MethAge performance in blood donor samples with a focus on outliers

Kotková L, Slavkovský R, Vrbková J, Drábek J

IMTM, Faculty of Medicine and Dentistry, Palacký University Olomouc
Contact: lucie.kotkova01@upol.cz

Introduction

MethAge is an epigenetic method for predicting chronological age from blood samples based on methylation level at seven CpGs in five genes (*MIR29B2CHG, FHL2, ELOVL2, CCDC102B,* and *PDE4C*).

The prediction model using Bayesian inference was developed on 200 samples from healthy Czech blood donors (ENIGMA cohort) and validated on 841 samples from the same cohort.

	Whole cohort	Training cohort	Validation cohort
Number of samples	1041	200	841
Percentage of women	19.1%	14.5%	20.2%
Mean age	39.6 years	40.9 years	39.3 years
Mean absolute error	3.55 years	2.38 years	3.83 years
Samples inside CI	80.2%	93.5%	77.1%

In this work, we focus on outliers in order to explore the underlying biological mechanism that leads to lower prediction precision in these samples.

We defined three outlier groups: Group A - prediction outside the method defined Bayesian credible interval; Group B - absolute prediction error > 7 years; and Group C - absolute prediction error > 10 years.

	Whole cohort	Outliers A	Outliers B	Outliers C
Number of samples	1041	206 (19.8%)	130 (12.5%)	38 (3.7%)

Age and sex
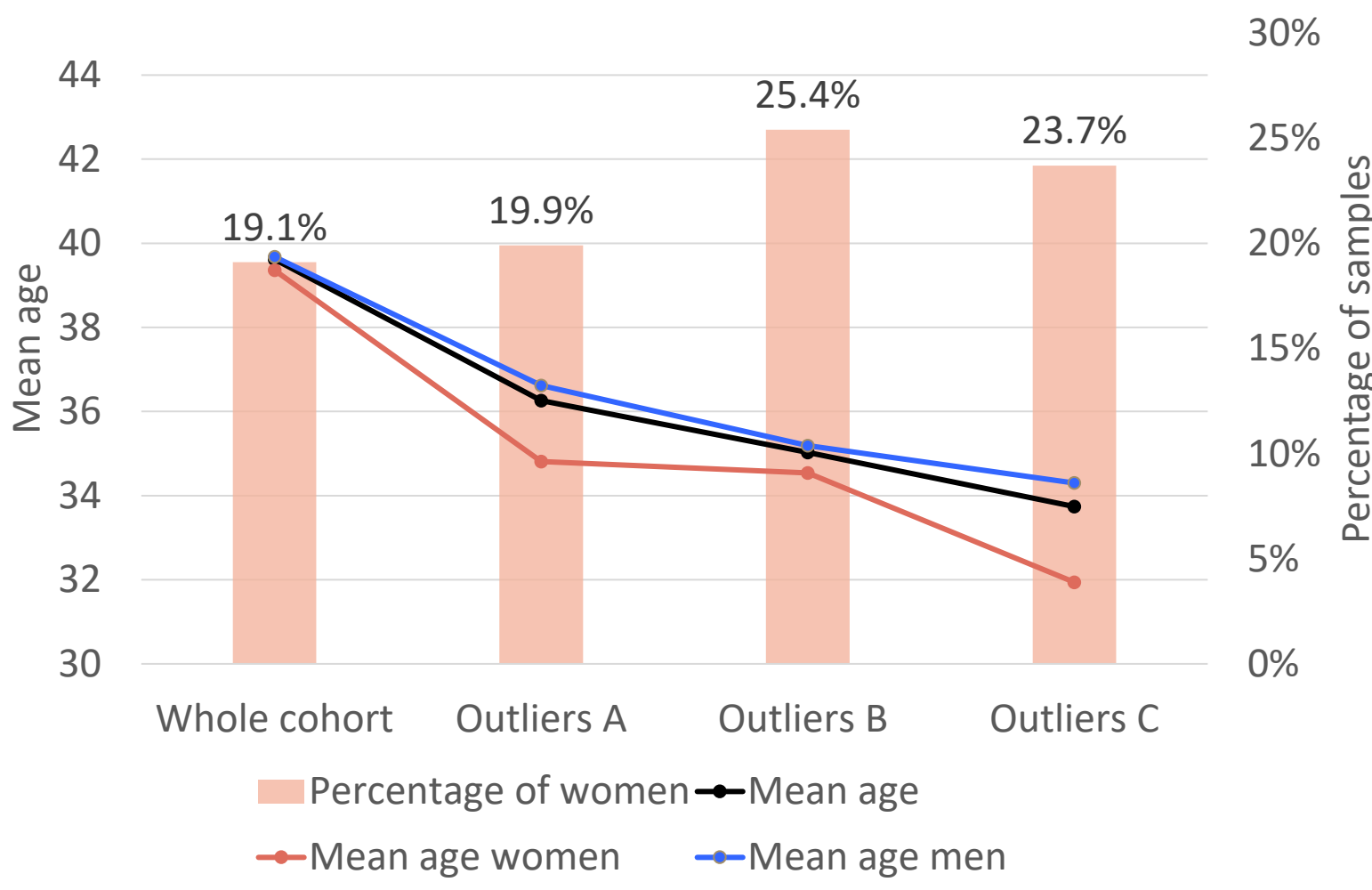
	Whole cohort	Outliers A	Outliers B	Outliers C
Predicted younger	33.6%	16.0%	13.1%	10.5%
Women predicted younger	26.6%	9.8%	9.1%	
Men predicted younger	35.3%	17.6%	14.4%	

Outlier groups tend to have:
- **A higher proportion of women**
- **Lower mean age**
- A lower proportion of individuals predicted younger than their chronological age

BMI
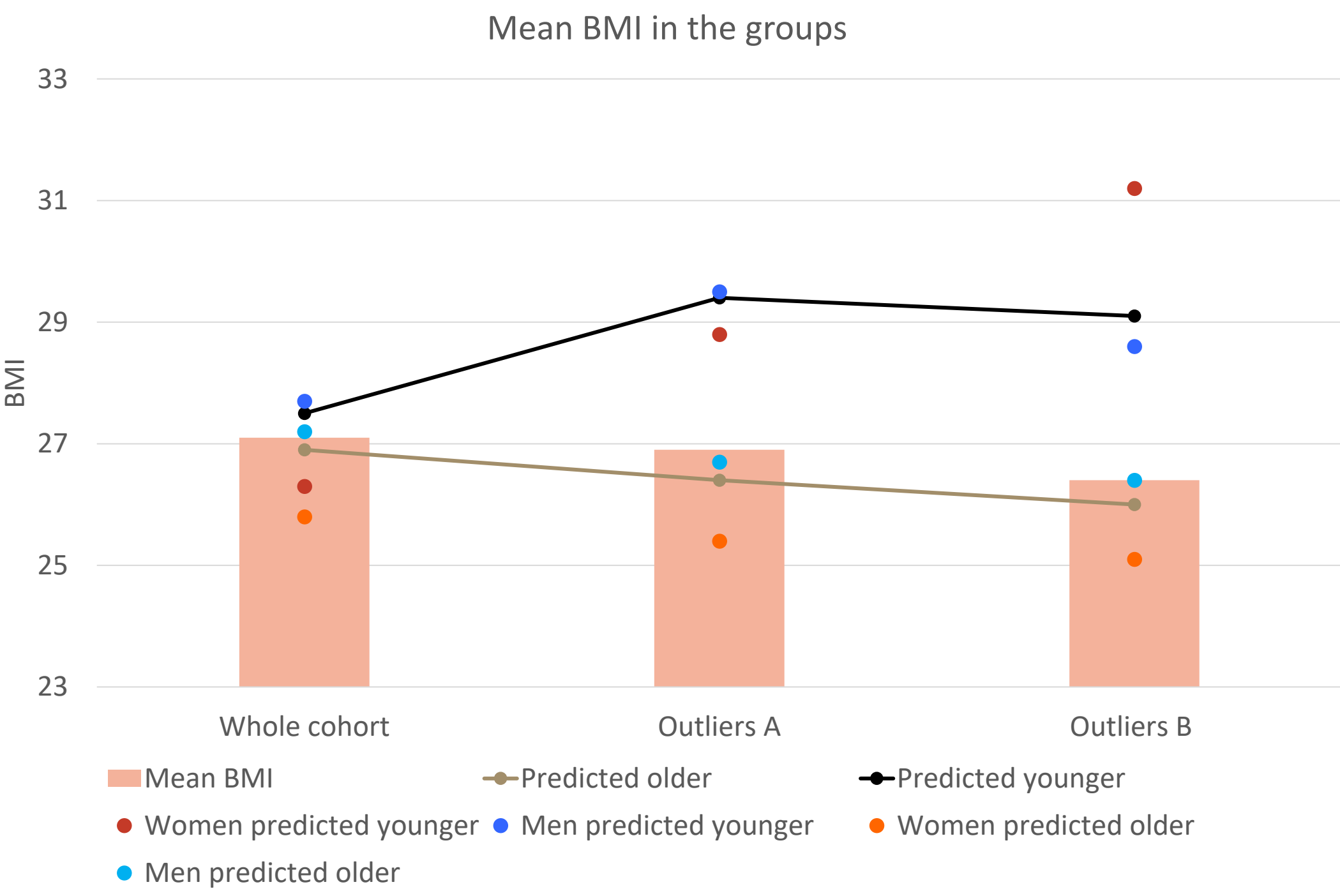
Among outliers, mean **BMI** tends to be lower in individuals predicted older and **higher in individuals predicted younger**. This difference is more pronounced in women.

Reproductive health (women only)

	Whole cohort	Outliers A	Outliers B	Outliers C
Data available	190	39	31	9
Pregnancies				
0	25.8%	46.2%	45.2%	66.7%
1	16.3%	15.4%	16.1%	-
multiple	57.9%	38.5%	38.7%	33.3%
Oral contraceptives use	3.50%	5.10%	6.50%	22.20%

Among outliers, we observed an increased proportion of women with no pregnancy history as well as a higher proportion of women using oral contraceptives (from them 57% had no pregnancy, 29% one and 14% two). These findings, however, are based on small numbers and should be interpreted with caution.

	Whole cohort	Outliers A	Outliers B	Outliers C
Medical history of				
- skin disease	1.2%	2.9%	3.1%	5.3%
- allergies	5.3%	4.4%	4.6%	10.5%
- infections (mononucleosis, tick-born encephalitis etc.)	2.3%	4.4%	3.8%	5.3%
Reported illnes in family	4.0%	7.8%	9.2%	15.8%
- from that women	45.2%	68.8%	66.7%	50.0%

Lifestyle factors

n = 120 samples with available data on smoking, alcohol intake, and diet
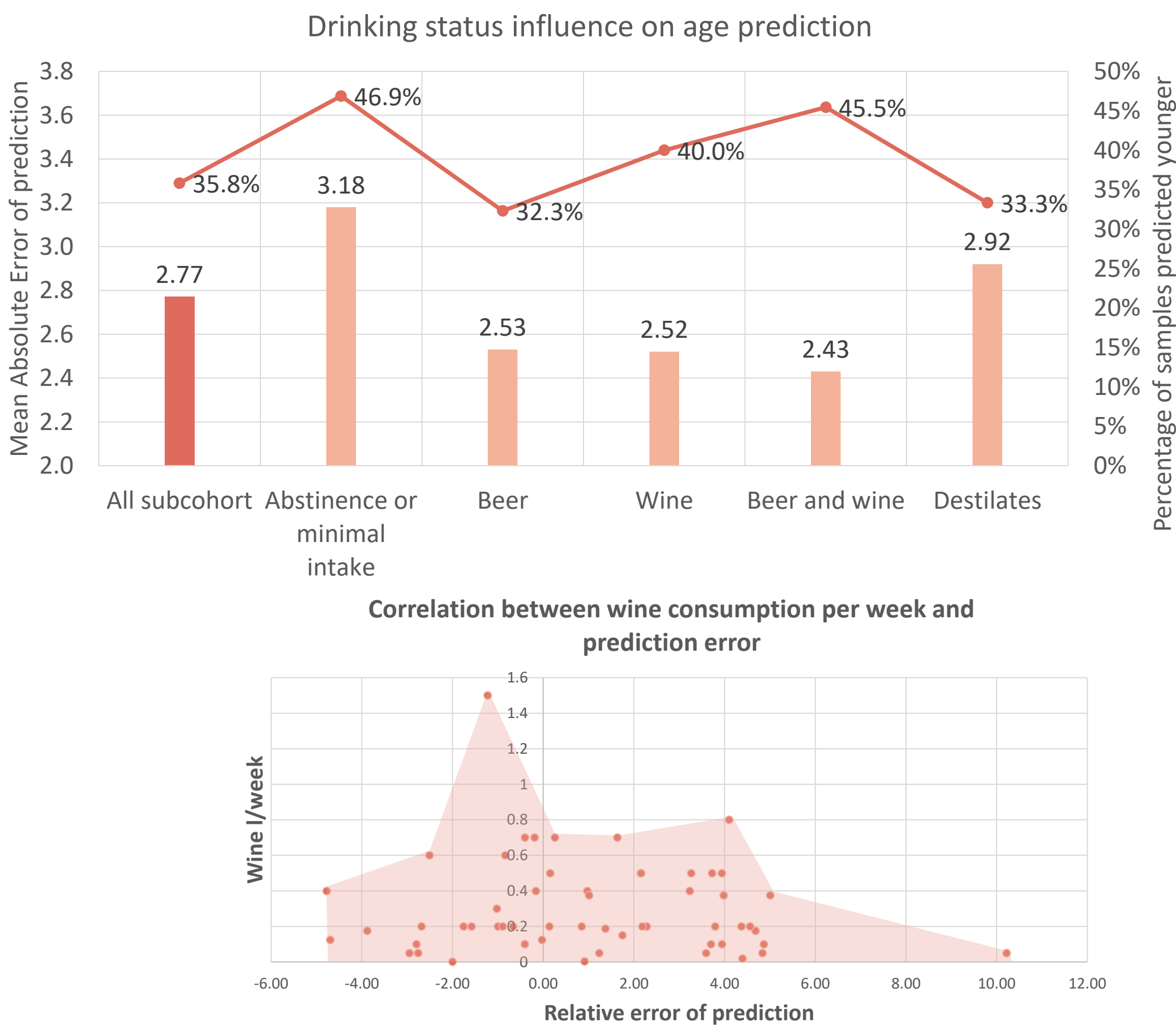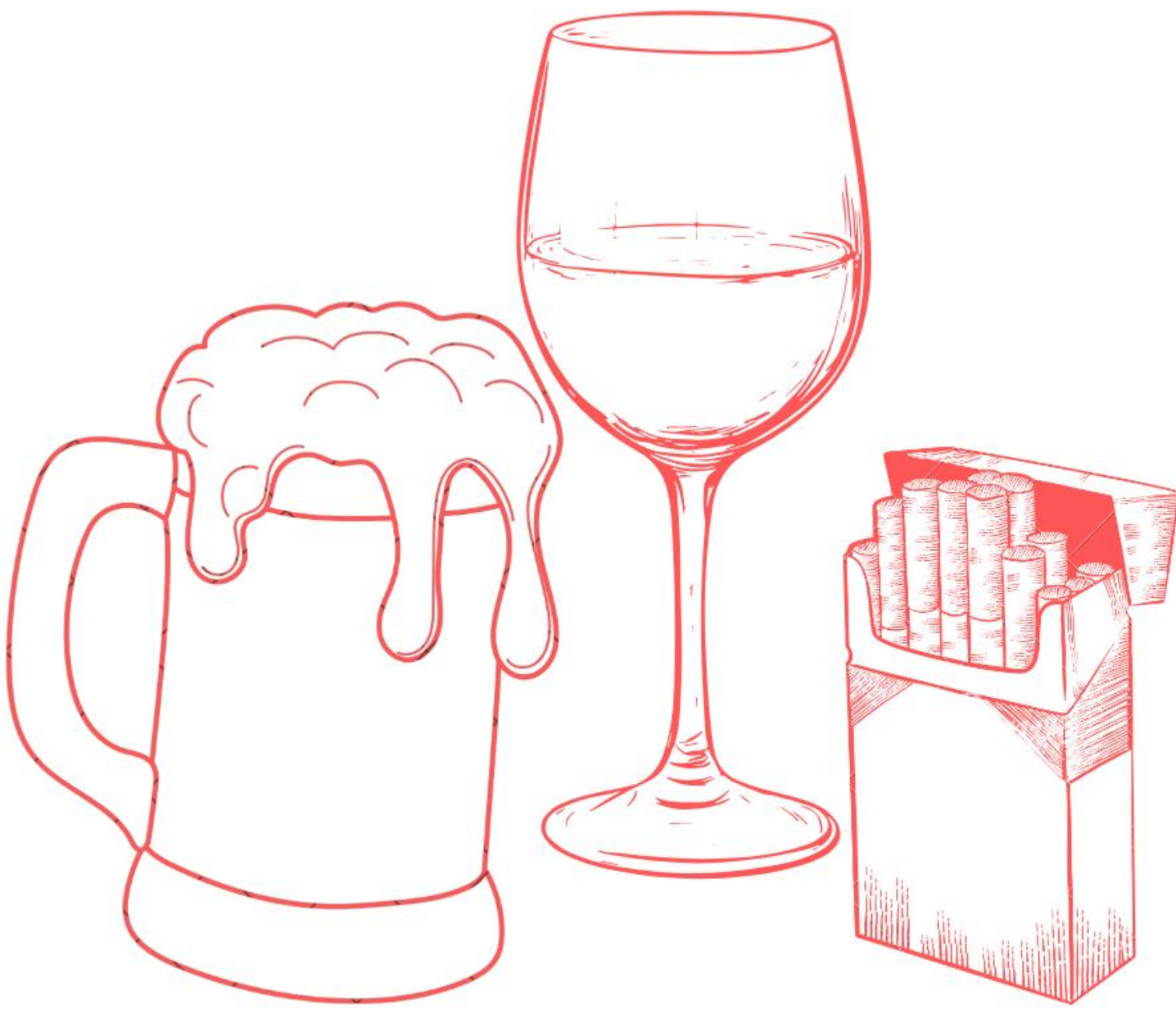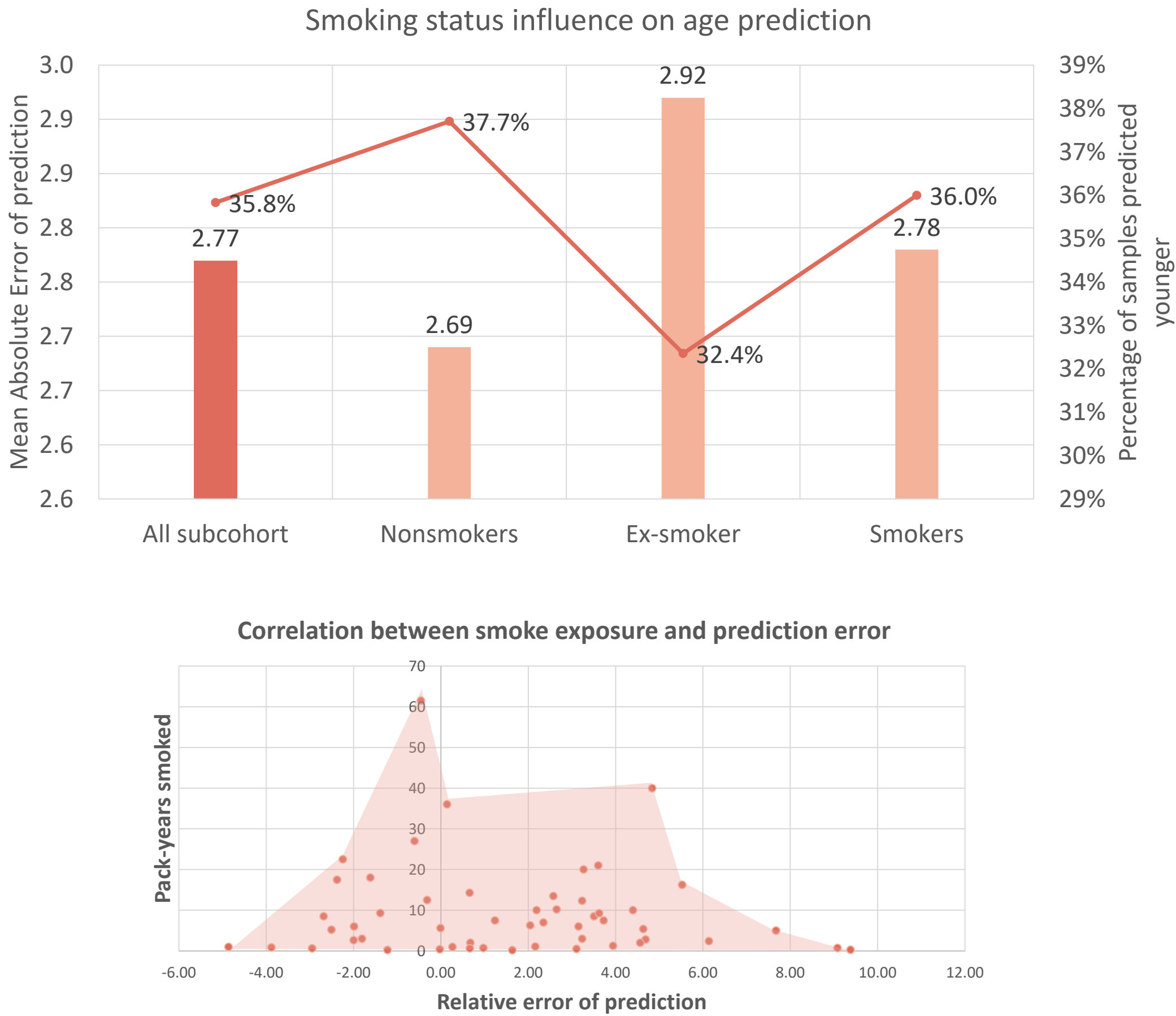
We observed the lowest mean absolute error (MAE) and the highest proportion of individuals predicted younger than their chronological age among non-smokers (n = 61), and the worst values for both metrics in ex-smokers (n = 34).

MAE increased in the group with no or minimal alcohol intake (n = 32), probably due to the higher proportion of individuals predicted younger in this group. A similar increase in the proportion of individuals predicted younger was observed among those who drink wine (n = 55), and especially among those who consume both wine and beer (n = 22). However, the expected clear dose–response relationship was not observed for either smoking or alcohol intake. Because only a few individuals reported specific dietary patterns, we could not perform a separate analysis for diet.

Conclusion

In the ENIGMA cohort, 80.2% of samples lie within the credible interval. Only 3.7% of samples were predicted with an error > ±10 years. Prediction precision is lower in women, and differences in other variables tend to be more prominent among them, probably due to the underrepresentation of women in the training set. Contrary to the literature, mean age decreases in outlier group B, suggesting that the Bayesian approach is particularly useful for samples from higher age categories. We also observe a higher BMI in individuals predicted younger. This may reflect higher muscle mass (and a generally healthier lifestyle), which cannot be distinguished from obesity when using BMI alone. Some potentially interesting categories (e.g. medical history) contain only a few individuals, which limits the statistical power of the analyses.