summary\_checkin

Team LSD: Adnan, Jenni, Stephen

February 13, 2019

Table of Contents

# Per- and polyfluoroalkyl substances (PFAS) and Body Mass

## 26 February 2019 Check-In

library(tidyverse)

## Warning: package 'dplyr' was built under R version 3.5.2

library(RNHANES)  
library(survey)

## Warning: package 'survey' was built under R version 3.5.2

## Warning: package 'survival' was built under R version 3.5.2

## NHANES Dietary Interviews

*Reference*: <https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR1IFF_I.htm>

Detailed information about each food/beverage item (including the description, amount of, and nutrient content) reported by each participant is included in the Individual Foods files.

* Drinking water variables:
  + Total plain water drank yesterday (gm)
  + Total tap water drank yesterday (gm)
  + Total bottled water drank yesterday (gm)
  + Tap water source
* Fish consumption during past 30 days (by type)

## PFAS

### Load and Inspect PFAS data

NHANES Codebook References: <https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/PFAS_H.htm> <https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/PFAS_I.htm>

pfas\_load <- nhanes\_load\_data("PFAS\_I", "2015-2016", demographics = TRUE)  
  
pfas\_survey <- nhanes\_survey\_design(pfas\_load, "WTSB2YR")  
  
pfas\_data <- nhanes\_load\_data("PFAS\_I", "2015-2016", demographics = TRUE) %>%   
 janitor::clean\_names() %>%   
 rename(pfdea = lbxpfde, pfhxs = lbxpfhs, me\_pfosa\_acoh = lbxmpah, pfna = lbxpfna, pfua = lbxpfua, pfdoa = lbxpfdo, n\_pfoa = lbxnfoa, sb\_pfoa = lbxbfoa, n\_pfos = lbxnfos, sm\_pfos = lbxmfos)

### 2015-2016 PFAS summary table

#pfas\_summary <-  
# pfas\_data %>%  
# select(seqn, pfdea, pfhxs, me\_pfosa\_acoh, pfna, pfua, pfdoa, n\_pfoa, sb\_pfoa, n\_pfos, sm\_pfos) %>%   
# gather(key = "analyte", value = "value", pfdea:sm\_pfos) %>%   
# group\_by(analyte) %>%   
# na.omit() %>%   
# summarise(n = n(),   
# mean = mean(value),  
# sd = sd(value))  
  
  
pfas\_summary <-  
 pfas\_load %>%  
 select(SEQN, LBXPFDE, LBXPFHS, LBXMPAH, LBXPFNA, LBXPFUA, LBXPFDO, LBXNFOA, LBXBFOA, LBXNFOS, LBXMFOS) %>%   
 gather(key = "variable", value = "value", LBXPFDE:LBXMFOS) %>%   
 group\_by(variable) %>%   
 na.omit() %>%   
 summarise(n = n(),   
 mean = mean(value),  
 sd = sd(value))  
  
  
# Weighted  
  
pfas\_inputs <- as.data.frame(matrix(c(  
 # CYCLE COLUMN COMMENT WEIGHTS  
 "2015-2016", "LBXPFDE", "LBDPFDEL", "WTSB2YR",  
 "2015-2016", "LBXPFHS", "LBDPFHSL", "WTSB2YR",  
 "2015-2016", "LBXMPAH", "LBDMPAHL", "WTSB2YR",  
 "2015-2016", "LBXPFNA", "LBDPFNAL", "WTSB2YR",  
 "2015-2016", "LBXPFUA", "LBDPFUAL", "WTSB2YR",  
 "2015-2016", "LBXPFDO", "LBDPFDOL", "WTSB2YR",  
 "2015-2016", "LBXNFOA", "LBDNFOAL", "WTSB2YR",  
 "2015-2016", "LBXBFOA", "LBDBFOAL", "WTSB2YR",  
 "2015-2016", "LBXNFOS", "LBDNFOSL", "WTSB2YR",  
 "2015-2016", "LBXMFOS", "LBDMFOSL", "WTSB2YR"  
), ncol = 4, byrow = TRUE), stringsAsFactors = FALSE)  
  
names(pfas\_inputs) <- c("cycle", "column", "comment\_column", "weights\_column")  
  
weighted\_median <-  
 nhanes\_quantile(pfas\_load, pfas\_inputs, quantiles = 0.5) %>%   
 select(column, weighted\_median = value)

## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.  
  
## Warning in callback(nhanes\_data, ret): No detection limit found from the  
## summary tables. Falling back to inferring detection limit from the fill  
## value.

weighted\_mean <-  
 nhanes\_survey(svymean, pfas\_load, pfas\_inputs, na.rm = TRUE) %>%   
 select(column, weighted\_mean = value)  
  
  
  
  
codes <- data.frame(  
 variable = c("LBXPFDE", "LBXPFHS", "LBXMPAH", "LBXPFNA", "LBXPFUA", "LBXPFDO", "LBXNFOA", "LBXBFOA", "LBXNFOS", "LBXMFOS"),  
 analyte = c("pfdea", "pfhxs", "me\_pfosa\_acoh", "pfna", "pfua", "pfdoa", "n\_pfoa", "sb\_pfoa", "n\_pfos", "sm\_pfos"))  
  
  
pfas\_table <-  
 pfas\_summary %>%   
 left\_join(weighted\_mean, by = c("variable" = "column")) %>%   
 left\_join(weighted\_median, by = c("variable" = "column")) %>%   
 left\_join(codes, by = "variable") %>%   
 subset(select = c(1, 7, 2, 3, 4, 5, 6))

## Warning: Column `variable` joining character vector and factor, coercing  
## into character vector

pfas\_table %>%  
 knitr::kable(digits = 2)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variable | analyte | n | mean | sd | weighted\_mean | weighted\_median |
| LBXBFOA | sb\_pfoa | 1993 | 0.07 | 0.02 | 0.07 | 0.07 |
| LBXMFOS | sm\_pfos | 1993 | 1.94 | 1.88 | 2.05 | 1.50 |
| LBXMPAH | me\_pfosa\_acoh | 1993 | 0.17 | 0.27 | 0.17 | 0.07 |
| LBXNFOA | n\_pfoa | 1993 | 1.81 | 1.63 | 1.90 | 1.50 |
| LBXNFOS | n\_pfos | 1993 | 5.10 | 6.84 | 4.72 | 3.20 |
| LBXPFDE | pfdea | 1993 | 0.26 | 0.45 | 0.24 | 0.10 |
| LBXPFDO | pfdoa | 1993 | 0.07 | 0.01 | 0.07 | 0.07 |
| LBXPFHS | pfhxs | 1993 | 1.61 | 1.75 | 1.74 | 1.20 |
| LBXPFNA | pfna | 1993 | 0.78 | 0.70 | 0.76 | 0.60 |
| LBXPFUA | pfua | 1993 | 0.16 | 0.26 | 0.14 | 0.07 |

## Body Mass

NHANES Codebook References: <https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/BMX_I.htm>

### Load and Inspect BMI data

bodymass\_data <- nhanes\_load\_data("BMX\_I", "2015-2016", demographics = TRUE)  
as\_tibble(bodymass\_data)

## # A tibble: 9,544 x 76  
## SEQN cycle SDDSRVYR RIDSTATR RIAGENDR RIDAGEYR RIDAGEMN RIDRETH1  
## <int> <chr> <int> <int> <int> <int> <int> <int>  
## 1 83732 2015~ 9 2 1 62 NA 3  
## 2 83733 2015~ 9 2 1 53 NA 3  
## 3 83734 2015~ 9 2 1 78 NA 3  
## 4 83735 2015~ 9 2 2 56 NA 3  
## 5 83736 2015~ 9 2 2 42 NA 4  
## 6 83737 2015~ 9 2 2 72 NA 1  
## 7 83738 2015~ 9 2 2 11 NA 1  
## 8 83739 2015~ 9 2 1 4 NA 3  
## 9 83740 2015~ 9 2 1 1 13 2  
## 10 83741 2015~ 9 2 1 22 NA 4  
## # ... with 9,534 more rows, and 68 more variables: RIDRETH3 <int>,  
## # RIDEXMON <int>, RIDEXAGM <int>, DMQMILIZ <int>, DMQADFC <int>,  
## # DMDBORN4 <int>, DMDCITZN <int>, DMDYRSUS <int>, DMDEDUC3 <int>,  
## # DMDEDUC2 <int>, DMDMARTL <int>, RIDEXPRG <int>, SIALANG <int>,  
## # SIAPROXY <int>, SIAINTRP <int>, FIALANG <int>, FIAPROXY <int>,  
## # FIAINTRP <int>, MIALANG <int>, MIAPROXY <int>, MIAINTRP <int>,  
## # AIALANGA <int>, DMDHHSIZ <int>, DMDFMSIZ <int>, DMDHHSZA <int>,  
## # DMDHHSZB <int>, DMDHHSZE <int>, DMDHRGND <int>, DMDHRAGE <int>,  
## # DMDHRBR4 <int>, DMDHREDU <int>, DMDHRMAR <int>, DMDHSEDU <int>,  
## # WTINT2YR <dbl>, WTMEC2YR <dbl>, SDMVPSU <int>, SDMVSTRA <int>,  
## # INDHHIN2 <int>, INDFMIN2 <int>, INDFMPIR <dbl>, BMDSTATS <dbl>,  
## # BMXWT <dbl>, BMIWT <dbl>, BMXRECUM <dbl>, BMIRECUM <dbl>,  
## # BMXHEAD <dbl>, BMIHEAD <dbl>, BMXHT <dbl>, BMIHT <dbl>, BMXBMI <dbl>,  
## # BMDBMIC <dbl>, BMXLEG <dbl>, BMILEG <dbl>, BMXARML <dbl>,  
## # BMIARML <dbl>, BMXARMC <dbl>, BMIARMC <dbl>, BMXWAIST <dbl>,  
## # BMIWAIST <dbl>, BMXSAD1 <dbl>, BMXSAD2 <dbl>, BMXSAD3 <dbl>,  
## # BMXSAD4 <dbl>, BMDAVSAD <dbl>, BMDSADCM <dbl>, file\_name <chr>,  
## # begin\_year <dbl>, end\_year <dbl>

### Inspect body mass data from 2015-2016

# bodymass\_data %>% nhanes\_detection\_frequency("BMXBMI", "BMXBMI", "WTMEC2YR") # not completely sure on weight  
bodymass\_data %>% nhanes\_sample\_size("BMXBMI", "BMXBMI", "WTMEC2YR")

## value cycle begin\_year end\_year file\_name column weights\_column  
## 1 8756 2015-2016 2015 2016 BMX\_I BMXBMI WTMEC2YR  
## comment\_column name  
## 1 BMXBMI sample size

bodymass\_data %>% nhanes\_quantile("BMXBMI","BMXBMI", "WTMEC2YR", quantiles = c(0.5, 0.95))

## value cycle begin\_year end\_year file\_name column weights\_column  
## 1 26.6 2015-2016 2015 2016 BMX\_I BMXBMI WTMEC2YR  
## 2 41.3 2015-2016 2015 2016 BMX\_I BMXBMI WTMEC2YR  
## comment\_column below\_lod quantile name  
## 1 BMXBMI FALSE 50% quantile  
## 2 BMXBMI FALSE 95% quantile

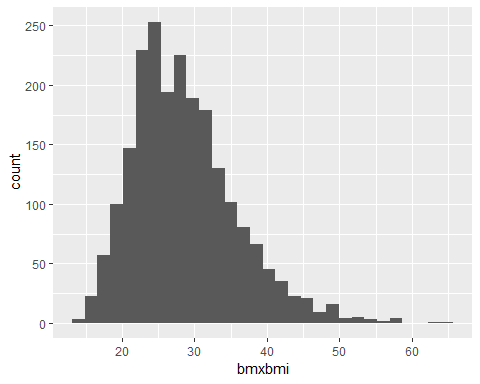
## Merged Dataset

pfas\_data\_clean = nhanes\_load\_data("PFAS\_I", "2015-2016", demographics = TRUE) %>%   
 janitor::clean\_names() %>%   
 select(seqn, cycle, sddsrvyr, riagendr, ridageyr, ridreth3, dmdeduc3, dmdeduc2, wtint2yr, wtmec2yr, lbxpfde:lbdmfosl) %>%   
 rename(pfdea = lbxpfde, pfhxs = lbxpfhs, me\_pfosa\_acoh = lbxmpah, pfna = lbxpfna, pfua = lbxpfua, pfdoa = lbxpfdo, n\_pfoa = lbxnfoa, sb\_pfoa = lbxbfoa, n\_pfos = lbxnfos, sm\_pfos = lbxmfos)  
  
bodymass\_data\_clean = bodymass\_data %>%   
 janitor::clean\_names() %>%   
 select(seqn, bmxbmi, bmxwt, bmiwt)  
  
pfas\_bodymass\_clean = left\_join(pfas\_data\_clean, bodymass\_data\_clean, by = "seqn")

# Histogram of BMI  
  
pfas\_bodymass\_clean %>%   
 ggplot(aes(x = bmxbmi)) +  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

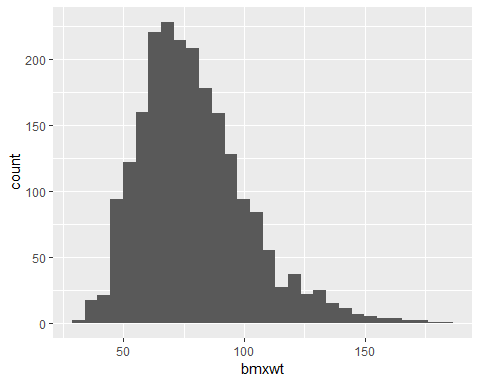
## Warning: Removed 23 rows containing non-finite values (stat\_bin).



# Histogram of Weight  
  
pfas\_bodymass\_clean %>%   
 ggplot(aes(x = bmxwt)) +  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 23 rows containing non-finite values (stat\_bin).



# Descriptive statistics BMI and weight  
  
stats\_bmi = as\_data\_frame(  
 summarise(pfas\_bodymass\_clean,  
 Mean = mean(bmxbmi, na.rm = TRUE),  
 SD = sd(bmxbmi, na.rm = TRUE),  
 Median = median(bmxbmi, na.rm = TRUE),  
 IQR = IQR(bmxbmi, na.rm = TRUE))  
 ) %>%   
 mutate(Variable = c("BMI"))  
  
stats\_wt = as\_data\_frame(  
 summarise(pfas\_bodymass\_clean,  
 Mean = mean(bmxwt, na.rm = TRUE),  
 SD = sd(bmxwt, na.rm = TRUE),  
 Median = median(bmxwt, na.rm = TRUE),  
 IQR = IQR(bmxwt, na.rm = TRUE))  
 ) %>%   
 mutate(Variable = c("Weight"))  
  
bodymass\_table = stats\_bmi %>%   
 bind\_rows(., stats\_wt) %>%   
 subset(select = c("Variable", "Mean", "SD", "Median", "IQR"))   
  
bodymass\_table %>%   
 knitr::kable()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean | SD | Median | IQR |
| BMI | 28.66707 | 7.148769 | 27.7 | 9.00 |
| Weight | 79.24974 | 22.403440 | 76.1 | 28.05 |

## Water

### Load water consumption

dietary\_day1 <- nhanes\_load\_data("DR1TOT\_I", "2015-2016") %>%   
 select(SEQN, DR1\_320Z, DR1\_330Z, DR1BWATZ, DR1TWS) %>%   
 janitor::clean\_names()

## Downloading DR1TOT\_I.XPT to C:\Users\jenni\AppData\Local\Temp\RtmpQ79ZJ1/DR1TOT\_I.XPT

dietary\_day2 <- nhanes\_load\_data("DR2TOT\_I", "2015-2016") %>%   
 select(SEQN, DR2\_320Z, DR2\_330Z, DR2BWATZ, DR2TWS) %>%   
 janitor::clean\_names()

## Downloading DR2TOT\_I.XPT to C:\Users\jenni\AppData\Local\Temp\RtmpQ79ZJ1/DR2TOT\_I.XPT

#### Link water consumption to SEQN

water\_matched <-   
 pfas\_data %>%   
 select(seqn) %>%   
 left\_join(dietary\_day1, by = "seqn") %>%   
 left\_join(dietary\_day2, by = "seqn") %>%   
 mutate(avg\_320z = (dr1\_320z + dr2\_320z) / 2,  
 avg\_330z = (dr1\_330z + dr2\_320z) / 2,  
 avgbwatz = (dr1bwatz + dr2bwatz) / 2,  
 avgtws = (dr1tws + dr2tws) / 2)

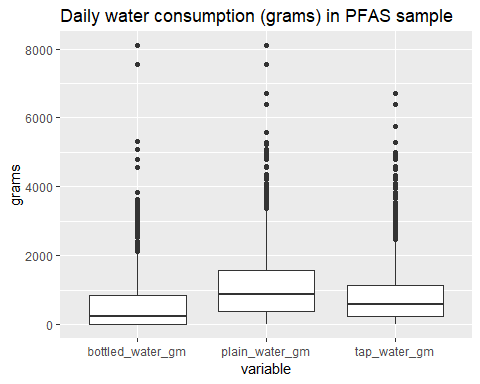
### Water consumption table (PFAS sample)

water\_table <-  
 water\_matched %>%   
 select(avg\_320z:avgtws) %>%   
 rename(plain\_water\_gm = avg\_320z,  
 tap\_water\_gm = avg\_330z ,  
 bottled\_water\_gm = avgbwatz,  
 tap\_water\_source = avgtws) %>%   
 summarise\_all(funs(mean, sd, median), na.rm = TRUE) %>%   
 gather(water\_consumption = plain\_water\_gm\_mean:tap\_water\_source\_median)   
  
water\_table %>%   
 knitr::kable()

|  |  |
| --- | --- |
| key | value |
| plain\_water\_gm\_mean | 1120.95273 |
| tap\_water\_gm\_mean | 838.24921 |
| bottled\_water\_gm\_mean | 546.67689 |
| tap\_water\_source\_mean | 8.15009 |
| plain\_water\_gm\_sd | 1027.23774 |
| tap\_water\_gm\_sd | 896.85726 |
| bottled\_water\_gm\_sd | 815.47016 |
| tap\_water\_source\_sd | 24.12917 |
| plain\_water\_gm\_median | 873.75000 |
| tap\_water\_gm\_median | 570.00000 |
| bottled\_water\_gm\_median | 240.00000 |
| tap\_water\_source\_median | 1.00000 |

### Water consumption boxplot

water\_matched %>%   
 select(avg\_320z:avgtws) %>%   
 rename(plain\_water\_gm = avg\_320z,  
 tap\_water\_gm = avg\_330z ,  
 bottled\_water\_gm = avgbwatz,  
 tap\_water\_source = avgtws) %>%   
 gather(key = "variable", value = "value", plain\_water\_gm:bottled\_water\_gm) %>%   
 group\_by(variable) %>%   
 na.omit() %>%   
 ggplot(aes(x = variable, y = value)) +  
 geom\_boxplot() +  
 labs(title = "Daily water consumption (grams) in PFAS sample", y = "grams")



water\_matched %>%  
 count(dr1tws)

## # A tibble: 6 x 2  
## dr1tws n  
## <dbl> <int>  
## 1 1 1279  
## 2 2 132  
## 3 3 16  
## 4 4 447  
## 5 99 147  
## 6 NA 149