# Part 1 – Introduction & Preliminary analysis

Data on 2,700 individuals were collected through an observational study. Information on age (categorised into 5-year groups, from 40-44 to 65-69), sex (female or male), education level (university graduate, or secondary education or less), smoking status (current smoker, ex-smoker, or never smoker), and waist to hip ratio was recorded for all 2,700 individuals. There are two outcomes of interest – a binary carotid plaque status (plaque absent or present), and a whole-number valued plaque count. Plaque status is recorded for all 2,700 individuals, and plaque count is recorded for 810 individuals.

Table 1 shows descriptive summary statistics for these variables, split by age & sex.

TABLE1 GOES HERE. MENTION AGE+SMOKING, AGE+EDUCATION, AND SMOKING+SEX

# Part 2 – Smoking & presence of carotid plaque

This section examines the relation between smoking and presence of carotid plaque. A logistic regression model is used to adjust for confounding due to age, sex, and education (a proxy for social class). First the data is aggregated by the confounders, and initial models are fit on the grouped data. Deviance is used to assess model fit, and more complex forms of the linear predictor are assessed through likelihood ratio tests. The model was than fit to the individual level data and used to estimate marginal odds ratios of carotid plaque presence in ex-smokers and current smokers, compared to never smokers. Adjusted marginal probabilities for presence of plaque were calculated using the margins command and by empirical standardisation. Finally, the model was used to estimate prevalence of carotid plaque presence if everyone in the sample never smoked.

An initial binomial model was fit to the grouped data. The model uses the individuals smoking status, age, sex, and level of education (all encoded as categorical variables), with no interactions between any of the variables. The deviance of this model is 63.6 on 62 degrees of freedom, comparing this to a distribution gives a p-value of 0.42 – there is no evidence of poor model fit.  
More complex models were also considered. Three interaction terms were separately added to the initial model – an interaction between smoking and age, between age and education, and between smoking and sex. A fourth model with interactions between all variables was also tried. None of the models had significantly better fit relative to the initial no-interaction model, so the additional terms were not included (p-values from likelihood ratio tests were 0.66, 0.89, 0.16, and 0.42 respectively).

The initial model was fit to the individual data and used to produce marginal probabilities of plaque presence. Marginal odds ratios are also calculated and are shown in table 2.

|  |  |  |
| --- | --- | --- |
| Smoking status | Marginal probability | Marginal odds ratio |
| Never smoker | 0.59 [0.57, 0.62] |  |
| Ex-smoker | 0.61 [0.58, 0.65] | 1.12 [0.90, 1.33] |
| Current smoker | 0.72 [0.68, 0.75] | 1.92 [1.43, 2.17] |
| **Table 2**: Estimated marginal probability of plaque presence by smoking status. Notes: 95% confidence intervals shown in square brackets. Estimates are adjusted for age, sex, and education. Estimates are rounded and may not line up with calculations quoted in the text. Odds ratio CIs based on percentile bootstrap estimates. | | |

The marginal probabilities in table 2 can be interpreted as prevalence of carotid plaque in the situation where all study participants have the same smoking status. For example, the prevalence would be 59% if no one in the study smoked. Assuming smoking has a causal effect on carotid plaque, this means that 3.5% (95% CI [1.0%, 6.0%]) of carotid plaque prevalence is due to smoking. If no one in the study ever smoked, and assuming that the effect of smoking is causal, there would be between 27 and 163 fewer carotid plaque cases.

# Part 3 – Waist to hip ratio & presence of carotid plaque

Next, waist to hip ratio (W2H) is added to the model from part 2, and the effect of adding W2H is quantified using a likelihood ratio test. The need for potential interactions were examined through lowess plots. Interactions between W2H and smoking, and W2H and age were examined using likelihood ratio tests but not added. This ‘no-interaction’ model is used to quantify the effect of increasing W2H has on presence of carotid plaque.  
An interaction term between W2H and sex is then added, in order to examine how the association between W2H and presence of carotid plaque varies by sex. This model is then used to examine the prevalence of carotid plaque if the study population all followed WHO guidelines on obesity.

A likelihood ratio test between the models with & without W2H was highly significant (19.9 on 1 degree of freedom, p < 0.0001), there is strong evidence that adding W2H improved model fit.   
Model fit was not improved by including interaction terms between W2H and smoking, or W2H and age (likelihood ratio test p-values 0.42 and 0.83 respectively).

The no-interaction model estimates that an increase in W2H of 0.1 leads to an increase in the odds ratio of having carotid plaque by 1.30 (95% CI [1.16, 1.46]), adjusting for age, education level, and smoking status. Including an interaction between sex and W2H does not significantly improve model fit (LR test p-value = 0.69) but is included to quantify how the effect of W2H varies by sex. This model estimates that a 0.1 unit increase in W2H leads to an increase in the odds ratio of having carotid plaque by 1.28 (95% CI [1.11, 1.47]) in females, and by 1.34 (95% CI [1.11, 1.62]) in males. These estimated odds ratios are adjusted for the same variables as the no-interaction model. There is no evidence that the effect of W2H varies by sex (coefficient Wald test p-value = 0.69).

The effect of reducing levels of obesity was quantified. Values of W2H which are classed by WHO as obese were reduced – males with W2H above 0.9 had their W2H value set to 0.9, and females with W2H above 0.85 had their W2H values set to 0.85. The ‘sex-interaction’ model was used since the WHO guidance varies by sex.

# Part 4 – Smoking and waist to hip ratio & plaque count

Models using the variables from the previous two sections are now fit to plaque count. From the results of part 3, a Poisson regression model including all five variables and no interactions is taken as an initial model.