

Supplementary Material 1. Detailed data analysis methods.**Regression assumptions and model validation:**

To ensure completeness of the statistical analysis, several diagnostic checks were performed:

- Multicollinearity: variance inflation factor values ranged between 1.010 and 1.096, with tolerance values above 0.90, confirming the absence of multicollinearity among predictors.
- Influential outliers: Standardized residuals ranged from -1.54 to $+1.54$, and Cook's distance values were all below 0.032 (range, 0.005–0.032), indicating no undue influence on the regression models.
- Model fit: The Hosmer-Lemeshow goodness-of-fit test ($\chi^2 = 3.766$, degrees of freedom [df] = 8, $p = 0.878$) demonstrated adequate calibration of the logistic regression models.
- Overall model significance: The Omnibus Test of Model Coefficients ($\chi^2 = 195.131$, df = 7, $p < 0.001$) confirmed statistical significance of the models.
- Explained variance: Nagelkerke R^2 was 0.268, indicating acceptable explanatory power for behavioral risk factors.

Co-occurrence analysis

Co-occurrence of risk behaviors was operationally defined as the presence of 2 or more unfavorable lifestyle risk factors (unhealthy diet, physical inactivity, tobacco use, alcohol intake).

- Pairwise co-occurrence was assessed descriptively to identify common behavioral clusters.
- Logistic regression was then applied to the binary outcome (≥ 2 vs. < 2 risk factors) to evaluate predictors of clustering.