

# 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 postal units (psu).

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 been 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 been carried out under the University 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 harm to research participants.

Other ethical issues include data custodianship ensuring that 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 privilege 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 identify the UPF intake. Each person's food diary entries are assessed against the Nova food classification from Rauber. Then the weight and energy content of the day's food is calculated 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 is 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.

Other variables in the data are compared across to assess how the data changes. t tests are again used for continuous data, and chi squared tests are used for categorical data.

### 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 Regression

Multivariable regression models are then developed to understand the interactions between variables and to develop a mathematical model. The optimal model is one which 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 particular variables, and the form in which they are best included.

### AIC and sensitivity Analysis

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 whether a model 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.

## Results Section

### Results Introduction

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.

### Descriptive Data Analysis

This first table highlights the data from the years 2008-2012 at the beginning of the data collection.

```
##           mean      SE
## Sodiummg 2181.9 16.835

##           mean      SE
## omsysval 124.09 0.4614

##           mean      SE
## Epcnt_4 49.147 0.3052

## [1] "tbl_svy"          "survey.design2" "survey.design"
```

Characteristic	N = 6,828 <sup>1</sup>
Sodium (mg) diet only	2,067 (783)
Epcnt_4	53 (15)
pcnt_4	32 (19)
omsysval	119 (18)
Unknown	3,321

<sup>1</sup>Mean (SD)

Characteristic	N = 5,693 <sup>1</sup>
Sodiummg	2,182 (837)
Epcnt_4	49 (15)
pcnt_4	26 (17)

<b>Characteristic</b>	<b>N = 6,828<sup>1</sup></b>
omsysval	124 (17)
Unknown	2,506

<sup>1</sup>Mean (SD)

The second table show the data from the second cohort 2017-19.

```
##          mean      SE
## Sodiummg 1914.3 18.409
```

```
##          mean      SE
## omsysval 121.61 0.5024
```

```
##          mean      SE
## Epcnt_4 45.92 0.3833
```

```
## [1] "tbl_svy"          "survey.design2" "survey.design"
```

<b>Characteristic</b>	<b>N = 3,558<sup>1</sup></b>
Sodium (mg) diet only	1,787 (701)
Epcnt_4	49 (15)
omsysval	118 (17)
Unknown	2,227

<sup>1</sup>Mean (SD)

<b>Characteristic</b>	<b>N = 4,270<sup>1</sup></b>
Sodiummg	1,914 (754)
Epcnt_4	46 (16)
omsysval	122 (17)
Unknown	2,436

<sup>1</sup>Mean (SD)

These tables show that there has been a change between the two cohorts. They also show that the weighted samples are different to raw samples identifying the importance of analysing the data across weighted groups.

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.

## Comparative analysis

### 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      13.230 1.78e-39
## 2: pcnt_4       9.062 1.68e-19
## 3:      Na      8.863 9.70e-19

##
## 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 = 9.0624, df = 6233.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.849281 7.526345
## sample estimates:
## mean of x mean of y
## 28.04142 21.85361

##
## 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 = 13.231, df = 6796.4, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 12.03483 16.22122
## sample estimates:
## mean of x mean of y
## 55.10377 40.97575
```

```
##
## Welch Two Sample t-test
##
## data: ndns_1_11[SurveyYear >= 9, Sodiummg * wti_UKY1to11] and
ndns_1_11[SurveyYear <= 4, Sodiummg * wti_UKY1to11]
## t = 8.8635, df = 7012.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 372.3000 583.7443
## sample estimates:
## mean of x mean of y
## 2297.148 1819.126
```

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      12.06 1.830e-32
## 2: Dia      12.42 2.892e-34

##
## Welch Two Sample t-test
##
## data: ndns_1_11[SurveyYear >= 9, omsysval * wti_UKY1to11] and
ndns_1_11[SurveyYear <= 4, omsysval * wti_UKY1to11]
## t = 12.057, df = 2207.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 45.87742 63.70054
## sample estimates:
## mean of x mean of y
## 167.5446 112.7556
```

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 other variables*

How are variables distributed between the two cohorts. The NDNS dataset was weighted to keep many of these the same between datasets. Continuous variables are assessed using ttests and categorical variables using chi squared tests to give p.values.



## Age and Sex

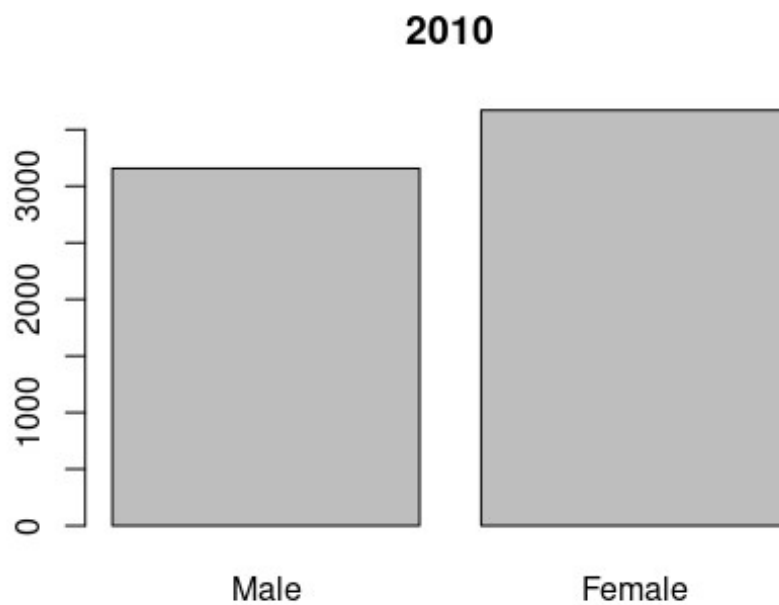
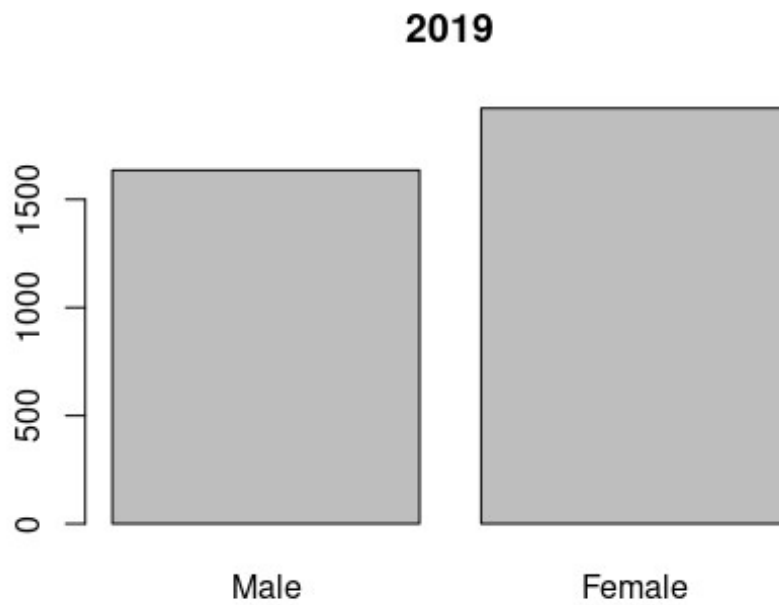
There is

##		name	pvalue	statistic
## 1:		Age	4.323e-38	12.9900
## 2:		Calciummg	6.929e-40	13.3100
## 3:		Totalsugarsg	1.966e-18	8.7840
## 4:		Glucoseg	1.444e-16	8.2840
## 5:		Fructoseg	2.055e-20	9.2910
## 6:		Sucroseg	3.788e-10	6.2710
## 7:		Lactoseg	6.278e-32	11.8300
## 8:	SOFTDRINKSLOWCALORIE		1.063e-23	10.0900
## 9:	SOFTDRINKSNOTLOWCALORIE		7.123e-01	-0.3688
## 10:	TEACOFFEEANDWATER		3.623e-53	15.5200

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	0.8205



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 9.325e-58      16.20
## 2: wtval 7.901e-44      14.00
## 3: bmival 2.238e-50      15.07
```

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

```
##      name      p.value statistic
## 1: vegetarn 7.709e-06      23.55
```

These values identify a significant difference in the number of vegetarians

```
##      name statistic      p.value
## 1: ethgrp5      68.41 4.919e-14
## 2: ethgrp2 6494.00 0.000e+00
```

```
##      name statistic      p.value
## 1: EIMD_2007_quintile      6.968 0.1376
## 2: EIMD_2010_quintile      5.139 0.2733
## 3: EIMD_2015_quintile      6.968 0.1376
```

```
##      name statistic      p.value
## 1: educfin      695.9 5.287e-146
```

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 0.0002328
## 2: agegad2 0.0021990
```

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

The regression models are examined for Sodium and UP against BP. These use the populations where participants have been excluded. (analysis including these makes no difference!!)

omsysval is compared to EnergykJ

```
## Stratified 1 - level Cluster Sampling design (with replacement)
## With (1695) clusters.
## svydesign(ids = ~area, weights = ~wti_UKY1toll, strata = ~astrata1,
```

```

##      data = ndns_1_11, nest = TRUE)
##
## Call:  svyglm(formula = omsysval ~ EnergykJ_4, design = ndns_1_11e)
##
## Coefficients:
## (Intercept)      EnergykJ_4
##   9.417e+01    1.691e-04
##
## Degrees of Freedom: 10106 Total (i.e. Null);  856 Residual
## (5548 observations deleted due to missingness)
## Null Deviance:      29900000
## Residual Deviance: 29890000  AIC: 114300

then sodiummg

## Stratified 1 - level Cluster Sampling design (with replacement)
## With (1695) clusters.
## svydesign(ids = ~area, weights = ~wti_UKY1tol1, strata = ~astrata1,

##      data = ndns_1_11, nest = TRUE)
##
## Call:  svyglm(formula = omsysval ~ Sodiummg, design = ndns_1_11e)
##
## Coefficients:
## (Intercept)      Sodiummg
##   85.078412    0.005583
##
## Degrees of Freedom: 10106 Total (i.e. Null);  856 Residual
## (5548 observations deleted due to missingness)
## Null Deviance:      29900000
## Residual Deviance: 29690000  AIC: 114300

## NULL

##      eff.p      AIC  deltabar
##  2.636406 44.744433  1.318203

```

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

In 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 education and IMD all of which may have an effect on BP. This model excludes UPF and Na.

```
## Stratified 1 - level Cluster Sampling design (with replacement)
## With (879) clusters.
## svydesign(ids = ~area, weights = ~wti_UKY1to11, strata = ~astrata1,

##      data = ndns_1_11, nest = TRUE)
##
## Call:  svyglm(formula = omsysval ~ Age + Sex + bmival + educfinh +
##      EIMD_2010_quintile,
##      design = ndns_1_11e, na.action = na.exclude)
##
## Coefficients:
##      (Intercept)              Age              SexFemale
##      63.8630              0.5685              -6.1558
##      bmival              educfinh2              educfinh3
##      1.4457              1.9234              -23.0252
##      educfinh4              educfinh5              educfinh6
##      -22.6198              -21.2258              -29.1864
##      educfinh7              educfinh8  EIMD_2010_quintile2
##      -18.3646              -18.7862              1.8094
##  EIMD_2010_quintile3  EIMD_2010_quintile4  EIMD_2010_quintile5
##      -0.9827              -4.9714              -3.9077
##
## Degrees of Freedom: 4428 Total (i.e. Null);  429 Residual
## (11226 observations deleted due to missingness)
## Null Deviance:      1.8e+07
## Residual Deviance: 15610000  AIC: 48010

## Anova table: (Rao-Scott LRT)
## svyglm(formula = omsysval ~ Age, design = ndns_1_11e, na.action =
## na.exclude)
##
```

	stats	DEff	df	ddf	p
## Age	1769230	5152.1	1.0	856	< 2.2e-16 ***
## Sex	66818	5415.5	1.0	855	0.0004998 ***
## bmival	1967932	7601.7	1.0	848	< 2.2e-16 ***
## educfinh	6865959	4810.8	7.0	741	< 2.2e-16 ***
## EIMD_2010_quintile	3619401	6794.8	4.0	429	< 2.2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 48008.04
```

This first model shows that all these variables, Age, Sex, education, IMD, and bmi, give statistically significant coefficients for the model which suggests that they do have an important part to play in any optimal model.

The next model adds Sodiummg.

```

## Stratified 1 - level Cluster Sampling design (with replacement)
## With (879) clusters.
## svydesign(ids = ~area, weights = ~wti_UKY1tol1, strata = ~astrata1,

##      data = ndns_1_11, nest = TRUE)
##
## Call:  svyglm(formula = omsysval ~ Age + Sex + Sodiummg + bmival +
educfinh +
##      EIMD_2010_quintile, design = ndns_1_11e, na.action =
na.exclude)
##
## Coefficients:
##      (Intercept)              Age              SexFemale
##      57.676434              0.570174             -4.612795
##      Sodiummg              bmival              educfinh2
##      0.002946              1.406515              3.624975
##      educfinh3              educfinh4              educfinh5
##      -21.989210             -22.047254             -20.906646
##      educfinh6              educfinh7              educfinh8
##      -28.635031             -18.152873             -18.510535
## EIMD_2010_quintile2 EIMD_2010_quintile3 EIMD_2010_quintile4
##      1.693260              -1.376258             -5.140956
## EIMD_2010_quintile5
##      -3.964199
##
## Degrees of Freedom: 4428 Total (i.e. Null);  428 Residual
## (11226 observations deleted due to missingness)
## Null Deviance:      1.8e+07
## Residual Deviance: 15580000  AIC: 48000

## Anova table: (Rao-Scott LRT)
## svyglm(formula = omsysval ~ Age, design = ndns_1_11e, na.action =
na.exclude)
##
##      stats      DEff      df ddf      p
## Age      1769230    5152.1    1.0 856 < 2.2e-16 ***
## Sex       66818     5415.5    1.0 855 0.0004998 ***
## Sodiummg  168043    7128.3    1.0 854 1.56e-06 ***
## bmival    1860042    7607.8    1.0 847 < 2.2e-16 ***
## educfinh  6862697    4779.8    7.0 740 < 2.2e-16 ***
## EIMD_2010_quintile 3597547    6707.6    4.0 428 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

This second model gives Sodiummg, educfinh, and IMD statistical significance. VitaminD shows no statistical significance, TotalEMJ and sqrt(pcmt) and ethgrp2 all have limited significance.

Now we add UPF as total energy from nova 4 or UPF

```

## Stratified 1 - level Cluster Sampling design (with replacement)
## With (879) clusters.

```

```
## svydesign(ids = ~area, weights = ~wti_UKY1tol1, strata = ~astrata1,
##         data = ndns_1_11, nest = TRUE)
##
## Call:  svyglm(formula = omsysval ~ Age + Sex + Sodiummg +
##             EnergykJ_4 +
##             bmival + educfinh + EIMD_2010_quintile, design = ndns_1_11e,
##             na.action = na.exclude)
##
## Coefficients:
##             (Intercept)                Age                SexFemale
##             5.624e+01                5.804e-01            -4.413e+00
##             Sodiummg                EnergykJ_4                bmival
##             2.104e-03                1.662e-04                1.410e+00
##             educfinh2                educfinh3                educfinh4
##             3.957e+00            -2.196e+01            -2.198e+01
##             educfinh5                educfinh6                educfinh7
##             -2.076e+01            -2.852e+01            -1.784e+01
##             educfinh8  EIMD_2010_quintile2  EIMD_2010_quintile3
##             -1.817e+01                1.714e+00            -1.330e+00
##             EIMD_2010_quintile4  EIMD_2010_quintile5
##             -5.130e+00            -3.962e+00
##
## Degrees of Freedom: 4428 Total (i.e. Null);  427 Residual
## (11226 observations deleted due to missingness)
## Null Deviance:      1.8e+07
## Residual Deviance: 15570000  AIC: 48000

## Anova table: (Rao-Scott LRT)
## svyglm(formula = omsysval ~ Age, design = ndns_1_11e, na.action =
## na.exclude)
##
```

	stats	DEff	df	ddf	p
## Age	1769230	5152.1	1.0	856	< 2.2e-16 ***
## Sex	66818	5415.5	1.0	855	0.0004998 ***
## Sodiummg	168043	7128.3	1.0	854	1.56e-06 ***
## EnergykJ_4	10540	6501.8	1.0	853	0.2058373
## bmival	1866294	7578.6	1.0	846	< 2.2e-16 ***
## educfinh	6854647	4774.7	7.0	739	< 2.2e-16 ***
## EIMD_2010_quintile	3593502	6701.1	4.0	427	< 2.2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

UPF does not seem significant...

but when removing sodiummg

```
## Stratified 1 - level Cluster Sampling design (with replacement)
## With (879) clusters.
## svydesign(ids = ~area, weights = ~wti_UKY1tol1, strata = ~astrata1,
##         data = ndns_1_11, nest = TRUE)
```

```
##
## Call: svyglm(formula = omsysval ~ Age + Sex + EnergykJ_4 + bmival
+
##      educfinh + EIMD_2010_quintile, design = ndns_1_1le, na.action =
na.exclude)
##
## Coefficients:
##      (Intercept)                Age                SexFemale
##      5.763e+01                5.893e-01            -4.908e+00
##      EnergykJ_4                bmival                educfinh2
##      3.235e-04                1.430e+00                3.517e+00
##      educfinh3                educfinh4                educfinh5
##      -2.239e+01               -2.217e+01            -2.077e+01
##      educfinh6                educfinh7                educfinh8
##      -2.866e+01               -1.765e+01            -1.797e+01
## EIMD_2010_quintile2  EIMD_2010_quintile3  EIMD_2010_quintile4
##      1.785e+00                -1.111e+00            -5.044e+00
## EIMD_2010_quintile5
##      -3.936e+00
##
## Degrees of Freedom: 4428 Total (i.e. Null);  428 Residual
## (11226 observations deleted due to missingness)
## Null Deviance:      1.8e+07
## Residual Deviance: 15580000  AIC: 48000

## Anova table: (Rao-Scott LRT)
## svyglm(formula = omsysval ~ Age, design = ndns_1_1le, na.action =
na.exclude)
##
##      stats      DEff      df ddf      p
## Age      1769230    5152.1    1.0 856 < 2.2e-16 ***
## Sex       66818     5415.5    1.0 855 0.0004998 ***
## EnergykJ_4 108159    6443.9    1.0 854 4.954e-05 ***
## bmival     1922691    7598.4    1.0 847 < 2.2e-16 ***
## educfinh    6850386    4791.3    7.0 740 < 2.2e-16 ***
## EIMD_2010_quintile 3600686    6759.3    4.0 428 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

the UPF becomes significant! This suggests that the effect of UPF is mediated by Sodium!!  
comparing AIC for tghese three models

```
lmM$aic
## [1] 48000.08

lmMna$aic
## [1] 48001.91

lmM2$aic
```



```
## [1] 48000.75
```

```
#AIC(lmM, lmMna, lmM2)
```

we find that the lowest AIC is given by the model without UPF!! Though all the models with UPF have a lower aic than the model without.

## Summary of Results

There is a table with summary values for the key 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 UPF.

## 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.

The regression models identify that sodium intake is an important contributor to any optimal model. That when sodium intake is present UPF intake is no longer significant, but still has some effect.