bp upf and na

Table of Contents

# BP and UPF and Na in NDNS Dissertation

# Method

## Introduction

This section will take the research question and explain how the data will be used to answer the question.

There will be a description of the study and how the data collection happened. This will be followed by a consideration of the process governing ethics in this research.

The data will be loaded and the relevant variables identified and extracted. Some data may need to be recalculated or to be processed to make a more useable form. The population will be reviewed and any groups which might influence the results removed. Once the data set has been prepared, analysis begins with description of the data. The second analysis section involves using linear regression to identify if there is a correlation between the BP and each of the key variables. Multivariable regression models are then generated. These models are finally examined to identify the relative importance of the different variables in developing an optimal model.

## Research Question

What proportion of the association between blood pressure (SBP/DBP) and UPF intake can be explained by the changes in salt intake in England between 2008-2012 and 2015-2019?

The question can be split into parts, What was intake of salt between 2008 and 2019? What was intake of UPF between 2008 and 2019? What was BP between 2008 and 2019? Did each of these change over that time and how? Did the changes in any one affect any other? What are the sizes of the changes?

All of these questions look for numbers as answers.

Answering the question starts with counting.

## National Dietary and Nutritional Survey

This survey is a collaboration between government departments responsible for health and for food production. They have engaged academic partners to deliver reports on diet and nutrition across the United Kingdom. The study is designed to be representative across the whole area.

### Study design

This is a rolling cohort study which each year selects a new cohort of participants. The sample is approximately 1000 per year with 50% adults. The design has a random selection across potal units.

Having taken up the study participants complete a 4 day food diary, and have an interview with a nurse which includes taking several measurements.

### NDNS Dataset

The data from the NDNS study contains items about each individual,and their household. It contains a table with each item of food as recorded in their diary. There is a table with the overall intake of each of a large range of nutrients for the whole period. This is calculated from the diary using nutritional tables which are published as part of the dataset. The dataset is available via the UK national Data service for research purposes.

NDNS began before Monteiro’s processing based classification was developed. There is no record of Nova food type in NDNS. This has bee calculated from the food descriptions. I have used a table from Rauber et al., but also one from Colombet (personal communication)

### University Research Governance and Ethical Review

The research has beencarried out under the Universtiy governance. A proposal was discussed and agreed within the department. The need for ethical review was considered using the university research tool. The fact that the data is anonymised and there was no contact with participants means that there is minimal risk of hram to research participants.

Other ethical issues include data custodianship ensurign that the the rights of the owners of the data and of the participants are still considered as part of the process of analysis and dissemination of the research.

Issues around the power structures which lead to privelage one research project or proposal over another are considered more in the positionality section.

## Data Processing

The storage of the data is in keeping with the research governance agreements of the University and the Data set owners. The data is read from its files using ‘r-studio’ with the processing being carried out using packages available from CRAN. I have used files which had been amalgamated into four batches. These are 2008-2012, 2013-2014, 2015-2016, 2017-2019.

Once the data labels are made consistent across the batches, weighting recalculation is done. This generates values which account for differences in population balance across the annual cohorts. These result from differences in compliance and uptake within and across the years.

The years are amalgamated and the nature of the variables is specified.

The food diaries need processing to identifiy the UPF intake. Each persons food diary entries are assessed against the Nova food classification from Rauber. Then the weight and energy content of the days food is calcualted by Nova group. This is added to the intake for the other 3 days and the total intake by nova group established.

The percentage of the total intake by weight and by energy is then calculated for each of the 4 nova categories. Nova group 4 or UPF intake is used for the study.

## Method for Data Analysis

### Descriptive data analysis

The data is summarised, with Mean median, and range for the key continuous variables. The key variables are omsysval, UPF intake and sodiummg. These variables are the ones which most relate to the research question. First for years 1-4 then for 9-11.

There are a number of related variables in the dataset. These were chosen for reliability and practicality. These variables are ones which can also influence BP. They include Age, Sex, BMI, height and weight. The population for years 1-4 are compared with those for years 9-11.

The omsysval is a validated measurement with significant quality assessment within the dataset. Raw systolic values are present in the dataset but are made up of data with issues around quality. In particular the systolic values are assessed for the effects of exercise, temperature and ill health. The variable omsysval is a quality assured mean value which is reliable across the dataset.

The sodium value is one calculated from intake based on food diaries and standard food nutrient values. This only reflects standard foods and is the result of assumptions about the content being consistent. Serum sodium values are available for the early dataset, but not the later one. There are also values for 24 urinary sodium which is probably a better indicator of dietary sodium for parts of the dataset, but again these are not found in both time periods. The Data is displayed in table x with minimum, mean, median, and maximum values. The exposure variables are sodium intake (Sodiummg), and ultra processed food intake (UPF\_4). The outcome variable are the mean systolic blood pressure (omsysval).

Key additional variables are considered in the later section looking at how they change across the populations. Age, sex, and bmi are important contributors. Education, NSSEC and IMD are also looked at.

### Exclusions

The relationship between salt and systolic blood pressure may be different in individuals with pathologically high BP. Those taking BP controlling medications may have a different relationship to sodium and UPF. These patients were excluded from the main analysis. Analysis was done with them included and this produced results in line with those presented, but of greater magnitude. This additional analysis is not presented here.

### Comparative Analysis

The second phase of analysis uses standard techniques to compare the means of the populations. A student’s t test is used as the populations are normally distributed continuous variables. The t tests compare the sodium, UPF, and systolic BP in the cohort from 2008-2012 with that from 2017-2019. This will show how the intake has changed, and the same for the outcome. These are not the same participants so matched analysis, or time series analysis is not directly applicable.

Plots will be given to show the values in each of the available eleven cohorts.

### Linear Regression

Analysis of the correlation between BP and sodium intake, and then UPF intake is done using linear regression. This will give an indicator of the direction, and strength of any relationship between the variables. There is also anova analysis to understand the statistical significance of these results.

### Multiple Repgresion

Multivariable regression models are then developed to understand the interactions between variables and to develop a mathematical model. The optimal model is one whch best explains the pattern of data, but which also makes practical sense for the wider understanding of relationships. Assessment techniques try to understand the importance of including particualar variables, and the form in which they are best included.

### AIC and sensitivty Anaylsis

This section compares models side by side using assessment techniques to identify the best way of describing the data. The ‘best’ in part is determined by the wether a mode is needed to predict more data, or just to understand the data available. Here it is about how best to describe the relationship between Na, UPF, and BP.

##Method Conclusion

This section has highlighted how the material for the study is brought together and how the governance and ethics fit with the data collection, processing and analysis to help us to derive the results which will be presented in the next section.

### Data preparation

The data is then arranged into a format which allows processing. This includes identifying continuous and categorical variables. It also includes naming the categories of the categorical variables.

The data is then combined into two comprehensive tables.

The food diary data needs more processing. In particular the NOVA categorisation is not in the data set. I have derived UPFNOVA from a paper which had a data table identifying the NDNS sub food groups by Rauber et al.

#### Processing the food diaries

The Nova group is attached to the foods in the food diaries from Rauber et al @rauberUltraprocessedFoodsExcessive2019b. The tables are reduced to the necessary variables.

To work out the gram weight amount of food intake by each individual, first the diary entries for each individual are totalled up. The total gram weight value of intake of each food is then worked out as a percentage of the total intake.

All these individual calculations are then built back up into tables. This is done for years 9-11 and then 1-4.

The process can be done for food level energy intake also.

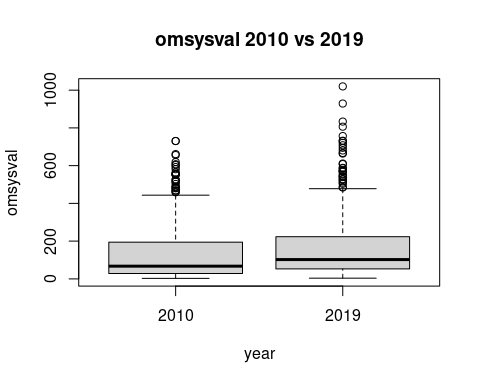
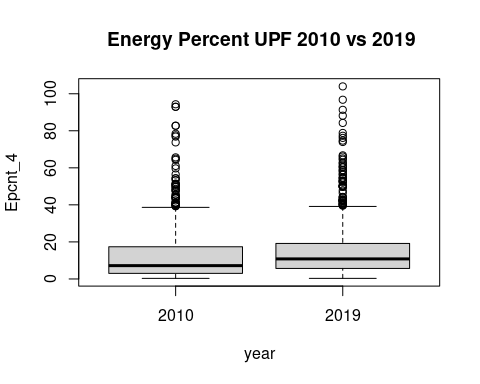
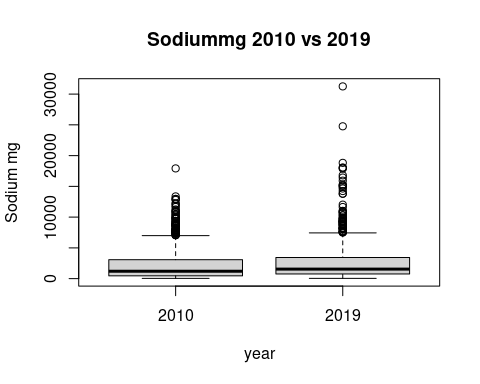
##Results Section ### Results Intrduction

The results derive from the method outlined above and follow the pattern described. I will try to discuss the results having already described the method.

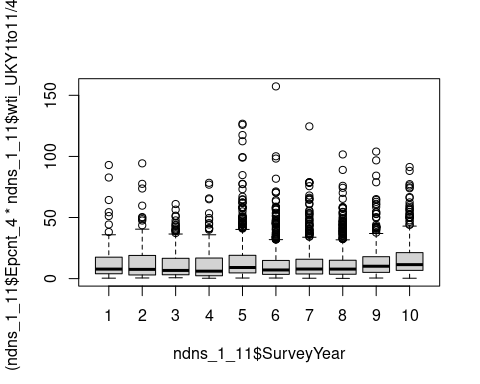
## Sodiummg pcnt\_4 Epcnt\_4 omsysval   
## Min. : 388.6 Min. : 0.86 Min. : 9.00 Min. : 77.5   
## 1st Qu.:1628.2 1st Qu.:14.14 1st Qu.:41.70 1st Qu.:108.0   
## Median :2041.1 Median :24.21 Median :52.40 Median :118.5   
## Mean :2122.2 Mean :28.81 Mean :51.88 Mean :118.8   
## 3rd Qu.:2540.2 3rd Qu.:40.36 3rd Qu.:61.95 3rd Qu.:129.0   
## Max. :6494.1 Max. :94.20 Max. :99.45 Max. :150.0

## Sodiummg pcnt\_4 Epcnt\_4 omsysval   
## Min. : 91.87 Min. : 0.21 Min. : 2.54 Min. : -8.00   
## 1st Qu.:1390.89 1st Qu.: 15.11 1st Qu.: 40.88 1st Qu.: 94.50   
## Median :1826.90 Median : 25.78 Median : 51.91 Median :109.00   
## Mean :1912.91 Mean : 30.07 Mean : 51.40 Mean : 90.95   
## 3rd Qu.:2324.24 3rd Qu.: 41.76 3rd Qu.: 61.91 3rd Qu.:121.00   
## Max. :7487.94 Max. :100.00 Max. :100.00 Max. :150.00

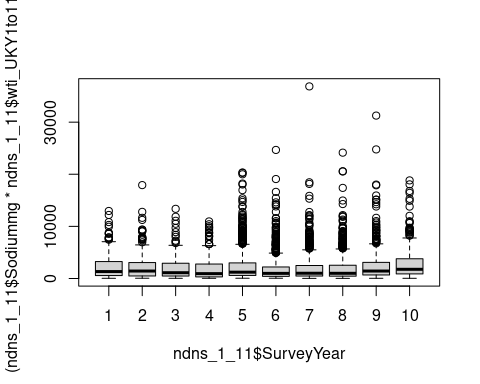
These tables show that there has been a change between the two cohorts. The boxplots show this graphically.



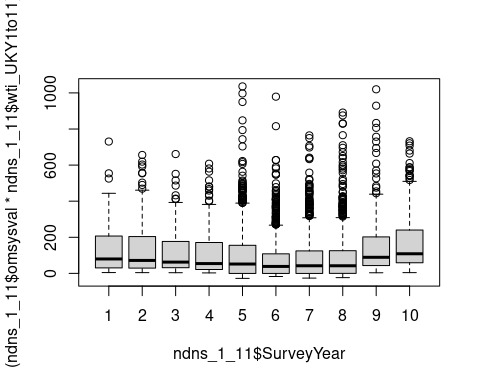
#keydata  
  
boxplot((ndns\_1\_11$Epcnt\_4\*ndns\_1\_11$wti\_UKY1to11/4)~  
 ndns\_1\_11$SurveyYear)



boxplot((ndns\_1\_11$Sodiummg\*ndns\_1\_11$wti\_UKY1to11)~  
 ndns\_1\_11$SurveyYear)



boxplot((ndns\_1\_11$omsysval\*ndns\_1\_11$wti\_UKY1to11)~  
 ndns\_1\_11$SurveyYear)

 These box plots show how the percentage of energy derived from UPF, the sodium intake, and the Systolic bp have changed over the years. The graphs show that there is not a visible difference between the years. Statistical analysis will follow.

## Statistical Comparison of key variables

## Comparison of key variables

### comparing UPF and Sodium intake calculated from diet

In order to confirm there has been a change in intake a t.test compares the means of the two samples. One compares the means of sodium in years 1-4 with sodium in years 9-11.

The second compares the means of pcnt UPF intake in over the same periods. A third compares the percentage energy provided by UPF.

## Var statistic p.value  
## 1: Epcnt\_4 4.353 1.412e-05  
## 2: pcnt\_4 3.709 2.137e-04  
## 3: Na 3.542 4.056e-04

##   
## Welch Two Sample t-test  
##   
## data: ndns\_1\_11[SurveyYear >= 9, pcnt\_4 \* wti\_UKY1to11] and ndns\_1\_11[SurveyYear <= 4, pcnt\_4 \* wti\_UKY1to11]  
## t = 3.7088, df = 2058.4, p-value = 0.0002137  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.708111 8.785751  
## sample estimates:  
## mean of x mean of y   
## 30.32624 24.57931

##   
## Welch Two Sample t-test  
##   
## data: ndns\_1\_11[SurveyYear >= 9, Epcnt\_4 \* wti\_UKY1to11] and ndns\_1\_11[SurveyYear <= 4, Epcnt\_4 \* wti\_UKY1to11]  
## t = 4.3527, df = 2021.6, p-value = 1.412e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 5.672926 14.976654  
## sample estimates:  
## mean of x mean of y   
## 59.17273 48.84794

##   
## Welch Two Sample t-test  
##   
## data: ndns\_1\_11[SurveyYear >= 9, Sodiummg \* wti\_UKY1to11] and ndns\_1\_11[SurveyYear <= 4, Sodiummg \* wti\_UKY1to11]  
## t = 3.5422, df = 2063.9, p-value = 0.0004056  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 186.5892 649.4641  
## sample estimates:  
## mean of x mean of y   
## 2574.327 2156.300

It seems the mean percentage UPF intake changes from 48.8% to 59.2% energy and this increase is statistically significant. The mean sodium intake has changed from 2156.30 mg to 2574.33 mg and is also statistically significant with a p value less than 0.05.

### what about outcome BP?

The next t tests compare mean systolic values in the two time periods and then the mean diastolic values.

## Var statistic p.value  
## 1: Sys 5.134 3.112e-07  
## 2: Dia 5.205 2.130e-07

##   
## Welch Two Sample t-test  
##   
## data: ndns\_1\_11[SurveyYear >= 9, omsysval \* wti\_UKY1to11] and ndns\_1\_11[SurveyYear <= 4, omsysval \* wti\_UKY1to11]  
## t = 5.1336, df = 2048.3, p-value = 3.112e-07  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 18.73190 41.89111  
## sample estimates:  
## mean of x mean of y   
## 151.9482 121.6367

There is a change in mean systolic from 122-152 mmHg with a p value of 3.112e -7.

In summary there is statistically significant change in UPF and Na intake and also in both systolic and diastolic pressures.

Has another factor affected the BP change ?

### Statistical analysis of Confounding variables

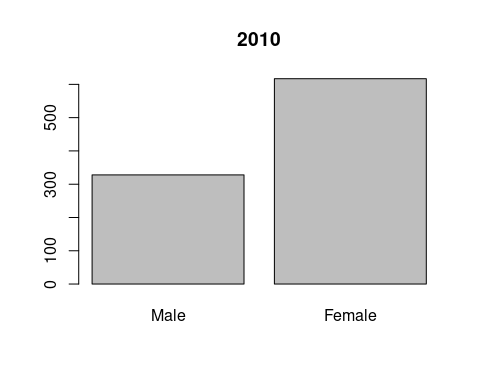
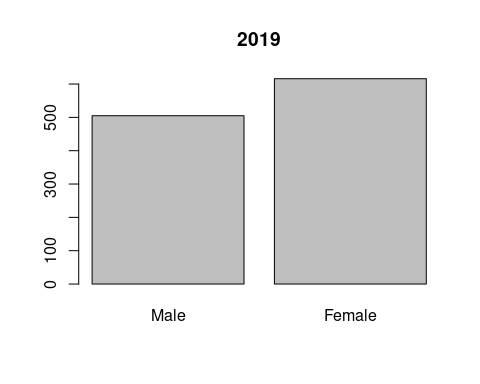
How are confounding variables distributed between the two datasets The NDNS dataset was weighted to keep many of these the same between datasets.

## name pvalue statistic  
## 1: Age 1.836e-01 1.33000  
## 2: Calciummg 1.128e-07 5.32400  
## 3: Totalsugarsg 3.621e-04 3.57200  
## 4: Glucoseg 2.839e-03 2.98800  
## 5: Fructoseg 5.978e-04 3.43800  
## 6: Sucroseg 3.907e-03 2.88900  
## 7: Lactoseg 1.024e-05 4.42300  
## 8: SOFTDRINKSLOWCALORIE 4.325e-08 5.50300  
## 9: SOFTDRINKSNOTLOWCALORIE 9.719e-01 0.03524  
## 10: TEACOFFEEANDWATER 3.895e-10 6.29000

The age of the two datasets has changed but not in a statistically significant way. Intake seems to be significantly different between the datasets.

There has been a change in the intake of soft drinks,not low calorie.

## name pvalue  
## 1: Sex 2.266e-06



There is a statistically significant change in the sex distribution of the two groups. This might be due to differences in the numbers of excluded participants. In particular there may be more younger people and women taking e.g. bblockers in one group.

## name pvalue statistic  
## 1: htval 2.094e-09 6.018  
## 2: wtval 9.372e-06 4.443  
## 3: bmival 1.636e-05 4.320

This table suggests that there is a significant difference between the height, and bmi of the groups.

## name p.value statistic  
## 1: vegetarn 0.0007869 14.29

These values identify a significant difference in the number of vegetarians

## name statistic p.value  
## 1: ethgrp5 31.41 2.527e-06  
## 2: ethgrp2 1199.00 3.131e-258

## name statistic p.value  
## 1: EIMD\_2007\_quintile 8.671 0.06986  
## 2: EIMD\_2010\_quintile 7.688 0.10370  
## 3: EIMD\_2015\_quintile 8.671 0.06986

## name statistic p.value  
## 1: educfin 181 1.216e-35

There are differences in ethnicity as divided into 5 subgroups. The differences in qimd, are not statistically significant. There is a difference in the age of finishing education.

## name p.value  
## 1: agegad1 0.003718  
## 2: agegad2 0.009063

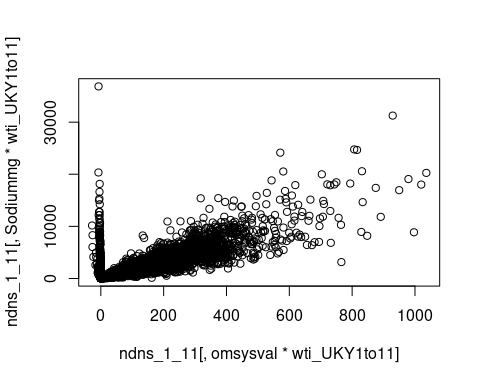
The age groups show some discrepancy with the p value significant only in the child age groups.

## Regression Analysis

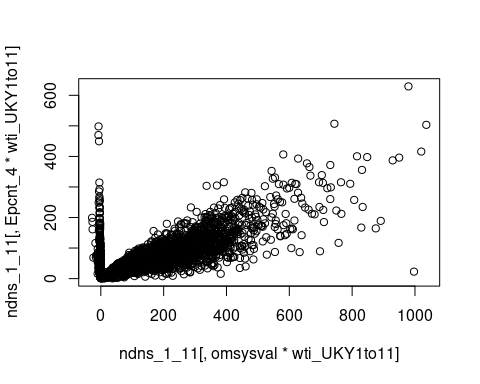
### linear regression

Simple linear regression equations look for the relationship between the dependant variable, and the independent variable. For these I am looking at the whole dataset Firstly I will plot omsysval and sodiummg, then omsysval and Epcnt, then omsysval and pcnt. These graphs show a patterned distribution.

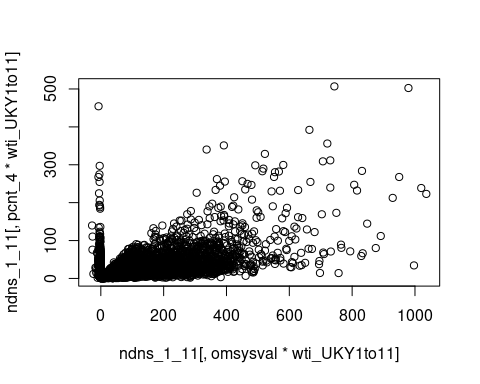
plot(ndns\_1\_11[,omsysval\*wti\_UKY1to11], ndns\_1\_11[,Sodiummg\*wti\_UKY1to11])



plot(ndns\_1\_11[,omsysval\*wti\_UKY1to11], ndns\_1\_11[,Epcnt\_4\*wti\_UKY1to11])



plot(ndns\_1\_11[,omsysval\*wti\_UKY1to11], ndns\_1\_11[,pcnt\_4\*wti\_UKY1to11])



The regression models are examined for specific variables. First omsysval against pcnt\_4.

##   
## Call:  
## lm(formula = (omsysval) ~ (pcnt\_4), data = ndns\_1\_11, weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) pcnt\_4   
## 110.1335 -0.3177

## Analysis of Variance Table  
##   
## Response: (omsysval)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## pcnt\_4 1 185836 185836 94.703 < 2.2e-16 \*\*\*  
## Residuals 5454 10702347 1962   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 59206.46

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: (omsysval) ~ (pcnt\_4)  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' pcnt\_4 ':  
## Coef. estimate: -0.31772   
## Standard Error: 0.03265   
## t-value: -9.73156   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.01707   
## Robustness Value, q = 1 : 0.12338   
## Robustness Value, q = 1 alpha = 0.05 : 0.09983   
##   
## For more information, check summary.

Then omsysval is compared to Epcnt\_4 showing a positive result which is statistically significant.

##   
## Call:  
## lm(formula = omsysval ~ Epcnt\_4, data = ndns\_1\_11, weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Epcnt\_4   
## 112.6456 -0.2203

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Epcnt\_4 1 67524 67524 34.034 5.726e-09 \*\*\*  
## Residuals 5454 10820659 1984   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 59266.44

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: omsysval ~ Epcnt\_4  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' Epcnt\_4 ':  
## Coef. estimate: -0.22033   
## Standard Error: 0.03777   
## t-value: -5.8339   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.0062   
## Robustness Value, q = 1 : 0.07594   
## Robustness Value, q = 1 alpha = 0.05 : 0.05109   
##   
## For more information, check summary.

sodiummg

##   
## Call:  
## lm(formula = omsysval ~ Sodiummg, data = ndns\_1\_11, weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Sodiummg   
## 93.500167 0.004256

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Sodiummg 1 62790 62790 31.634 1.954e-08 \*\*\*  
## Residuals 5454 10825393 1985   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 59268.83

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: omsysval ~ Sodiummg  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' Sodiummg ':  
## Coef. estimate: 0.00426   
## Standard Error: 0.00076   
## t-value: 5.62446   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.00577   
## Robustness Value, q = 1 : 0.07331   
## Robustness Value, q = 1 alpha = 0.05 : 0.0484   
##   
## For more information, check summary.

There are relationships between Na and g pcnt as well as E pcnt and omsysval .

#### Conclusion

The linear regression models show that there are statistically significant positive correlations between the systolic BP and each of the key variables.

### multi variable regression

This uses a model of variables. It can highlight the contributions of each variable. The

intention is to develop an optimal model which mathematically describes the situation.

In particular the research question asks about the relationship between Sodium and UPF intake with BP. The models will reflect this question with models looking to include or exclude particular variables. Comparisons between these models are then made using sensitivity analysis, identifying how sensitive the model is to sodium, or other factors

This first model looks at the relationships between BP and Age and Sex

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + bmival, data = ndns\_1\_11,   
## weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Age SexFemale bmival   
## 62.7371 0.4289 -2.2184 0.9943

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 750614 750614 413.5180 < 2e-16 \*\*\*  
## Sex 1 9191 9191 5.0636 0.02447 \*   
## bmival 1 201414 201414 110.9601 < 2e-16 \*\*\*  
## Residuals 5296 9613247 1815   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 57094.45

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: omsysval ~ Age + Sex + bmival  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' Age ':  
## Coef. estimate: 0.42891   
## Standard Error: 0.0305   
## t-value: 14.06491   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.03601   
## Robustness Value, q = 1 : 0.17549   
## Robustness Value, q = 1 alpha = 0.05 : 0.15307   
##   
## For more information, check summary.

The next model looks at a large number of variables

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Sodiummg + sqrt(pcnt\_4) +   
## wtval + TotalEMJ + ethgrp2 + VitaminDµg + educfinh + EIMD\_2010\_quintile,   
## data = ndns\_1\_11, weights = wti\_UKY1to11, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Age SexFemale   
## 69.847616 0.385109 2.503620   
## Sodiummg sqrt(pcnt\_4) wtval   
## -0.001216 0.026429 0.368138   
## TotalEMJ ethgrp2Non-white VitaminDµg   
## 0.698362 0.649504 0.683179   
## educfinh2 educfinh3 educfinh4   
## 14.027519 -9.059706 -9.917399   
## educfinh5 educfinh6 educfinh7   
## -10.740833 -15.242392 -9.539394   
## educfinh8 EIMD\_2010\_quintile2 EIMD\_2010\_quintile3   
## -9.421504 -1.832457 -0.794902   
## EIMD\_2010\_quintile4 EIMD\_2010\_quintile5   
## -2.824420 -2.914142

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 579247 579247 277.3863 < 2.2e-16 \*\*\*  
## Sex 1 7570 7570 3.6250 0.0570255 .   
## Sodiummg 1 25074 25074 12.0071 0.0005383 \*\*\*  
## sqrt(pcnt\_4) 1 88 88 0.0423 0.8371440   
## wtval 1 204826 204826 98.0857 < 2.2e-16 \*\*\*  
## TotalEMJ 1 6793 6793 3.2532 0.0713975 .   
## ethgrp2 1 323 323 0.1547 0.6941361   
## VitaminDµg 1 6941 6941 3.3241 0.0683833 .   
## educfinh 7 24205 3458 1.6559 0.1153540   
## EIMD\_2010\_quintile 4 4605 1151 0.5513 0.6981006   
## Residuals 2698 5634052 2088   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 28341.61

These models can be compared with others with different variables to understand how they help predict values more or less effectively.

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Sodiummg + Epcnt\_4 + bmival +   
## ethgrp2 + VitaminDµg + educfinh + EIMD\_2010\_quintile, data = ndns\_1\_11,   
## weights = wti\_UKY1to11, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Age SexFemale   
## 67.893153 0.478153 -1.573087   
## Sodiummg Epcnt\_4 bmival   
## 0.001093 0.075204 0.991683   
## ethgrp2Non-white VitaminDµg educfinh2   
## 0.011396 0.751497 12.167776   
## educfinh3 educfinh4 educfinh5   
## -13.281820 -12.559376 -12.865655   
## educfinh6 educfinh7 educfinh8   
## -17.634303 -11.120282 -10.484127   
## EIMD\_2010\_quintile2 EIMD\_2010\_quintile3 EIMD\_2010\_quintile4   
## -1.575845 -0.683227 -2.792052   
## EIMD\_2010\_quintile5   
## -2.673494

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 554444 554444 262.3753 < 2.2e-16 \*\*\*  
## Sex 1 9207 9207 4.3569 0.036955 \*   
## Sodiummg 1 20129 20129 9.5254 0.002047 \*\*   
## Epcnt\_4 1 1755 1755 0.8303 0.362259   
## bmival 1 127911 127911 60.5305 1.026e-14 \*\*\*  
## ethgrp2 1 66 66 0.0313 0.859476   
## VitaminDµg 1 8760 8760 4.1456 0.041841 \*   
## educfinh 7 32003 4572 2.1635 0.034542 \*   
## EIMD\_2010\_quintile 4 4209 1052 0.4980 0.737264   
## Residuals 2677 5656965 2113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 28143.16

this model has sodium and gram percent

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Sodiummg + sqrt(pcnt\_4) +   
## bmival, data = ndns\_1\_11, weights = wti\_UKY1to11, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Age SexFemale Sodiummg sqrt(pcnt\_4)   
## 62.911122 0.402169 -1.444537 0.001905 -0.627627   
## bmival   
## 0.974467

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 750614 750614 413.9073 < 2.2e-16 \*\*\*  
## Sex 1 9191 9191 5.0684 0.0244068 \*   
## Sodiummg 1 23617 23617 13.0228 0.0003106 \*\*\*  
## sqrt(pcnt\_4) 1 1784 1784 0.9836 0.3213491   
## bmival 1 188681 188681 104.0436 < 2.2e-16 \*\*\*  
## Residuals 5294 9600579 1813   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 57091.46

This model has Sodium and energy pcnt

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Sodiummg + Epcnt\_4, data = ndns\_1\_11,   
## weights = wti\_UKY1to11, na.action = na.exclude)  
##   
## Coefficients:  
## (Intercept) Age SexFemale Sodiummg Epcnt\_4   
## 73.950180 0.579049 -0.834323 0.003251 0.009189

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 842158 842158 458.8942 < 2.2e-16 \*\*\*  
## Sex 1 8856 8856 4.8258 0.02808 \*   
## Sodiummg 1 33451 33451 18.2276 1.993e-05 \*\*\*  
## Epcnt\_4 1 100 100 0.0547 0.81504   
## Residuals 5451 10003617 1835   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 58844.09

this model has Age sex and g pcnt only

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + sqrt(pcnt\_4), data = ndns\_1\_11,   
## weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Age SexFemale sqrt(pcnt\_4)   
## 81.88159 0.58329 -2.46709 -0.04109

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 842158 842158 457.4451 < 2e-16 \*\*\*  
## Sex 1 8856 8856 4.8105 0.02833 \*   
## sqrt(pcnt\_4) 1 22 22 0.0117 0.91374   
## Residuals 5452 10037147 1841   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 58860.35

the pcnt\_4 is not statistically significant

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Epcnt\_4, data = ndns\_1\_11,   
## weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Age SexFemale Epcnt\_4   
## 79.35795 0.59455 -2.38325 0.03909

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 842158 842158 457.5299 < 2e-16 \*\*\*  
## Sex 1 8856 8856 4.8114 0.02831 \*   
## Epcnt\_4 1 1881 1881 1.0220 0.31208   
## Residuals 5452 10035287 1841   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 58859.34

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: omsysval ~ Age + Sex + Epcnt\_4  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' Epcnt\_4 ':  
## Coef. estimate: 0.03909   
## Standard Error: 0.03867   
## t-value: 1.01095   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.00019   
## Robustness Value, q = 1 : 0.0136   
## Robustness Value, q = 1 alpha = 0.05 : 0   
##   
## For more information, check summary.

In this model Epcnt is not statistically significant.

What has removing the sodium done to anova and AIC?

This last model is just sodium with Age and sex

##   
## Call:  
## lm(formula = omsysval ~ Age + Sex + Sodiummg, data = ndns\_1\_11,   
## weights = wti\_UKY1to11)  
##   
## Coefficients:  
## (Intercept) Age SexFemale Sodiummg   
## 74.408750 0.576690 -0.834906 0.003285

## Analysis of Variance Table  
##   
## Response: omsysval  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 842158 842158 458.9737 < 2.2e-16 \*\*\*  
## Sex 1 8856 8856 4.8266 0.02807 \*   
## Sodiummg 1 33451 33451 18.2307 1.99e-05 \*\*\*  
## Residuals 5452 10003718 1835   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] 58842.14

## Sensitivity Analysis to Unobserved Confounding  
##   
## Model Formula: omsysval ~ Age + Sex + Sodiummg  
##   
## Null hypothesis: q = 1 and reduce = TRUE   
##   
## Unadjusted Estimates of ' Sodiummg ':  
## Coef. estimate: 0.00328   
## Standard Error: 0.00077   
## t-value: 4.26975   
##   
## Sensitivity Statistics:  
## Partial R2 of treatment with outcome: 0.00333   
## Robustness Value, q = 1 : 0.05618   
## Robustness Value, q = 1 alpha = 0.05 : 0.03079   
##   
## For more information, check summary.

This model shows age and sodium mg as being statistically significant. Sex appears less significant.

## This final set analyses the whole dataset together across the key variables

Then tests them across two sets of UPF data one calculated using Rauber, the other from ZC. First for gram percent UPF 4

the second set compares Energy percent upf between the two datasets

the next section compares the AIC values for the given models from above

library(AICcmodavg)  
Cand.models = list("no sodium no epcnt"= lm1AS, "sodium"= lm4c,"sodium epcnt" = lmallE,"Epcnt" = lm4b)  
selectionTable <- aictab(cand.set = Cand.models)  
selectionTable

##   
## Model selection based on AICc:  
##   
## K AICc Delta\_AICc AICcWt Cum.Wt LL  
## no sodium no epcnt 5 57094.46 0.00 1 1 -28542.22  
## sodium 5 58842.16 1747.70 0 1 -29416.07  
## sodium epcnt 6 58844.10 1749.65 0 1 -29416.04  
## Epcnt 5 58859.35 1764.89 0 1 -29424.67

confint(lm4c)

## 2.5 % 97.5 %  
## (Intercept) 70.257062777 78.560436864  
## Age 0.523236243 0.630144624  
## SexFemale -3.149589552 1.479778434  
## Sodiummg 0.001776581 0.004792852

confset(cand.set = Cand.models)

##   
## Confidence set for the best model  
##   
## Method: raw sum of model probabilities  
##   
## 95% confidence set:  
## K AICc Delta\_AICc AICcWt  
## no sodium no epcnt 5 57094.46 0 1  
##   
## Model probabilities sum to 1

evidence(aic.table = selectionTable)

##   
## Evidence ratio between models 'no sodium no epcnt' and 'sodium':  
## Inf

evidence(selectionTable, model.high = "sodium",  
 model.low = "no sodium no epcnt")

##   
## Evidence ratio between models 'sodium' and 'no sodium no epcnt':  
## 0

modavg(Cand.models,parm = "Sodiummg")

##   
## Multimodel inference on "Sodiummg" based on AICc  
##   
## AICc table used to obtain model-averaged estimate:  
##   
## K AICc Delta\_AICc AICcWt Estimate SE  
## sodium 5 58842.16 0.00 0.73 0 0  
## sodium epcnt 6 58844.10 1.95 0.27 0 0  
##   
## Model-averaged estimate: 0   
## Unconditional SE: 0   
## 95% Unconditional confidence interval: 0, 0

importance(cand.set = Cand.models, parm = "Sodiummg", second.ord = TRUE,nobs = NULL)

##   
## Importance values of 'Sodiummg':  
##   
## w+ (models including parameter): 0   
## w- (models excluding parameter): 1

## Summary

NDNS datasets have been downloaded and adapted into a form to approach the research question.

The key variables of BP, ‘omsysval’ and ‘omdiaval’ are taken directly from the data. The diary entries are identified by NOVA type.

The total weight of each nova type is calculated for each individual. The percentage of the total weight food intake per person is then calculated. This gives the derived value ‘pcnt\_4’, which is the percentage by weight of intake which is NOVA 4 or UPF.

The total energy in kJ of each nova type is calculated for each individual. The percentage of the total energy food intake per person is then calculated. This gives the derived value ‘Epcnt\_4’, which is the percentage of intake which is NOVA 4 or UPF.

There is a table with summary values for these variables across the dataset.

Statistical analysis of the key variables shows the change in all the variables between the two time periods.

Confounding variables are analysed and show if there has been a significant change in the balance of the populations.

Regression shows a degree of association between the BP and UPF intake by weight and by energy. It also shows the same for sodium intake.

Using Anova analysis of different multi variable regression models the key variables are significant for sodium in several models, and sometimes for energy percentage. Sodium intake shows the strongest association in the latter 9-11 cohort.

Of the other variables bmi and educfinh can also be statistically significant at times.

## Conclusion

The percentage by weight of NOVA group 4 foods increased from 2008 to 2019. The percentage by energy of NOVA group 4 foods increased from 2008 to 2019. The mean sodium intake in mg increased between the two time periods. The systolic and diastolic BP have increased between the two time periods.

In each period there is a correlation between systolic BP and sodium intake. In each period there is a correlation between systolic BP and UPF intake.

The regression models identify that age and sex are statistically significant contributors to the BP. Only those models from the later time period show sodium as being statistically significant in importance. Combining the data shows the energy percentage of UPF as being a statistically significant contributor.