Results:

H1

An ordinal logistic regression was conducted to predict smoking status (1 = never smoked, 2 = former smoker, 3 = current smoker) from age and sex. Age was modeled using a natural spline function (df = 3) to account for potential non-linear effects, and sex was included as a covariate. A likelihood-ratio test comparing the linear age model with the spline model indicated that the non-linear specification provided a significantly better fit, $\Delta\chi^2(2) = 8{,}729$, $p < .001$. The proportional odds assumption was met, as indicated by a non-significant nominal test.

Results showed a significant non-linear association between age and smoking status. The first spline component was negatively associated with smoking status, $\beta = -0.192$, SE = 0.011, $z = -18.20$, $p < .001$, whereas the second spline component was not statistically significant, $\beta = 0.035$, SE = 0.026, $z = 1.33$, $p = .183$. The third spline component showed a strong negative association, $\beta = -1.481$, SE = 0.017, $z = -86.66$, $p < .001$, confirming a pronounced non-linear age effect.

Sex was a highly significant predictor, with males having substantially higher odds of belonging to higher smoking categories compared to females, $\beta = 3.520$, SE = 0.007, $z = 508.48$, $p < .001$. The estimated threshold parameters were 3.009 for the transition from never smokers to former smokers and 4.247 for the transition from former smokers to current smokers.

Visual inspection of model-based predicted probabilities indicated a non-linear age pattern, with the highest probability of current smoking observed around 40 years of age, followed by a gradual decline at older ages. This pattern was more pronounced among males.

H2

A multiple linear regression analysis was conducted to examine differences in body weight across smoking status while adjusting for age and sex. The overall model was statistically significant, $F(4, 991{,}341) = 140{,}000$, $p < .001$, explaining 36.1% of the variance in body weight ($R^2 = .361$).

Age was negatively associated with body weight, such that weight decreased by approximately 0.13 kg per year of age, $b = -0.13$, SE < 0.01, $p < .001$. Sex emerged as a strong predictor, with males weighing on average 13.75 kg more than females, $b = 13.75$, SE = 0.03, $p < .001$.

Smoking status was also significantly associated with body weight, showing both a significant linear trend, $b = 0.37$, SE = 0.02, $p < .001$, and a significant quadratic trend, $b = -0.75$, SE = 0.02, $p < .001$, indicating a non-linear relationship across smoking categories. Former smokers exhibited higher body weight compared to both never smokers and current smokers.

Effect sizes were examined using partial eta squared (η^2_p) to assess practical relevance. Sex showed a large effect on body weight ($\eta^2_p = .34$), whereas age showed a small-to-moderate effect ($\eta^2_p = .06$). Smoking status exhibited a very small effect ($\eta^2_p = .001$). Although all predictors reached statistical significance due to the large sample size, only sex demonstrated a practically meaningful association with body weight.

H3

A multivariate analysis of covariance (MANCOVA) was conducted to examine the association between smoking status and lipid profile (HDL and LDL cholesterol), adjusting for age and sex. Using Pillai's trace as the multivariate test statistic, a significant multivariate effect of smoking status was observed, $V = .041$, $F(4, 1{,}982{,}588) = 10{,}325$, $p < .001$. Significant multivariate effects were also found for age, $V = .021$, $F(2, 991{,}293) = 10{,}868$, $p < .001$, and sex, $V = .056$, $F(2, 991{,}293) = 29{,}604$, $p < .001$.

Follow-up linear regression analyses were conducted separately for HDL and LDL cholesterol, additionally adjusting for alcohol consumption (DRK_YN).

For HDL cholesterol, smoking status was significantly associated with HDL levels, $F(2, 991{,}293) = 1{,}388$, $p < .001$. Both a significant linear trend, $b = -1.61$, SE = 0.03, $p < .001$, and a significant quadratic trend, $b = -0.83$, SE = 0.03, $p < .001$, were observed. Increasing age was associated with lower HDL cholesterol, $b = -0.12$, SE < 0.01, $p < .001$, and males had substantially lower HDL levels than females, $b = -10.06$, SE = 0.04, $p < .001$. Alcohol consumption was associated with higher HDL cholesterol, $b = 4.75$, SE = 0.03, $p < .001$. The model explained 12.9% of the variance in HDL cholesterol ($R^2 = .129$).

For LDL cholesterol, smoking status was also significantly associated with LDL levels, $F(2, 991{,}293) = 6.07$, $p = .002$, and remained significant after Holm correction ($p = .007$). A small but significant linear trend was observed, $b = -0.23$, SE = 0.08, $p = .003$, whereas the quadratic trend was not significant, $b = 0.07$, SE = 0.08, $p = .36$. Increasing age was associated with slightly

higher LDL cholesterol, b = 0.05, SE < 0.01, p < .001. Male sex was associated with marginally higher LDL levels, b = 1.05, SE = 0.09, p < .001, whereas alcohol consumption was associated with lower LDL cholesterol, b = −3.03, SE = 0.08, p < .001. The model explained a very small proportion of variance in LDL cholesterol (R^2 = .003).

Overall, smoking status showed a statistically significant association with lipid profile at both the multivariate and univariate levels. However, the association was substantially stronger for HDL cholesterol than for LDL cholesterol. Visualization suggested sex-specific patterns, with smoking status being more strongly related to HDL cholesterol among women than among men, whereas corresponding differences in LDL cholesterol were comparatively small.