalent information about seafood, which can be combined from seafood high in n-3 fatty acid+ sea food low in n-3 fatty acid.

Which one do we use, in terms of consistency, I suggest only using ounce equivalent and disregard frequency information?

(2) Total fruit includes citrus, melons, and berries, other fruits and fruit juice

(3) Total vegetables include dark green vegetables, total red and orange vegetables, other red and total starchy vegetables and other vegetables.

(4) Consumption of red meat: Can we use Meat variable to define red meat? Meat variable Beef, veal, pork, lamb, and game meat; excludes organ meat and cured meat (oz. eq.) (5) We have consumption of legumes

(6) Total dairy includes consumption of milk, yogurt and cheese

(7) Consumption of processed meat: Can we use cured meat to define processed meat? Cured meat includes Frankfurters, sausages, corned beef, and luncheon meat that

are made from beef, pork, or poultry (oz. eq.)

(8) We have whole grain

(9) We don’t have bread, coffee, desert and sweet, and we remove them

(10) We have consumption of cheese

(11) We have consumption of poultry

(12) We have consumption of nuts and seeds

(13) We have consumption of eggs

(14) We have consumption of total meats: total meats defined as meat+ cured meat+ organ meat + poultry + from seafood high in n-3 fatty acid+ sea food low in n-3 fatty acid.

(15) We don’t have consumption of beef, methionine, blood dishes in detail

(16) We don’t have consumption of berries in detail

(17) Because Meat variable is the summation of beef+pork+lamb+veal+game meat, in order to create a variation in the definition of red meat variable, I created four other meat variable: beef+veal; beef+veal+pork; beef+veal+lamb; beef+veal+pork+lamb. Creating these variables has limitations because we search words like: BEEF, PORK, VEAL, LAMB (case insensitive), however, some food have red meat contained in it but omitted by description of food, such as **Chili con carne with beans**, this food has no BEEF or PORK or VEAL or LAMB in the description but this food is considered having grounded beef and PF\_MEAT is non-zero.

(18) Can we discuss the missingness:

Age: 20+&eligible 21852, 2022 death

**ALQ130: 5455 missing. This is the best I can get.**

I combined ALQ120Q - How often drink alcohol over past 12 months and ALQ130 to create a new variable ALQ130.

**BPQ050A: 15000+ missing: taking prescribed medicine for HBP**

I could not do anything about it. I don’t think imputation make sense, so I suggest that we remove this variable.

**RHQ060(**age at last period missing much than others) premenopausal, postmenopausal

This variable has 50% missing for women, so we cannot create menopausal status effectively. In complete cases, 5000 women will be cut into 2500.

**OCD241: 9326 missing values for occupation group code (to create cleric worker, manual worker…)**, **so I used other variables**

Solution: Type of work done last week have fewest missing values and can use it to combine with Hours worked last week at all jobs to create non-worker, light worker, moderate worker, heavily worker. Then the missing value could be reduced to **1175.** How to create that variable?

**PAD660: 17195 missing values for Minutes vigorous recreational activities, so I used other variables**

Solution: I used Vigorous recreational activities (Yes or No), and Moderate recreational activities (Yes or No), to create a new variable called Activity (Yes or No), then the missing values could be reduced to **0.**

**SMQ040: 12040 missing so many missing values**

**SMD650: 17024 missing so many missing values**

**Solution:** I will categorize participants into 3 categories, (1) Light or Non-smoker includes people who never smoked regularly, people who smoked less than 100 cigarettes, and people who smoked more than 100 cigarettes but quit smoking (2) Moderate smokers includes people who smoked more than 100 cigarettes and are current smoker and smoked <67 cigarettes a month (3) Heavy smokers includes people who smoked more than 100 cigarettes and are current smoker and smoked>=67 cigarettes a month. In this way, the missing values could be reduced to **50**

* **0 Pack Years – Never Smokers**
* **1 – 20 Pack Years – Light Smokers**
* **1 – 40 Pack Years – Moderate Smokers**
* **More than 40 Pack Years – Heavy Smokers**
* **20 cigarettes per pack**

40 pack years = 67 cigarettes per month

https://www.pdcenterlv.com/blog/what-is-a-cigarette-smoking-pack-year/#:~:text=Pack%20Years%20and%20Smoker%20Classification&text=0%20Pack%20Years%20%E2%80%93%20Never%20Smokers,40%20Pack%20Years%20%E2%80%93%20Heavy%20Smokers

**Complete cases analysis: 1041 death 11006 people**

(19) Refused or Don't Know will be treated as missing values.

(20) Borderline diabetes treated as Yes or No?

(21) 0 Not at all 4163 4163

1 Several days 957 5120

2 More than half the days 234 5354

3 Nearly every day 216 5570 for depression?

(22) 1 Excellent, 654 654

2 Very good, 1747 2401

3 Good, 2631 5032

4 Fair, or 1270 6302

5 Poor? 255 6557

(23) How we quantify OCQ180 and create categories?

(24) Can we calibrate the unit of measure? And can we discuss how to categorize each variable one by one?

(25) How can we conduct specification curve? Any packages or code example for this?

The below questions are invalid:

(1) Let’s define dietary variables:

Consumption of fruits

[During the past month], how often did {you/SP} eat fruit? Include fresh, frozen or canned fruit. Do not include juices. [You can tell me per day, per week or per month.]

Q: Do we need to include juices for consumption of fruit?

(2) Consumption of vegetables

Q: How do we define consumption of vegetables? In NHANES, we have DTD090Q - How often eat leafy/lettuce salad?, DTD100Q - How often eat fried potatoes?, How often eat non-fried potatoes?, DTD120Q - How often eat beans?, DTD130Q - How often eat other vegetables? [During the past month], not including what you just told me about [lettuce salads, potatoes, cooked dried beans], how often did {you/SP} eat other vegetables? [You can tell me per day, per week or per month.] It looks like NHANES define vegetables as salad+potatoes+beans+other vegetables?

(3) Consumption of fish

Q: How do we define consumption of fish or we define seafood?

In NHANES, we have Shellfish: Clams, crabs, crayfish, lobsters, mussels, oysters, scallops, shrimp, other shellfish, other unknown shellfish eaten

Fish: breaded fish products eaten, tuna, bass, catfish, cod, flatfish, haddock, mackerel, perch, pike, pollock, porgy, salmon, sardines, sea bass, shark, swordfish, trout, walleye, other fish, other unknown fish eaten

Do we need Fish or We need Fish+Shellfish=Seafood?

We use Yes or No, or we use times/week?

(4) We use times/week, or times/day or times/year? If there are multiple vegetables or fish, are we using the sum? We also have Yes or No information for fish.

(5) Consumption of unprocessed poultry

We do not have this information, unless using individual food and query

(6) Consumption of legume

It looks like legume is approximately beans, and is a type of vegetables, do we still need it? If Yes, then I will use consumption of beans.

(7) Consumption of total dairy

The closest I found is DTD030Q - How often drink milk or on cereal? Can we use it?

(8) Consumption of nuts and seeds

We do not have this information in NHANES unless using individual food and query

(9) Consumption of eggs

We do not have this information in NHANES unless using individual food and query

(10) Consumption of whole grain

I only found consumption of cooked whole grains, including brown rice or other cooked whole grains, such as bulgur, cracked wheat, or millet? But not include white rice. Can we use it?

(11) Consumption of total meat

Looks like we only have red meat + processed meat and seafood (such as bacon, lunch meats, or hot dogs?) Because we do not have chicken, turkey, can we still define total meat?

(12) Consumption of coffee

We have How often drink sweetened coffee/tea? Can we use it?

(13) Consumption of desert/sweet

We have chocolate or candy, pastries, cookies/cake, ice cream, popcorn, Can we combine them to define desert and sweet?

(14) We don’t have Consumption of beef, Consumption of methionine, Consumption of berries, Consumption of blood dishes information. Can we remove them? Only appear once in literature.