scientific reports

Check for updates

OPEN

Exploring latent classes of complete blood count profiles and their association with smoking status in the Bandar Kong cohort study

Ali Mouseli[1], Pejman Shahrokhi[2], Mehdi Sharafi[3]✉, Zahra Mastaneh[4], Abdollah Gharibzadeh[5], Leila Jamali[6] & Arash Shekari[7]

Smoking is a significant modifiable risk factor influencing various health outcomes, including hematologic indices. This study investigates the association between smoking status and latent classes of complete blood count (CBC) profiles in older adults. Data were analyzed from the baseline phase of the Bandar Kong Cohort Study, including adults aged ≥ 35 years. Latent class analysis (LCA) was employed to identify subgroups based on CBC indices. Logistic regression was applied to examine associations between smoking status and class membership, adjusting for potential confounders. Four latent classes were identified: Class 1 (microcytic normochromic anemia; prevalence: 39.3%), Class 2 (beta-thalassemia minor; 13.8%), Class 3 (iron deficiency anemia; 34.4%), and Class 4 (mixed anemia; 12.5%). Current smoking was significantly associated with higher odds of membership in Class 2 (OR = 1.51, 95% CI = 1.05–2.16) and Class 4 (OR = 1.50, 95% CI = 1.08–2.10) compared to Class 1. Former smoking showed no significant associations. Smoking is significantly associated with specific CBC profiles, particularly beta-thalassemia minor and mixed anemia. These findings underscore the importance of considering smoking history in anemia diagnosis and management while highlighting the potential benefits of smoking cessation in mitigating adverse hematologic effects.

Keywords Latent class analysis, Blood cell count, Smoking, Beta-thalassemia, Anemia, Bandar kong cohort study, PERSIAN cohort

Cigarette smoking remains one of the most significant preventable causes of morbidity and mortality worldwide. According to the 2017 Global Burden of Disease Study, smoking was identified as the second most critical risk factor for premature mortality and disability globally[1]. In 2015, smoking contributed to 6.4 million deaths worldwide, with projections suggesting that tobacco-related fatalities rise to 8 million annually by 2030[2]. This health burden is compounded by the economic costs associated with smoking-related diseases, which, in high-income countries, account for 6.5% of total healthcare spending and 2.2% of GDP[3]. In Iran, smoking prevalence studies show that approximately 20% of men are smokers, with notable increases observed among young adults[4,5]. Additionally, a 2016 national survey reported a prevalence of 24.4% in men and 3.8% in women[6].

The adverse effects of smoking extend to various health domains, including cardiovascular diseases, respiratory disorders, multiple cancers, and reproductive complications. However, its impact on the hematopoietic system and blood parameters has received relatively limited attention[7,8]. While some studies report that smoking elevates red blood cell count (RBC), hematocrit (HCT), hemoglobin (HGB), mean corpuscular volume (MCV),

[1]Social Determinants in Health Promotion Research Center, Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. [2]Department of Internal Medicine,Faculty of Medicine , Hormozgan University of Medical Sciences, Bandar Abbas, Iran. [3]Cellular and Molecular Research Center , Gerash University of Medical Sciences, Gerash, Iran. [4]Department of Health Information Management and Technology, School of Allied Medical Sciences, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. [5]Social Determinants in Health Promotion Research Center,Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. [6]Cardiovascular Diseases Research Institute, Shiraz University of Medical Sciences, Shiraz, Iran. [7]Cardiovascular Diseases Research Institute, Tehran Heart Center, Tehran University of Medical Sciences, Tehran, Iran. ✉email: mehdisharafi_2002@yahoo.com

and mean corpuscular hemoglobin (MCH), others have found conflicting results, particularly concerning white blood cell count (WBC) and platelet count (PLT)[9–12]. Nicotine is hypothesized to affect leukocyte count through hormonal stimulation or by triggering respiratory inflammation, leading to cytokine release[13,14]. The variability in these findings reflects the need for further investigation, particularly in understudied populations, such as those in Iran[8,15,16].

Hormozgan, the location of this study, has one of the highest anemia prevalence rates in Iran, with significant contributions from conditions like beta-thalassemia and sickle cell anemia[16]. Anemia remains a public health concern in this region due to its multifactorial etiology, including nutritional deficiencies, genetic disorders, and chronic diseases. Hence, there is a need to explore the interaction between smoking and hematological indices in regions with high anemia prevalence.

The present study focuses on the association between cigarette smoking and latent classes of complete blood count (CBC) components in a region with a notably high prevalence of anemia with various etiologies. Unlike previous studies that primarily assessed the effects of smoking on individual CBC components, this research employs latent class analysis (LCA) to uncover hidden patterns within hematological parameters. LCA is a robust statistical method that identifies unobserved subgroups within a population, offering deeper insights into the complex interactions between smoking and hematological indices. This innovative approach not only yields a more comprehensive understanding of the systemic impact of smoking but also indicates the importance of considering regional health challenges, such as anemia, in designing targeted public health interventions.

Materials and methods

Study population

This cross-sectional study utilized baseline data from the Bandar Kong cohort of the larger PERSIAN Cohort study, which commenced in 2016 to investigate chronic non-communicable diseases (NCDs) and cardiovascular risk factors. The baseline data were collected after two years, and the cohort is designed for a 15-year follow-up with repeated surveys every five years. The primary objective of the Bandar Kong Cohort Study is to assess various risk factors associated with NCDs. A total of 4,200 participants were initially enrolled from 6,000 permanent residents aged 35–70 years in Bandar Kong, southern Iran, in line with PERSIAN Cohort arrangements[17,18]. The inclusion criteria for this study required participants to be aged 30 to 70 years, to have the ability to provide informed consent and complete questionnaires, and to have lived in Bandar Kong for at least 10 years. Participants with no interest to attend or mental or physical disabilities were excluded. The cohort is designed to comprehensively evaluate health-related factors in the population using a combination of demographic, socioeconomic, anthropometric, nutritional, and lifestyle data, as well as biological samples. The cohort profile has been previously published, providing a detailed description of the study design and population[19].

Measurements

Blood samples and laboratory analyses were integral to this study. A 25 mL blood sample was collected from each participant during the baseline assessment, adhering to protocols established by the PERSIAN Cohort. All blood samples were processed in a central laboratory equipped with advanced biochemistry analyzers (BT1500, Italy, 2017) and hematology counters (Mindray, China, 2017). Following collection, blood samples were aliquoted into specific cryo tubes for long-term storage at −80 °C, including whole blood, plasma, buffy coat, red blood cells, and serum, as per the PERSIAN Cohort guidelines. These stored samples allow for future analysis beyond the scope of this study.

Complete blood count (CBC) parameters were measured, including White Blood Cell Count (WBC), Red Blood Cell Count (RBC), Hemoglobin (HGB), Hematocrit (HCT), Platelet Count (PLT), Lymphocyte-to-Monocyte Ratio (LMR), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and Red Cell Distribution Width - Coefficient of Variation (RDW-CV). Each parameter was classified based on its normal range, with values categorized as within or above the normal range.

In addition to hematological parameters, demographic and anthropometric data were recorded. Demographic variables included sex, age, years of education, and marital status. Anthropometric measurements, including body weight, height, waist circumference (WC), and hip circumference (HC), were performed by trained staff using standardized equipment and procedures. Body weight was measured to the nearest 10 g using a portable digital scale (RGZ 160, China), while height was measured to the nearest 0.5 cm using a wall-mounted stadiometer (Height Rod Wall, FAZZINI S208, Italy, 2015). Waist and hip circumferences were measured using a SECA 201 retractable tape (SECA GmbH, Germany). Derived metrics, such as body mass index (BMI) and waist-to-hip ratio, were calculated to evaluate participants' nutritional and metabolic status. Moreover, dietary intakes and patterns over the previous year were assessed using a validated, 132-item food frequency questionnaire (FFQ) specific to the Iranian general population. In the current study, the daily energy intake, which is a general indicator derived from FFQ, was selected to reflect the dietary pattern.

The smoking status of participants was assessed using a validated framework from the National Institutes of Health (NIH), which categorizes individuals into four groups: (1) Current smokers: Those actively smoking at the time of assessment; (2) Former smokers: Those who smoked at least 100 cigarettes during their lifetime but were not currently smoking; (3) Never smokers: Those who had never smoked or had only experimented with smoking briefly; (4) Passive smokers: Non-smokers exposed to environmental tobacco smoke (ETS), also referred to as secondhand smoke.

Additionally, the medical history of chronic diseases, including diabetes, hypertension, and thyroid disease, was collected using a detailed questionnaire. This information was obtained through face-to-face interviews conducted by trained interviewers, ensuring data reliability and accuracy.

Statistical analysis

Latent Class Analysis (LCA) was employed to identify distinct latent classes based on participants' CBC indices. LCA is a statistical technique that identifies subgroups within a population, where the subgroup memberships are not directly observed but inferred from the data. The LCA method was used to explore the association between latent classes of CBC indices and smoking status while controlling for relevant demographic and health factors.

LCA was performed on the following CBC parameters: WBC, RBC, HGB, HCT, PLT, LMR, MCV, MCH, MCHC, and RDW-CV. These parameters were categorized into "upper" and "under" normal ranges, based on established clinical guidelines. The LCA models were estimated using maximum likelihood estimation, testing between 2 and 5 latent classes. The model selection process was based on fit indices including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the G^2 statistic. The AIC and BIC are widely used model selection criteria that balance model fit with complexity, penalizing models with more parameters to avoid overfitting. The BIC, in particular, is known to favor simpler models and has been shown to be the most accurate criterion for selecting the optimal number of latent classes in Monte Carlo simulations. A model with the lowest AIC and BIC values was selected as the best-fitting model.

Each participant was assigned to the latent class with the highest predicted probability of membership. Following this, multinomial logistic regression was used to examine the association between smoking status and latent class membership. The regression models were adjusted for potential confounders, including age, sex, education, marital status, chronic diseases (diabetes, hypertension, thyroid disease), anthropometric measurements (BMI, WC, HC), and FFQ's energy intake. Odds Ratios (OR) with 95% confidence intervals (CI) were reported for the association between smoking status and membership in the identified latent classes.

For comparisons across latent classes based on sex, Chi-square tests were used for categorical variables, while independent t-tests were applied to continuous variables. The differences in demographic and health characteristics across latent classes were assessed using ANOVA for continuous variables and Chi-square tests for categorical variables.

All statistical analyses were performed using R (version 4.1.1), with the *poLCA* package employed to conduct the LCA. Descriptive statistics were used to summarize the data, including means with standard deviations (SD) for continuous variables and frequencies with percentages for categorical variables.

It should be noticed that missing data were absent or minimal in our cohort, with < 5% missingness for any variable. Additionally, missingness apparently depended only on observed characteristics (MAR assumption). Hence, we implemented Multiple Imputation (MI), using the missRanger package in R, to avoid biases introduced by simpler methods. As a regulatory process in Bandar Kong Cohort Study, ten imputed datasets were generated, which incorporated all analysis variables to retain relationships.

Result

A total of 4,015 participants were included in this study, with a mean age of 48.28 ± 9.40 years. Among the study population, women comprised the majority (57.5%), and 89.41% of participants were married. The prevalence rates of diabetes mellitus, hypertension, and thyroid disease were 16.2%, 21.0%, and 9.6%, respectively. A significant disparity was observed in smoking habits, as men were far more likely to be current smokers compared to women (20.1% vs. 0.3%; $P < 0.001$). Anthropometric indices differed between sexes, with women showing significantly higher mean BMI, WC, and HC compared to men ($P < 0.001$ for all measures), while men showed significantly higher daily energy intakes ($P < 0.001$) (Table 1).

Men had a higher prevalence of normal values for HGB and HCT ($P < 0.001$). The majority of participants had normal WBC and RBC counts, with no significant differences observed between men and women ($P = 0.399$ and $P = 0.578$, respectively). However, PLT and LMR varied significantly between sexes, with women showing a lower prevalence of subnormal platelet counts but a higher prevalence of subnormal LMR values ($P < 0.001$). Women had a higher prevalence of MCH and MCHC values outside the normal range, with significant deviations toward lower and upper limits ($P < 0.001$). Similarly, RDW-CV values differed between sexes, with women showing a higher prevalence of abnormal RDW-CV values compared to men ($P < 0.001$) (Table 1).

Table 2 presents the results of the LCA for models with two to five classes, presenting key statistical metrics used to evaluate model fit. Among the models, the four-class solution showed the lowest AIC and BIC values compared to the three-class solution, indicating better fit and parsimony. Additionally, the four-class model exhibited clinical interpretability superior to the other models, supporting its selection as the most appropriate representation of the latent structure within the data.

Table 3 yields a detailed breakdown of the probabilities of blood indices across the four latent classes derived from the LCA. These classes, representing distinct hematological patterns in adults, were classified as follows:

Class 1 (Microcytic normochromic Anemia - MC-NC anemia) This class included 39.3% of participants and was characterized by predominantly normal RBCs but lower-than-normal MCV (95.6%). HGB levels were slightly below normal in 9.7% of individuals. Other blood indices such as PLT and WBC were within normal ranges for the majority.

Class 2 (Beta-Thalassemia minor) Accounting for 13.8% of participants, this group showed a distinctive pattern with nearly all individuals having RBC counts above the normal range (85.8%) but reduced HGB (99.2%) and HCT (84.6%) levels. MCV was lower in 71% of individuals, and MCH levels were also within the normal range for most, except in 4% who were on the lower side.

Variable	Male (n = 1,705)	Female (n = 2,310)	P [1]
Age, years, M ± SD	48.3 ± 9.5	48.20 ± 9.2	0.566 [2]
Education years, M ± SD	7.2 ± 4.8	4.7 ± 4.4	< 0.001 [2]
Marital status, n (%)			< 0.001
Single	39 (2.3)	386 (16.7)	
Married	1,666 (97.7)	1,924 (83.3)	
Tobacco smoking, n (%)			< 0.001
Never smoker	962 (56.4)	1,956 (84.7)	
Current smoker	343 (20.1)	6 (0.3)	
Ex-Smokers	231 (13.5)	2 (0.1)	
Passive smoker	169 (9.9)	346 (15.0)	
Diabetes, n (%)			< 0.001
Yes	230 (13.5)	424 (18.4)	
No	1,475 (86.5)	1,886 (81.6)	
Hypertension, n (%)			< 0.001
Yes	291 (17.1)	554 (24.0)	
No	1,414 (82.9)	1,756 (76.0)	
Thyroid disease, n (%)			< 0.001
Yes	56 (3.3)	333 (14.4)	
No	1,649 (96.7)	1,977 (85.6)	
BMI, kg/m^2, M ± SD	25.7 ± 4.4	27.7 ± 5.2	< 0.001 [2]
WC, cm, M ± SD	90.3 ± 11.1	96.2 ± 11.8	< 0.001 [2]
HC, cm, M ± SD	97.7 ± 8.1	101.8 ± 10.2	< 0.001 [2]
FFQ's energy intake, kcal, M ± SD	3,250.84 ± 1,005.26	2,559.21 ± 812.40	< 0.001 [2]
WBC, n (%)			0.399
Normal, 4.5–11 cells/μL	1,505 (88.3)	2,047 (88.6)	
Under normal range, < 4.5 cells/μL	162 (9.5)	225 (9.7)	
Upper normal range, > 11 cells/μL	38 (2.2)	38 (1.6)	
RBC, n (%)			0.578
Normal, 4.2–5.4 million cells/μL in female and 4.7–6.1 million cells/μL in male	1,225 (71.8)	1,670 (72.3)	
Under normal range, < 4.2 million cells/μL in female and < 4.7 million cells/μL in male	368 (21.6)	474 (20.5)	
Upper normal range, > 5.4 million cells/μL in female and > 6.1 million cells/μL in male	112 (6.6)	166 (7.2)	
HGB, n (%)			< 0.001
Normal, 12.3-15-3 g/dL in female and 13.8–17.2 g/dL in male	976 (57.2)	923 (40.0)	
Under normal range, < 12.3 g/dL in female and < 13.8 g/dL in male	684 (40.1)	1,355 (58.7)	
Upper normal range, > 15–3 g/dL in female and > 17.2 g/dL in male	45 (2.6)	32 (1.4)	
HCT, n (%)			< 0.001
Normal, 36.1–44.3% in female and 40.7–50.3% in male	1,054 (61.8)	1,207 (52.3)	
Under normal range, < 36.1% in female and < 40.7% in male	600 (35.2)	1,048 (45.4)	
Upper normal range, > 44.3% in female and > 50.3% in male	51 (3.0)	55 (2.4)	
PLT, n (%)			< 0.001
Normal, 150,000-450,000 cells/μL	1,623 (95.2)	2,194 (95.0)	
Under normal range, < 150,000 cells/μL	75 (4.4)	60 (2.6)	
Upper normal range, > 450,000 cells/μL	7 (0.4)	56 (2.4)	
LMR, n (%)			< 0.001
Normal, 4.12–26.67 in female and 3.46–26.50 in male	316 (18.5)	302 (13.1)	
<3.83	1,389 (81.5)	2,008 (86.9)	
MCV, n (%)			< 0.001
Normal, 90-92.5fL in female and 89.80–93.60 in male	215 (12.6)	145 (6.3)	
Under normal range, < 90 fL in female and < 89.80 in male	1,386 (81.3)	2,059 (89.1)	
Upper normal range, > 92.5 fL in female and > 93.60 in male	104 (6.1)	106 (4.6)	
MCH, n (%)			< 0.001
Normal, 27–30 pg/cell	691 (40.5)	770 (33.3)	
Under normal range, < 27 pg/cell	653 (38.3)	1,303 (56.4)	
Upper normal range, > 30 pg/cell	361 (21.2)	237 (10.3)	
Continued			

Variable	Male (n = 1,705)	Female (n = 2,310)	P [1]
MCHC, n (%)	1,070 (62.8)	1,342 (58.1)	< 0.001
Normal, 31.8–34.3 g/dL			
Under normal range, < 31.8 g/dL	282 (16.5)	655 (28.4)	
Upper normal range, > 34.3 g/dL	353 (20.7)	313 (13.5)	
RDW-CV, n (%)			< 0.001
Normal, 11–15%	1,406 (82.5)	1,638 (70.9)	
Upper normal range, > 15%	299 (17.5)	672 (29.1)	

Table 1. Demographic, clinical, and hematological characteristics of study participants by sex. [1] Chi-square test. [2] Independent t-test. BMI, Body mass Index; WC, Waist circumference; HC, Hip circumference; WBC, White blood cells; RBC, Red blood cells; HGB, Hemoglobin; HCT, Hematocrit; PLT, Platelet count; LMR, Lymphocyte-to-monocyte ratio; MCV, mean corpuscular volume; MCH, Mean corpuscular hemoglobin; MCHC, Mean corpuscular hemoglobin concentration; RDW, Red cell distribution width - Coefficient of variation; FFQ, Food frequency questionnaire.

Class	Number of estimated parameters	AIC	BIC	X^2	G^2	MLL
2	37	47030.08	47263.1	72198.12	7826.961	−23478.04
3	56	44946.44	45299.12	63722.3	5705.321	−22417.22
4	75	43664.74	44137.07	44803.91	4385.615	−21757.37
5	94	42883.39	43475.38	43895.81	3566.27	−21347.7

Table 2. Model fit statistics for LCA with varying class numbers. MLL, Maximum log-likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; G^2, Likelihood ratio/deviance statistic; X^2, Chi-square goodness of fit.

Class 3 (Microcytic hypochromic anemia - MC-HC anemia [Iron Deficiency Anemia, IDA]) Representing 34.4% of participants, this class had marked reductions in HGB (100%), HCT (76.7%), and MCV (100%) levels. The majority of individuals also displayed a high RDW-CV (54.6%), which is consistent with IDA.

Class 4 (Mixed Anemia) This class, which included 12.5% of participants, showed a mixed pattern of abnormalities consistent with coexisting IDA and megaloblastic anemia. Approximately 27.8% of individuals had reduced RBC counts, and 86% showed upper-normal or elevated MCH levels, suggestive of mixed anemia. Increased RDW-CV (14.8%) and a combination of other aberrations in blood indices further differentiated this class.

Table 4 presents the study population's socio-demographic and clinical characteristics and blood indices, stratified by latent classes. The mean age across the four classes differed statistically ($P < 0.001$), although the variations in absolute values are unlikely to hold clinical relevance. Women were significantly overrepresented in classes 1, 2, and 3 compared to men ($P < 0.001$), whereas men predominated in class 4, which might suggest distinct gender-based susceptibility patterns across the latent classes. Tobacco smoking showed significant variation between classes, with a higher prevalence of current smokers observed in classes 2 and 4 (11.2% and 15.4%, respectively, $P < 0.001$), suggesting the potential influence of smoking on hematological patterns in these groups. The prevalence of chronic conditions also varied notably across the classes. In class 1, the frequency of diabetes (18.5%) and hypertension (23.0%) was higher than in other classes, suggesting an association between chronic disease burden and the hematological profile characterized by microcytic normochromic anemia. Similarly, thyroid disease was more prevalent in class 3 (12.1%, $P = 0.002$), which aligns with the hematological presentation of microcytic hypochromic anemia observed in this class. Moreover, although BMI, WC, and HC showed statistically significant differences across the classes ($P < 0.001$), these variations lack clinical relevance given their narrow ranges. Thus, their contribution to distinguishing the latent classes might be of limited practical importance. Furthermore, the average daily energy intake was high among participants classified in class 4, which was statistically significant compared to that of other classes ($P < 0.001$).

Table 5 shows the results of multinomial logistic regression analyses assessing the association between tobacco smoking categories and latent class membership, adjusted for potential confounders such as age, sex, education, marital status, thyroid disease, diabetes, hypertension, BMI, WC, HC, and energy intake. Current smokers had significantly higher odds of belonging to both class 2 and class 4 compared to class 1. Specifically, the odds of being in class 2 were 1.49 times higher (OR = 1.49, 95% CI = 1.04–2.13, $P = 0.029$), and the odds of being in class 4 were also 1.49 times higher (OR = 1.49, 95% CI = 1.07–2.08, $P = 0.017$). However, no significant associations were observed for former smokers or passive smokers with any latent class.

Items	Latent classes			
	Class 1	Class 2	Class 3	Class 4
WBC				
Under normal range	0.079	0.149	0.092	0.098
Upper normal range	0.015	0.019	0.024	0.015
RBC				
Under normal range	0	0.858	0.146	0.278
Upper normal range	0.115	0	0.07	0.005
HGB				
Under normal range	0.097	0.992	1	0
Upper normal range	0.022	0	0	0.076
HCT				
Under normal range	0.067	0.846	0.767	0.082
Upper normal range	0.051	0	0	0.045
PLT				
Under normal range	0.037	0.036	0.021 0.036	0.047
Upper normal range	0.005	0.009		0.001
LMR				
< 3.83	0.858	0.847	0.854	0.790
MCV				
Under normal range	0.956	0.71	1	0.395
Upper normal range	0.002	0.111	0	0.257
MCH				
Under normal range	0.401	0.04	0.992	0
Upper normal range	0	0.21	0	0.86
MCHC				
Under normal range	0.149	0.084	0.494	0.008
Upper normal range	0.163	0.19	0.062	0.393
RDW-CV				
Upper normal range	0.137	0.051	0.546	0.148

Table 3. Blood indices (item-response) probabilities in the four-class latent model WBC, White blood cells; RBC, Red blood cells; HGB, Hemoglobin; HCT, Hematocrit; PLT, Platelet count; LMR, Lymphocyte-to-monocyte ratio; MCV, Mean corpuscular volume; MCH, Mean corpuscular hemoglobin; MCHC, Mean corpuscular hemoglobin concentration; RDW, Red cell distribution width - Coefficient of variation.

Discussion

We assessed the relationship between smoking habits and distinct anemia subtypes identified through LCA in a cohort of Iranian adults from a region with the highest anemia prevalence in the country. The main finding was the significant association between smoking and specific anemia phenotypes, with current smokers showing a higher likelihood of belonging to Class 2 (beta-thalassemia minor) and Class 4 (mixed anemia) compared to Class 1. No significant associations were observed for as former or passive smokers.

Our study showed notable sex differences across the identified latent classes of anemia. Females were predominantly classified in Class 3 (IDA), consistent with previous findings that anemia is more prevalent in females[20]. The most common anemia type in this population was hypochromic-microcytic anemia, often linked to iron deficiency, chronic diseases, and thalassemia[21,22]. Conversely, males were more prevalent in Class 4, which might be partially explained by the significantly higher smoking prevalence among males in our cohort (20.1% vs. 0.3% in females). This aligns with national data reporting that smoking rates in Iran are substantially higher in men compared to women (26.0% vs. 2.7%)[6]. The association between smoking and Class 4 could, therefore, account for the male predominance in this class.

In our study, notable differences in the prevalence of NCDs were observed across the anemia subtypes. Class 1, characterized by microcytic normochromic anemia, exhibited a higher prevalence of hypertension and type 2 diabetes mellitus. This association aligns with the phenotype of this class, which typically reflects anemia of chronic disease (ACD). Chronic conditions such as diabetes are known to be linked with mild-to-moderate anemia through mechanisms like elevated proinflammatory cytokines, including interleukin-6, which possess antierythropoietic effects, and reduced erythropoietic hormone levels due to diabetic nephropathy[23]. Additionally, thyroid diseases were more prevalent in Class 3, predominantly comprising individuals with IDA. This finding can be explained by the higher prevalence of IDA among females, coupled with evidence suggesting a bidirectional relationship between IDA and thyroid dysfunction. Iron deficiency can impair thyroid hormone synthesis by affecting enzymes such as thyroid peroxidase, which catalyzes the iodination of thyroglobulin, a precursor for thyroid hormones. Furthermore, IDA has been associated with disruptions in the hypothalamic–

Variable	Latent classes				P [1]
	Class 1	Class 2	Class 3	Class 4	
	n = 1,577 (39.3%)	n = 556 (13.8%)	n = 1,382 (34.4%)	n = 500 (12.5%)	
Sex, n (%)					
Male	749 (47.5)	240 (43.2)	401 (29.0)	315 (63.0)	< 0.001
Female	828 (52.5)	316 (56.8)	981 (71.0)	185 (37.0)	
Age, years, M ± SD	48.8 ± 9.6	48.5 ± 9.4	47.6 ± 9.1	48.1 ± 9.1	0.007 [2]
Education years, M ± SD	5.8 ± 4.8	6.0 ± 4.6	5.2 ± 4.5	7.0 ± 5.2	< 0.001 [2]
Marital status, n (%)					
Single	168 (10.7)	38 (6.8)	185 (13.4)	34 (6.8)	< 0.001
Married	1,409 (89.3)	518 (93.2)	1,197 (86.6)	466 (93.2)	
Tobacco smoking, n (%)					
Never smoker	1,129 (71.6)	392 (70.5)	1,062 (76.8)	335 (67.0)	< 0.001
Current smoker	131 (8.3)	62 (11.2)	79 (5.7)	77 (15.4)	
Former smokers	110 (7.0)	29 (5.2)	61 (4.4)	33 (6.6)	
Passive smoker	207 (13.1)	73 (13.1)	180 (13.0)	55 (11.0)	
Diabetes, n (%)					
Yes	292 (18.5)	74 (13.3)	225 (16.3)	63 (12.6)	0.002
No	1,285 (81.5)	482 (86.7)	1,157 (83.7)	437 (87.4)	
Hypertension, n (%)					
Yes	363 (23.0)	111 (20.0)	283 (20.5)	88 (17.6)	0.048
No	1,214 (77.0)	445 (80.0)	1,099 (79.5)	412 (82.4)	
Thyroid disease, n (%)					
Yes	134 (8.5)	51 (9.2)	167 (12.1)	37 (7.4)	0.002
No	1,443 (91.5)	505 (90.8)	1,215 (87.9)	463 (92.6)	
BMI, kg/m^2, M ± SD	27.3 ± 4.8	26.0 ± 4.9	26.8 ± 5.1	26.7 ± 4.6	< 0.001 [2]
WC, cm, M ± SD	95.0 ± 11.3	91.2 ± 12.2	93.4 ± 12.5	93.3 ± 11.2	< 0.001 [2]
HC, cm, M ± SD	100.7 ± 9.5	98.6 ± 9.5	100.0 ± 9.9	99.9 ± 9.0	< 0.001 [2]
FFQ's energy intake, kcal, M ± SD	2,889.19 ± 980.29	2,831.15 ± 964.68	2,747.63 ± 894.04	3,053.70 ± 1,042.67	< 0.001 [2]

Table 4. Socio-demographic and clinical characteristics of the study population by latent class. [1] Chi-square test. [2] ANOVA. BMI, Body mass Index; WC, Waist circumference; HC, Hip circumference; WBC, White blood cells; RBC, Red blood cells; HGB, Hemoglobin; HCT, Hematocrit; PLT, Platelet count; LMR, Lymphocyte-to-monocyte ratio; MCV, Mean corpuscular volume; MCH, Mean corpuscular hemoglobin; MCHC, Mean corpuscular hemoglobin concentration; RDW, Red cell distribution width - Coefficient of variation; FFQ, Food frequency questionnaire.

Tobacco smoking category [1]	Latent classes (OR [95% CI], *P*)			
	Class 1 [2]	Class 2	Class 3	Class 4
Current smoker	Reference	**1.51 [1.05–2.16], 0.024**	1.03 [0.75–1.42], 0.838	**1.50 [1.08–2.10], 0.015**
Former smokers	Reference	0.93 [0.59–1.47], 0.767	1.07 [0.75–1.53], 0.675	0.80 [0.52–1.23], 0.320
Passive smoker	Reference	1.01 [0.75–1.35], 0.936	0.90 [0.72–1.13], 0.383	0.88 [0.63–1.23], 0.488

Table 5. Odds ratios for the association between tobacco smoking categories and latent class membership, adjusted for confounders. The models were adjusted for age, sex, education years, marital status, thyroid disease, diabetes, hypertension, BMI, WC, HC, and energy intake. [1] Never smoker category was set as the reference group. [2] Class 1 was set as the reference group.

pituitary–thyroid axis and reduced erythropoiesis, both of which contribute to thyroid dysfunction[24,25]. Generally, this complex interplay might represent the intricate relationship between anemia subtypes and NCDs, which was observed in our study.

Our study showed a significant association between current smoking and membership in Class 4, which is characterized by the coexistence of IDA and macrocytosis, in both univariable and adjusted analyses. This association might indicate the complex relationship between smoking and hematological abnormalities, with several potential underlying physiological mechanisms. Class 4 was distinguished by higher probabilities of exceeding normal ranges for MCV, MCH, MCHC, and HGB, aligning with findings from previous studies linking smoking to macrocytosis. In the study by Shakiba et al.[26], an investigation conducted within the RaNCD branch of the PERSIAN and showed a robust positive correlation between smoking and increased levels of

HCT, HGB, MCHC, MCH, MCV, and WBC counts. This positive association persisted even after adjusting for numerous confounders. Similarly, Pedersen et al.[10] analyzed data from over 100,000 subjects, suggesting that smoking contributes to modest but significant increases in RBC indices, further solidifying the relationship with showing that these increases occurred regardless of the presence of genetic predispositions associated with smoking, such as variations in the CHRNA3 gene. Additionally, Takahashi et al.[27] showed a significant positive correlation between smoking and MCV and MCH in male workers. Interestingly, RBC counts in smokers were either reduced or unchanged, a finding that closely aligns with our results. Similarly, Schmitt et al.[28] observed that smokers tend to show higher RBC indices (MCV, MCH, and MCHC) despite normal or decreased RBC counts. The link between smoking and macrocytosis was further corroborated by O'Reilly et al.[29], who identified smoking as a significant correlate of elevated MCV, even after adjusting for alcohol use, vitamin B12 and folate deficiency, and liver disease. Their analysis implied that smoking accounted for over half of the cases of macrocytosis in their study population. Our results further support these associations by showing that, while the RBC count in Class 4 remained normal or decreased, the RBC indices were elevated.

The above-mentioned finding is consistent with the notion that smoking is associated with qualitative abnormalities in erythropoiesis, potentially contributing to macrocytosis without a corresponding increase in RBC count. That is, while hypoxia typically stimulates erythropoietin production, studies indicate that serum erythropoietin levels are often decreased in smokers due to negative feedback mechanisms from elevated HGB and RBC counts; however, increased erythropoietin receptor expression in smokers may enhance erythropoiesis[30–32]. These abnormalities might be linked to impaired erythropoiesis or increased RBC turnover, as suggested by the mechanism of eryptosis (programmed RBC death) in smokers, associated with increased MCV and MCH levels despite the absence of significant pancytopenia[28,33]. Oxidative stress related to chronic smoking may exacerbate both IDA and macrocytosis. Chronic smoking generates reactive oxygen species (ROS) and toxic substances like cyanide and acetaldehyde that can damage bone marrow, alter erythrocyte membrane and disrupt RBC morphology, potentially contributing to an increase in MCV and MCH[34–36]. This is corroborated by Eisenga et al.[37], which have reported a dose-dependent relationship between smoking and elevated MCV by analyzing both questionnaire-based smoking data and urinary cotinine levels in large European cohorts. Additionally, smoking-induced inflammation may impair iron homeostasis by increasing hepcidin levels, which sequester iron in macrophages and could limit its bioavailability for erythropoiesis, potentially exacerbating IDA in smokers. This inflammatory pathway might also interfere with the proper synthesis of HGB, relating to the anemic state observed in Class 4. Furthermore, although less likely[28,29], smoking-related alterations in folate metabolism, possibly due to increased hepatic enzyme activity induced by smoking-related hydrocarbons, could further contribute to megaloblastic anemia. The decrease in serum folate and potential alterations in vitamin B12 metabolism linked to cyanide might explain the mixed anemia phenotype observed in Class 4, as these deficiencies are known to disrupt RBC production and maturation[38–41].

In our study, we observed a significant association between current smoking status and membership in class 2 compared to class 1. Additionally, in the univariable analysis, a positive smoking history, particularly being a current smoker, was more frequent in individuals classified as class 2 compared to those in classes 1 and 3. This suggests that smoking is more prevalent among individuals with more pronounced clinical or laboratory features of beta-thalassemia minor. The observed association might be partially explained by the physiological and hematologic impacts of smoking on individuals with beta-thalassemia minor. Evidence shows that smokers with beta-thalassemia minor exhibit significantly elevated levels of carboxyhemoglobin (HbCO) compared to non-smokers with the condition and healthy smokers. This increase in HbCO reflects both the inhalation of carbon monoxide from smoking and endogenous CO production due to heightened heme catabolism in beta-thalassemia. Elevated HbCO levels can impair oxygen transport and exacerbate hypoxia, potentially contributing to more severe clinical manifestations, as observed in class 2 members[42]. Furthermore, findings from another study suggest that smoking induces significant alterations in hematologic parameters, including increases in RBC count, HGB levels, HCT, MCV, and MCH in smokers with beta-thalassemia minor. These changes might be compensatory responses to the combined effects of chronic hypoxia and oxidative stress caused by smoking. These hematologic shifts, in turn, could contribute to the increased likelihood of current smokers being classified in class 2, which might represent individuals with more marked disease characteristics[43]. Nonetheless, although the direct link between smoking and beta-thalassemia minor is less commonly discussed, the evidence may suggest the potential for smoking to exacerbate the clinical features of beta-thalassemia through mechanisms such as impaired oxygenation, oxidative stress, and hematologic adaptations. These findings align with our results, indicating a need for further research to explore the role of smoking in modulating the clinical presentation of beta-thalassemia minor.

From clinical point of view, significant association between current smoking status and membership in both class 2 and class 4, with respective prevalence of 13.8% and 12.5% may carry important clinical relevance for the diagnosis, management, and prevention of anemia in adults, particularly among smokers. The identification of smoking as a significant factor in specific anemia subtypes may direct the need for clinicians to incorporate smoking history into their diagnostic and treatment protocols. In cases of beta-thalassemia minor, smoking may compound the clinical manifestations of the disease by exacerbating hypoxia and oxidative stress. Additionally, in mixed anemia, smoking-related macrocytosis may complicate the anemia profile and necessitate more comprehensive diagnostic algorithms. This is suggested that idiopathic macrocytosis in smokers could mask early signs of myelodysplastic syndromes (MDS) or other underlying conditions[44]; therefore, physicians should remain vigilant for signs of idiopathic macrocytosis or elevated MCV in smokers, considering smoking as a differential factor in cases where other causes are excluded. Notably, since these effects are largely reversible, integrating smoking cessation counseling into routine anemia care could mitigate adverse hematologic effects, improve treatment outcomes, and reduce the overall burden of anemia in this population. Further research is

warranted to explore the mechanistic pathways through which smoking influences anemia subtypes, particularly in populations with coexisting hematologic conditions.

A notable strength of this study is the use of LCA, which enabled a nuanced exploration of blood index profiles, as well as effectively managing potential collinearity among variables through its pattern-based approach to subgroup identification rather than direct parameter estimation[45,46]. It also further supported by examining inter-variable correlations prior to model inclusion. Unlike traditional approaches that focus on single blood markers, LCA allowed us to simultaneously evaluate multiple indices, uncovering heterogeneity in anemia subtypes that might otherwise be overlooked.

However, the study had several limitations. The cross-sectional design precludes causal inferences, making it unclear whether smoking directly contributes to the observed anemia subtypes or whether other confounding factors, such as dietary habits or coexisting health conditions, play a role. Longitudinal studies are needed to assess causality and the impact of smoking cessation on blood indices over time. The smoking status variable, as provided by the cohort team, was limited to a validated 4-status NIH classification and lacked granular data on smoking intensity (cigarettes/day) and duration, which prevented dose-response analysis. Additionally, while we observed significant associations between smoking and specific anemia profiles, the reliance on self-reported smoking data introduces the possibility of reporting bias, potentially underestimating the true prevalence of smoking. Misclassification is another potential limitation. For instance, the presence of various anemia types might lead to diagnostic overlap, complicating the interpretation of anemia subtypes identified through LCA. In addition, lack of confirmatory diagnostic tests for beta-thalassemia might have resulted in under- or overestimation of this condition within the study population. While sex-specific modeling would be methodologically preferable due to sex disparities in hematological indices, the low prevalence of female smokers could potentially result in sparse data that precluded stable stratification; hence, we limited our analysis to only adjusted for sex. The study's findings are restricted to adults aged 35 years and older, further limiting the generalizability of our conclusions to younger populations. Moreover, other unmeasured factors, such as medication use, and lifestyle factors like opium consumption, alcohol intake, and detailed dietary habits, were not accounted for, introducing the possibility of residual confounding. These variables could influence both smoking behavior and anemia profiles. Finally, the absence of similar studies for comparison restricts the contextual interpretation of our findings, highlighting the need for further research to validate and expand upon our results.

Conclusion

This study found a significant association between smoking status and distinct latent classes of CBC profiles. By employing LCA, we uncovered nuanced patterns of blood indices and showed that current smoking is associated with specific anemia subtypes, particularly beta-thalassemia minor and mixed anemia. These findings suggested the importance of incorporating smoking history into the diagnosis of anemia and the potential benefits of smoking cessation, as a modifiable risk factor, in mitigating hematological abnormalities and anemia management.

Data availability

Data can be inquired from the corresponding author.

Received: 5 February 2025; Accepted: 21 July 2025
Published online: 28 July 2025

References

1. GBD 2016 Risk Factors Collaborators. Global, regional, and National comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet* **390** (10100), 1345–1422. https://doi.org/10.1016/S0140-6736(17)32366-8 (2017).
2. WHO global report on trends in. prevalence of tobacco use 2000–2025, third edition. Geneva: World Health Organization; (2019).
3. Goodchild, M., Nargis, N. & Tursan d'Espaignet, E. Global economic cost of smoking-attributable diseases. *Tob. Control.* **27** (1), 58–64. https://doi.org/10.1136/tobaccocontrol-2016-053305 (2018).
4. Moosazadeh, M., Ziaaddini, H., Mirzazadeh, A., Ashrafi-Asgarabad, A. & Haghdoost, A. A. Meta-analysis of smoking prevalence in Iran. *Addict. Health*. **5** (3–4), 140–153 (2013).
5. Abdollahpour, I., Mansournia, M. A., Salimi, Y. & Nedjat, S. Lifetime prevalence and correlates of smoking behavior in Iranian adults' population; a cross-sectional study. *BMC Public. Health*. **19** (1), 1056. https://doi.org/10.1186/s12889-019-7358-0 (2019).
6. Varmaghani, M. et al. Prevalence of smoking among Iranian adults: findings of the National steps survey 2016. *Arch. Iran. Med.* **23** (6), 369–377. https://doi.org/10.34172/aim.2020.29 (2020).
7. Inal, B. et al. Effects of smoking on healthy young men's hematologic parameters. *North. Clin. Istanb*. **1** (1), 19–25. https://doi.org/10.14744/nci.2014.39974 (2014).
8. Higuchi, T. et al. Current cigarette smoking is a reversible cause of elevated white blood cell count: Cross-sectional and longitudinal studies. *Prev. Med. Rep.* **4**, 417–422. https://doi.org/10.1016/j.pmedr.2016.08.009 (2016).
9. Aldosari, K. H. et al. The influence and impact of smoking on red blood cell morphology and buccal microflora: A case-control study. *J. Clin. Lab. Anal.* **34** (6), e23212. https://doi.org/10.1002/jcla.23212 (2020).
10. Pedersen, K. M. et al. Smoking and increased white and red blood cells. *Arterioscler. Thromb. Vasc Biol.* **39** (5), 965–977. https://doi.org/10.1161/ATVBAHA.118.312338 (2019).
11. Khan, M. I., Bukhari, M. H., Akhtar, M. S. & Brar, S. Effect of smoking on red blood cells count, hemoglobin concentration and red cell indices. *PJMHS* **8** (2), 361–364 (2014).
12. Kung, C. M., Wang, H. L. & Tseng, Z. L. Cigarette smoking exacerbates health problems in young men. *Clin. Invest. Med.* **31** (3), E138–E149. https://doi.org/10.25011/cim.v31i3.3471 (2008).
13. Yarnell, J. W. et al. Some long term effects of smoking on the haemostatic system: a report from the caerphilly and speedwell collaborative surveys. *J. Clin. Pathol.* **40** (8), 909–913. https://doi.org/10.1136/jcp.40.8.909 (1987).

14. Tell, G. S., Grimm, R. H. Jr, Vellar, O. D. & Theodorsen, L. The relationship of white cell count, platelet count, and hematocrit to cigarette smoking in adolescents: the Oslo youth study. *Circulation* **72** (5), 971–974. https://doi.org/10.1161/01.cir.72.5.971 (1985).
15. Malenica, M. et al. Effect of cigarette smoking on haematological parameters in healthy population. *Med. Arch.* **71** (2), 132–136. https://doi.org/10.5455/medarh.2017.71.132-136 (2017).
16. Zamani, M. et al. Prevalence and determinants of anemia among Iranian population aged ≥ 35 years: A PERSIAN cohort-based cross-sectional study. *PLoS One*. **17** (2), e0263795. https://doi.org/10.1371/journal.pone.0263795 (2022).
17. Postchi, H. et al. *Prospective Epidemiologic Research Studies in Iran, PERSIAN Cohort, Study Protocol* (editor. Tehran. Mir Mah publication, 2016). Edition one.
18. Poustchi, H. et al. Prospective epidemiological research studies in Iran (the PERSIAN cohort Study): Rationale, Objectives, and Design. *Am. J. Epidemiol.* **187** (4), 647–655. https://doi.org/10.1093/aje/kwx314 (2018).
19. Nejatizadeh, A. et al. Cohort profile: Bandar Kong prospective study of chronic non-communicable diseases. *PLoS One*. **17** (5), e0265388. https://doi.org/10.1371/journal.pone.0265388 (2022).
20. Baker, W. F. Jr Iron deficiency in pregnancy, obstetrics, and gynecology. *Hematol. Oncol. Clin. North. Am.* **14** (5), 1061–1077. https://doi.org/10.1016/s0889-8588(05)70171-4 (2000).
21. Aydogan, G. et al. Causes of hypochromic microcytic anemia in children and evaluation of laboratory parameters in the differentiation. *J. Pediatr. Hematol. Oncol.* **41** (4), e221–e223. https://doi.org/10.1097/MPH.0000000000001382 (2019).
22. Iolascon, A., De Falco, L. & Beaumont, C. Molecular basis of inherited microcytic anemia due to defects in iron acquisition or Heme synthesis. *Haematologica* **94** (3), 395–408. https://doi.org/10.3324/haematol.13619 (2009).
23. Barbieri, J. et al. Anemia in patients with type 2 diabetes mellitus. *Anemia* **2015**, 354737. https://doi.org/10.1155/2015/354737 (2015).
24. Zimmermann, M. B. & Köhrle, J. The impact of iron and selenium deficiencies on iodine and thyroid metabolism: biochemistry and relevance to public health. *Thyroid* **12** (10), 867–878. https://doi.org/10.1089/105072502761016494 (2002).
25. Maggio, M. et al. The role of the multiple hormonal dysregulation in the onset of anemia of aging: focus on testosterone, IGF-1, and thyroid hormones. *Int. J. Endocrinol.* **2015**, 292574. https://doi.org/10.1155/2015/292574 (2015).
26. Shakiba, E. et al. Tobacco smoking and blood parameters in the Kurdish population of Iran. *BMC Cardiovasc. Disord.* **23** (1), 401. https://doi.org/10.1186/s12872-023-03433-2 (2023).
27. Takahashi, M. et al. Idiopathic Macrocytosis in Smokers. 人間ドック (Ningen Dock). 27(4):689 – 96. https://doi.org/10.11320/ningendock.27.689 (2012).
28. Schmitt, M. et al. Smoking is associated with increased eryptosis, suicidal erythrocyte death, in a large population-based cohort. *Sci. Rep.* **14** (1), 3024. https://doi.org/10.1038/s41598-024-53258-y (2024).
29. O'Reilly, M. A. et al. Smoking as an independent risk factor for macrocytosis in middle-aged adults: a population-based observational study. *Am. J. Hematol.* **90** (9), E196–E197. https://doi.org/10.1002/ajh.24085 (2015).
30. EisengaMF et al. Active smoking and hematocrit and fasting Circulating erythropoietin concentrations in the general population. *Mayo Clin. Proc.* **93** (3), 337–343. https://doi.org/10.1016/j.mayocp.2018.01.005 (2018).
31. Singh, V., Tanwar, A. S., Hungund, A. S., Hungund, S. A. & Nagaraja, C. Comparison of serum erythropoietin levels in smokers and nonsmokers with periodontitis: A biochemical study. *J. Indian Soc. Periodontol.* **20** (3), 249–253. https://doi.org/10.4103/0972-124X.181242 (2016).
32. Izzotti, A. et al. Genomic and transcriptional alterations in mouse fetus liver after transplacental exposure to cigarette smoke. *FASEB J.* **17** (9), 1127–1129. https://doi.org/10.1096/fj.02-0967fje (2003).
33. Blumenthal, I. Carbon monoxide poisoning. *J. R Soc. Med.* **94** (6), 270–272. https://doi.org/10.1177/014107680109400604 (2001).
34. Yanbaeva, D. G., Dentener, M. A., Creutzberg, E. C., Wesseling, G. & Wouters, E. F. Systemic effects of smoking. *Chest* **131** (5), 1557–1566. https://doi.org/10.1378/chest.06-2179 (2007).
35. Padmavathi, P., Reddy, V. D., Kavitha, G., Paramahamsa, M. & Varadacharyulu, N. Chronic cigarette smoking alters erythrocyte membrane lipid composition and properties in male human volunteers. *Nitric Oxide*. **23** (3), 181–186. https://doi.org/10.1016/j.niox.2010.05.287 (2010).
36. Buckley, C. M., Harrington, J. M., O'Shea, S., Perry, I. J. & Cahill, M. R. Cigarette smoking is an under recognised cause of macrocytosis. *Blood* **122**(21):4660. https://doi.org/10.1182/blood.V122.21.4660.4660
37. Eisenga, M. F. et al. Active smoking and macrocytosis in the general population: two population-based cohort studies. *Am. J. Hematol.* **94** (2), E45–E48. https://doi.org/10.1002/ajh.25346 (2019).
38. Mannino, D. M., Mulinare, J., Ford, E. S. & Schwartz, J. Tobacco smoke exposure and decreased serum and red blood cell folate levels: data from the third National health and nutrition examination survey. *Nicotine Tob. Res.* **5** (3), 357–362. https://doi.org/10.1080/1462220031000094330 (2003).
39. Northrop-Clewes, C. A. & Thurnham, D. I. Monitoring micronutrients in cigarette smokers. *Clin. Chim. Acta*. **377** (1–2), 14–38. https://doi.org/10.1016/j.cca.2006.08.028 (2007).
40. Helman, N. & Rubenstein, L. S. The effects of age, sex, and smoking on erythrocytes and leukocytes. *Am. J. Clin. Pathol.* **63** (1), 35–44. https://doi.org/10.1093/ajcp/63.3.35 (1975).
41. Yokoyama, A. et al. Macrocytosis, a new predictor for esophageal squamous cell carcinoma in Japanese alcoholic men. *Carcinogenesis* **24** (11), 1773–1778. https://doi.org/10.1093/carcin/bgg142 (2003).
42. Palarie, D. L., Preda, M. & Palarie, I. Effect of smoking on concentration of carboxyhemoglobin for persons with beta-thalassemia minor. *Asian J. Chem.* **20** (2), 1303–1310 (2008).
43. Tarazi, I. S., Sirdah, M. M., El Jeadi, H. & Al Haddad, R. M. Does cigarette smoking affect the diagnostic reliability of hemoglobin alpha 2 delta 2 (HbA2)? *J. Clin. Lab. Anal.* **22** (2), 119–122. https://doi.org/10.1002/jcla.20228 (2008).
44. Leifert, J. A. Anaemia and cigarette smoking. *Int. J. Lab. Hematol.* **30** (3), 177–184. https://doi.org/10.1111/j.1751-553X.2008.01067.x (2008).
45. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36** (1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x (2013).
46. Sinha, P., Calfee, C. S. & Delucchi, K. L. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Crit. Care Med.* **49** (1), e63–e79. https://doi.org/10.1097/CCM.0000000000004710 (2021).
47. Nejatizadeh, A. et al. Cohort profile: Bandar Kong prospective study of chronic non-communicable diseases. PLoS One. ;17(5):e0265388. doi: 10.1371/journal.pone.0265388. Erratum in: PLoS One. 2024;19(12):e0315430. (2022). https://doi.org/10.1371/journal.pone.0315430

Acknowledgements

We would also like to thank Hormozgan University of Medical Sciences for supporting this research.

Author contributions

AM, MSH, and AGH: providing the main idea of study and methodology, final analysis, developing the idea and revising the final manuscript, ZM, LJ, ASH, and MSH: developing the idea and revising the final manuscript, contributed to data analysis and revising the final manuscript. PSH, LJ, and ASH revised the final manuscript. All authors approved the final version of the manuscript that was submitted.

Funding
Hormozgan University of Medical Sciences.

Declarations

Ethics approval and consent to participate
The authors fully adhered to ethical standards regarding issues such as plagiarism, informed consent, misconduct, data fabrication and falsification, duplicate publication or submission, and redundancy. This research was conducted following the ethical guidelines outlined in the Declaration of Helsinki and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. Additionally, the study received approval from the Research Ethics Committee of Hormozgan University of Medical Sciences (IR.HUMS.REC.1402.363). This study utilized baseline data from the Kong Cohort Study, and informed consent was obtained from all participants. The cohort profile of the Bandar Kong study has been published previously[47].

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.