

Construction of a healthy lifestyle index using Italian national survey data

Manuela Scioni^{1*}, Chiara Baldan¹, Alessia Ghirardo¹, and Giovanna Boccuzzo¹

Submission History

Submitted:	31/01/2025
Accepted:	23/09/2025
Published:	01/12/2025

¹Department of Statistical Sciences, University of Padua, Padua, Italy, 35121

Abstract

Introduction

Lifestyle choices encompassing dietary habits, physical activity levels, alcohol consumption, and tobacco use have been consistently shown to significantly impact individual health outcomes and overall well-being.

Objectives

This study proposes a novel composite index to measure the adoption of healthy lifestyles among the Italian population aged 18 years and over.

Methods

The Healthy Lifestyle Composite Index (HLCI) is constructed by aggregating four key dimensions: diet, physical activity, alcohol consumption, and tobacco use. The dimensions are structured as ordinal variables derived from the comprehensive Aspects of Daily Life (AVQ) multipurpose household survey conducted annually by the Italian National Statistical Institute (ISTAT). A formative approach is employed, involving defining the dimensions, determining weights through the Analytic Hierarchy Process based on expert evaluations, and specifying an aggregation procedure using a weighted Borda rule.

Results

The resulting HLCI provides a score from 0 to 100, with higher values indicating healthier lifestyles. Analysis of the HLCI and its dimensions using the 2022 AVQ data (n=32,600) reveals an average score of 61.77, with substantial variation across demographic groups. Descriptive analysis of the HLCI revealed significantly higher scores for females compared to males, driven by better performance in the alcohol and tobacco consumption dimensions. An inverted U-shaped trend emerged for age, with the youngest (18-19 years) and oldest (75+) groups exhibiting higher HLCI values. Educational level was positively associated with HLCI, with graduates scoring highest, excelling in physical activity. Geographically, the North-East region had the highest HLCI. Quantile regression on the first decile highlighted at-risk profiles with extremely low HLCI values, such as 35-44-year-old separated/divorced males with middle school education residing in South Italy.

Conclusion

Constructed using reliable data from an annually updated national survey, the HLCI allows for monitoring lifestyle dynamics across different demographic groups and geographic regions. The findings highlight specific segments of the population that may benefit from targeted interventions promoting a healthier lifestyle.

5 Bullet points:

- Proposal of a new Healthy Lifestyle Composite Index (HLCI) to measure adoption of healthy lifestyles in the Italian population.
- HLCI aggregates four dimensions: diet, physical activity, alcohol consumption, and tobacco use, using data from an annual national survey.
- HLCI employs a formative approach with expert-weighted dimensions and a weighted Borda aggregation rule to calculate the 0-100 score.
- Analysis of 2022 survey data shows average HLCI of 61.77 with variations across demographics like age, marital status, and educational level.
- Monitoring healthy lifestyle dynamics using regularly updated institutional data to target promotion efforts effectively.

Keywords

healthy lifestyle; lifestyle factors; composite index; health promotion

*Corresponding Author:

Email Address: manuela.scioni@unipd.it (Manuela Scioni)

Introduction

According to the 1948 definition by the World Health Organization (WHO), health can be defined as 'a state of complete physical, mental, and social well-being, and not merely the absence of disease or infirmity' [1]. The WHO further elucidates that an individual's health status is influenced by a variety of factors encompassing environmental, economic, and social conditions, as well as the individual's characteristics and behaviours [2].

While individuals may have limited control over many determinants of their health, they retain agency over their behaviours and habits. Factors such as smoking, dietary choices, alcohol consumption, and physical activity represent aspects of an individual's lifestyle that can be consciously managed to promote overall health. Notably, the link between these four behaviours and health outcomes has garnered significant attention within the scientific community [3–5]. Numerous studies have investigated the relationship between these behaviours and the risk of mortality [6, 7], various types of cancer [8, 9], and several diseases, including cardiovascular diseases [10] and hypertension [11].

The purpose of this paper is to propose a composite index that measures the healthy lifestyle of the Italian population aged 18 years and over, and then it aims to analyse this population according to the adopted lifestyles.

In literature many composite indices are designed to evaluate healthy lifestyle patterns with the objective of establishing the association between the index and a specific disease or death. The manner in which these four factors are measured to construct such indices varies significantly across studies, depending on the available data. For instance, Patrão et al. [12] developed a lifestyle index based on health behaviours in the ELSA-Brazil study. This index considered eating behaviours (assessed through fruit and vegetable consumption); physical activity; smoking; and alcohol consumption. Each behaviour was classified as healthy or unhealthy, and participants were assigned a score ranging from 0 (no healthy behaviours) to 5 (all healthy behaviours present). Spencer et al. [13] used data from the Western Australian Abdominal Aortic Aneurysm (AAA) Screening Programme to create a lifestyle score predicting survival in healthy men with a mean age of 71 years and no cardiovascular history. This score was based on eight behaviours, with one point allocated for each healthy behaviour. The selected factors included smoking history, physical activity levels, alcohol consumption, BMI, and dietary choices (in particular, the frequency of eating fish and meat, salt usage, and milk preference). Scores ranged from 0 to 8, with higher scores indicating a healthier lifestyle. Adjibade et al. [14] introduced a Healthy Lifestyle Index (HLI) using data from the NutriNet-Santé study. Their aim was to evaluate the association between the HLI and the risk of incident depressive symptoms. This index combined factors such as smoking status, alcohol consumption, physical activity, diet quality, and BMI. Participants were assigned one point for each healthy lifestyle component they adhered to, resulting in an HLI ranging from 0 to 5 points. McKenzie et al. [8] generated a Healthy Lifestyle Index Score (HLIS) with five components: smoking status, physical activity, alcohol consumption, diet, and BMI, to assess associations with all cancer, and the

alcohol-, tobacco-, obesity- and reproductive-related cancer groupings. The study included 391,608 men and women from the multinational European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. The HLIS was constructed from these five factors assessed at baseline by assigning scores of 0 to 4 to categories of each factor, with higher values indicating healthier behaviours. The resulting healthy lifestyle index ranged from 0 to 20. An adaptation of this index was developed by van der Meer et al. [15] to investigate the relationship between the score and the risk of renal cell cancer, using data from The Netherlands Cohort Study on Diet and Cancer. Fukunaga et al. [16] used data derived from a cluster randomised controlled trial to investigate the association between a Healthy Lifestyle Index (HLI) and hypertension among community-dwelling adults in Sri Lanka. To construct the HLI, five modifiable lifestyle factors were employed: BMI, physical activity, smoking status, alcohol consumption, and fruit and vegetable consumption. Each lifestyle factor was dichotomised into a low-risk group (indicating adherence to a healthy lifestyle) or a high-risk group (indicating non-adherence to a healthy lifestyle). The low-risk group received a score of 1, while the high-risk group received a score of 0. These scores were then summed to calculate a composite score ranging from 0 to 5.

In these studies, the behaviours associated with a healthy lifestyle were measured in different ways. Most studies applied thresholds or scoring systems to classify or rate individuals' adherence to healthy behaviours. The overall lifestyle score was then typically calculated by summing these individual behaviour scores, although the specific methodology varied across studies.

The methodology for constructing the Healthy Lifestyle Composite Index (HLCI) we propose follows the traditional formative approach, which involves: (1) defining the dimensions, (2) determining the weights assigned to each dimension through expert elicitation, and (3) specifying the aggregation procedure for the dimensions. Data from the comprehensive Aspects of Daily Life (AVQ) multipurpose household survey conducted annually by the Italian National Statistical Institute (ISTAT) were used to define the four dimensions included in the HLCI - namely diet, physical activity, alcohol consumption, and tobacco use - which are structured as ordinal variables; we then apply weights to these dimensions, reflecting our belief that they are not equally important, and employ a weighted Borda aggregation method.

The advantage of drawing from institutional sources for the definition of composite indices is considerable: a composite index is primarily useful for comparative purposes, both spatially and temporally, and it is therefore essential that it can be computed based on reliable and consistently updated sources. To our knowledge, no lifestyle indices based on ISTAT data have been proposed to date, and we believe that our proposal can contribute to enriching the framework of fair and sustainable well-being indices in Italy. Based on data available annually, the index is easily updatable and can therefore constitute a useful tool for evaluating the behavioral dynamics of Italians regarding their lifestyles.

ISTAT, thanks to the BES (Well-being and sustainability) project, produces a set of 152 indicators every year, with the aim of evaluating the progress of society not only from an economic perspective but also from a social and environmental

point of view [17]. However, it is important to note that this set of indicators is not an attempt to create a composite indicator, nor does it specifically focus on healthy lifestyle. Rather, it provides a broad overview of various aspects of well-being and sustainability. Moreover, the BES project utilises data from multiple sources beyond the AVQ survey, including other ISTAT surveys.

Data and methods

The Aspects of Daily Life (AVQ) survey is an annual survey conducted by the Italian National Institute of Statistics (ISTAT) since 1993. For this study, we utilised data from the 2022 survey, involving approximately 24,000 households and 50,000 individuals, selected through a two-stage stratified sampling to ensure representativeness at both national and regional levels. The AVQ is a cross-sectional survey, meaning that each year a new sample of households is interviewed, limiting the analysis to a single reference year. The survey covers various aspects of daily life, including health conditions, lifestyle habits, use of health services, family and social relationships, education, work, leisure time activities, social and political participation, safety, and subjective well-being. AVQ data are made publicly available by ISTAT in anonymised form, in compliance with privacy regulations. ISTAT follows rigorous ethical protocols in data collection and dissemination, conforming to international standards for official statistical surveys¹.

In this article, we put forward a proposal for constructing a Healthy Lifestyle Composite Index (HLCI) measuring the adoption of healthy lifestyles among the Italian population aged 18 and over using individual data. The approach taken is formative [18], wherein the HLCI is the representation of different aspects of lifestyles which, when taken together, form an individual's lifestyle profile.

The construction of composite indices at the micro level, i.e., for individuals, poses an additional methodological challenge, given that the dimensions and the elementary indicators being calculated for individuals can be both quantitative and, more often, qualitative ordinal in nature. This implies the need for more articulated approaches to normalising and aggregating them compared to standard approaches (for example: weighted arithmetic mean).

The steps addressed in constructing the index are therefore [19]: defining the dimensions that comprise the index and the elementary indicators within each dimension, identifying the appropriate approach to defining weights and the method of aggregating the dimensions that compose the index.

Defining the dimensions

Four dimensions have been included in the HLCI – namely diet, physical activity, alcohol consumption, and tobacco use. In defining the dimensions, we undertook a process that aimed to optimally match the available data from the AVQ survey with guidelines and existing literature. These dimensions were

structured as ordinal variables, following the ordinal nature of several elementary indicators.

Diet

A proper dietary regime is closely linked to a health condition and the pursuit of healthy longevity. An inappropriate diet, coupled with a sedentary lifestyle, can lead to overweight, a condition characterised by excessive body fat accumulation which can significantly impact the health of a population. Moreover, the composition and quality of diet, not merely its impact on body weight, can play a protective role against the onset of cardiovascular, neurodegenerative, and neoplastic conditions [20].

To define the diet dimension, we considered the available information from the AVQ survey regarding various categories of food and beverages (Table 1). The consumption frequency for various food groups was recorded through the following possible answers: more than once a day, once a day, several times a week, less than once a week, never. The consumption frequency for non-alcoholic beverages, particularly water and carbonated drinks, is instead assessed through the question: 'In what quantity do you habitually consume the following beverages?' with categorical responses ranging from no consumption to more than a litre a day. Starting from these responses, the daily consumption frequency of foods is calculated through an approximation (see Table S1 in Supplementary Appendix 1) and, if necessary, the overall consumption frequency for a food category is calculated by summing the consumption frequency of its sub-food items. Then, for each food/beverage category a threshold value above/below which the amount of food is considered at risk is identified from the dietary guidelines provided by the Council for Research in Agricultural and Analysis of Agricultural Economy (CREA), last revised in 2018 [21, 22]. Finally, the diet dimension is created from the sum of individual risky behaviours among the 10 food categories and non-alcoholic beverages considered, defining the following categories: 0, 1, 2, 3, 4 and 5+.

Physical activity

Physical activity is defined as any bodily movement produced by skeletal muscles that requires energy expenditure [23]. Physical activity confers benefits for the following health outcomes: improved all-cause mortality, cardiovascular disease mortality, incident hypertension, incident site-specific cancers, incident type-2 diabetes, mental health (reduced symptoms of anxiety and depression); cognitive health, and sleep; measures of adiposity may also improve [23].

Regarding the dimension related to physical or sports activity, the assessment is based on the frequency of engagement in at least one such activity. In the AVQ survey, each individual is asked the following questions:

- Do you engage in one or more sports activities regularly or occasionally during your leisure time? (Five or more times a week, Three or four times a week, Twice a week, Once a week, Two or three times a month, Once a month, Occasionally during the year)

¹For further details on the survey methodology and content, please refer to ISTAT's methodological notes and the official website (<https://www.istat.it/en/microdata/multipurpose-survey-on-households-aspects-of-daily-life/>)

Table 1: Variables considered in the Diet Dimension and associated risk thresholds

Food category	Foods	Risk thresholds
Fruits and Vegetables	Leafy vegetables cooked and raw (spinach, salads, chicory, cabbage, broccoli) Tomatoes (excluding preserves), eggplants, bell peppers, fennel, zucchinis, artichokes, carrots, squashes, cauliflower, peas, and other fresh legumes.	Fruits and vegetables $\leq$ 2/day
Milk and Dairy Products	Fruits Milk Cheeses, dairy products	Milk and dairy products >10 /week
Red Meat and Meat Products	Sheep meats Beef meats Pork meats (excluding cured meats) Cured meats	Red Meat and meat products > 2 /week
White Meat	Chicken, turkey, rabbit meats	White Meat >3 /week
Fish	Fish	Fish < 2 /week
Eggs	Eggs	Theoretical threshold: Eggs > 4 /week Survey-based threshold: Eggs ≥ 7 /week
Sweets	Sweets (filled cakes, snacks, ice creams, etc.)	Sweets >1 /week
Salty snacks	Salty snacks (chips, popcorn, pretzels, olives)	Salty snacks > 1 /week
Water	Mineral water	Water < 1 litre/day
Carbonated Beverages	Non-alcoholic carbonated beverages (excluding mineral water)	Carbonated beverages ≥ 1 glass/day

- Do you engage in any physical activities during your leisure time, such as walking for at least 2 kilometres, swimming, cycling, or similar activities at least once a year? (Once or more a week, Once or more a month, Less frequently)

Since the health benefits derived from engaging in moderate physical activities are comparable to those obtained from participating in any sports [24], it was decided to assign equal importance to both sports and physical activities.

For individuals who declare participating in at least one of the above-mentioned activities, the following frequency of engagement in physical or sports activities are derived: 5+/week, 3-4/week, 1-2/week, 2-3/month, 1/month or less often (see Table S2 in Supplementary Appendix 1). Individuals who do not engage in any sport or physical activity are classified in the 'Less often' category. In conclusion, the Physical Activity dimension is defined as an ordinal variable with six categories. While it is true that sporting activities and physical activities are quantified differently, this distinction originates from the structure of the AVQ questionnaire itself. Despite this differentiation, we find the resulting categorisation to be coherent: intensive sports practitioners are placed in distinct categories, while regular physical activity (once or more weekly) is equated with weekly sports participation, and less frequent physical activity is categorised alongside sporadic sports engagement.

Alcohol consumption

The consumption of alcoholic products is associated with the development of numerous non-communicable chronic diseases (NCDs) and it can create dependency. Established effects include increased high-density lipoprotein cholesterol

and antithrombotic activity [25]. Overall, the evidence does not indicate any safe level of alcohol consumption [25].

In the context of alcohol consumption assessment, the focus lies on quantifying the daily intake of alcoholic units. An alcoholic unit is defined as 12 grams of ethanol, equivalent to the alcohol content typically found in:

- A 125 ml glass of wine with an average alcohol content of 12%.
- A 330 ml can of beer with an average alcohol content of 4.5%.
- A 80 ml aperitif with an average alcohol content of 38%.
- A 40 ml serving of high-alcohol spirits, containing 40% alcohol.

In order to distinguish individuals having daily alcohol intake from those who consume it sporadically, the AVQ survey poses the question: 'In what quantity do you typically consume the following alcoholic beverages?' If an individual reports daily consumption within any of the specified alcoholic beverage categories, s/he is further asked: 'How many glasses of wine, beer, alcoholic aperitifs, liqueurs, or spirits do you usually consume per day?' For those who reported drinking alcohol more often than sporadically but less than daily, a value of 2 drinks per week was assigned. To derive the total daily alcoholic unit count, we aggregated the quantities of wine, beer, aperitifs, and spirits consumed daily by the respondents. The dimension related to alcohol consumption refers to the total number of alcoholic units consumed per day, grouped in the following categories: 0, (0-1], (1-2], 3+. In this notation, parentheses indicate exclusion of the boundary value, while square brackets indicate inclusion. For example, (0-1] represents more than 0 and up to and including 1 unit.

Tobacco consumption

Tobacco smoking represents a major global public health concern, with the WHO estimating it is responsible for 8 million deaths annually [26]. Cigarette smoking is the primary risk factor for cancers (strongly associated with lung, oral cavity, throat, oesophagus, pancreas, colon, bladder, prostate, kidney, breast, ovaries, and certain types of leukaemia) and non-neoplastic respiratory diseases, such as chronic obstructive pulmonary disease (COPD) and it is also a significant cardiovascular risk factor (hypertension, stroke, and heart attack). Moreover, smoking negatively impacts the reproductive systems of both men and women, reducing fertility [27].

In the context of assessing smoking behaviour, we focus on the number of cigarettes smoked per day. In the ISTAT questionnaire, each individual is asked the following question: 'Do you currently smoke?' For those who report current smoking, additional questions are posed. These include: 'How frequently do you smoke?' with a focus on daily or occasional smoking, and 'What is your primary smoking choice?' with options such as cigarettes, cigars, or pipes for respondents to choose from. Only individuals who predominantly smoke cigarettes are further asked: 'On average, how many cigarettes do you smoke per day?'. The classification of individuals within this dimension is as follows:

- Non-smokers (including never-smokers and former smokers) are assigned a daily cigarette consumption of zero. For those who primarily smoke cigarettes, the exact average number of cigarettes reported by the respondent is recorded.
- For individuals who primarily smoke pipes or cigars, daily product consumption is not available. To prevent underestimation of the tobacco-related risk for these individuals, which accounts for approximately 3.1% of smokers, a proxy measure was employed. The average number of cigarettes consumed daily by those who smoke cigarettes regularly (10.75) and the average number for those who report occasional cigarette use (2.96) were calculated. These values were attributed to those who report predominantly smoking cigars or pipes, either on a daily or occasional basis, respectively.

The total number of cigarettes was grouped into 6 categories: 0, 1-4, 5-9, 10-12, 13-15 and 16+.

Determining the weights

Many methodological approaches are proposed in the literature for calculating weights. Decancq and Lugo [28] distinguish three classes of approaches to set the weights in the case of multidimensional indices: data-driven, normative and hybrid. Data-driven weights (based on statistical analyses) are a function of the distribution of the dimensions analysed and are not based, at least explicitly, on any value judgment. Normative weights, on the other hand, are set from value judgments (e.g. expert opinion and equal or arbitrary). Finally, hybrid approaches' weights are both data-driven and based on some form of evaluation. In this context of analysis and given the need for medical expertise, we believe that the choice of weights should be entrusted to experts in the field.

Table 2: Saaty's scale for pairwise comparison

Importance	Definition
1	A and B are equally important
3	A is slightly more important than B
5	A is moderately more important than B
7	A is significantly more important than B
9	A is absolutely more important than B

The methodology employed in this study for deriving weights is the Analytic Hierarchy Process (AHP). Its originator, Thomas Saaty, an American mathematician, authored the first book on AHP in 1987 [29]. Saaty's process allows incorporating experts' judgments and evaluations in a structured way through a pairwise comparison process facilitating the evaluation of the relative importance between the dimensions and simplifying the task for experts. Moreover, it includes a consistency index to assess the logical consistency of the judgments expressed by experts during the comparisons. For each pair of dimensions, each expert is asked to determine which one is more important and by what degree on a scale from 1 to 9 (Table 2). The questions posed to the experts are listed in Section S2 in Supplementary Appendix 1.

Four experts have been involved in the evaluation process. The experts are PhD medical doctors, three of whom are specialised in Epidemiology and Biostatistics, and one has an extensive clinical experience. All of them are experts in Public Health, healthcare service organisation and actively engaged in research related to lifestyle, health promotion, and the assessment of community preventive interventions.

For each expert who performed the pairwise comparison of the dimensions, an individual AHP preference matrix $A_i, i = 1, \dots, 4$ was constructed. The corresponding inconsistency index was computed as a transformation of the Perron-Frobenius eigenvalue, here denoted as $\lambda_{\max}$:

$$CI(A_i) = \frac{\lambda_{\max} - k}{k - 1}$$

where k is the matrix order, in this case equal to 4. To make this quantification of the inconsistency more intelligible, the CI was divided by the so-called Random Index RI_k , which is the average CI calculated on a large number of randomly generated matrices of order k . This division yields the Consistency Ratio $CR = CI/RI_k$ [29]. By assessing consistency for each individual matrix, it is possible to detect and address inconsistent judgments, by excluding them from further analysis. As proposed by Saaty [29], a threshold of 10% for the inconsistency level was established. Consequently, one expert whose responses exceeded this limit (18%) was excluded from the subsequent analyses to maintain the overall consistency of the judgments. The resulting sample size was of three experts, which can be considered an appropriate number, since the problem is rather simple and involves only one level of comparisons [30].

Assuming that all the experts have the same expertise level, individual evaluations have been aggregated in a new matrix representing the overall preferences of the expert panel through the geometric mean, which satisfies the Arrow's four conditions [31]. The weights attributed to

the 4 dimensions correspond to the components of the principal eigenvector associated with the largest eigenvalue of the matrix constructed from these aggregated individual evaluations.

Aggregation process

To aggregate the four dimensions composing the Index, a weighted Borda Rule was employed. The Borda rule is appropriate for working with ordinal variables. Originally, Borda scores were used to determine rankings in voting situations (see, e.g., [32–35]). Nowadays, they are applied in a variety of fields such as discrete multi-criteria analysis, composite indices, artificial intelligence, database queries, and Internet multiple search engines [36]. In the context of the Borda rule, and all scoring methods in general, the intensity of preference is measured by the scores given according to the rank positions.

Each dimension has a varying number of levels ordered from the healthiest behaviour to the riskiest (lower is better), as follows:

- Diet, measured in daily risky behaviours: 0, 1, 2, 3, 4, 5+;
- Physical activity, measured by frequency: 5+ times a week, 3-4 times a week, 1-2 times a week, 2-3 times a month, 1 time a month, less frequently;
- Alcohol consumption, measured by the number of alcoholic units consumed daily: 0, (0-1], (1-2], 3+;
- Tobacco consumption, measured by the number of cigarettes smoked daily: 0, 1-4, 5-9, 10-12, 13-15, 16+.

For each dimension, individuals can be ranked into different levels. The number of levels varies by dimension. The scores are obtained by assigning a value of 0 to the last rank, 1 to the second-to-last rank, and so on, with the top rank receiving a value equal to the number of levels minus one. Therefore, for the diet, physical activity, and tobacco consumption dimensions, the possible scores are 0, 1, 2, 3, 4, and 5, with a score of 0 representing the least healthy behaviour and a score of 5 corresponding to the healthiest behaviour. To handle the alcohol dimension, which has 4 categories, we adopted a proportional scoring approach, in order to preserve the overall scoring scale. With the minimum value always being 0 and the maximum value always being 5, for the intermediate positions, we distributed the remaining points equidistantly: 3.34 points for the second position and 1.67 points for the third.

The scores, calculated for each individual, are subsequently multiplied by the relative weight assigned to the corresponding dimension and computed using Saaty's approach [29]. After performing this procedure for each of the dimensions, the four weighted scores are summed to obtain the overall score for each individual. So, the score associated to the i -th individual, s_i , can be expressed by the following formula:

$$s_i = \sum_{j=1}^4 w_j \times s_{ij}$$

where:

- w_j is the weight associated to the j -th dimension
- s_{ij} is the score obtained by the i -th individual with respect the j -th dimension.

Finally, to normalise the obtained score, a Min-Max normalisation approach was applied, using the theoretical minimum and maximum values. The normalised score, s'_i , was scaled to the interval [0, 100], according the following:

$$s'_i = \frac{s_i - \min(s)}{\max(s) - \min(s)}$$

where 0 corresponds to the absolute worst lifestyle and 100 to the best.

The Borda Rule offers a simple and intuitive approach, whose computational efficiency becomes particularly advantageous when dealing with large datasets, as in our case. The Borda rule is often criticised for its violation of the independence of irrelevant alternatives (IIA), as the addition or removal of a non-winning candidate can affect the outcome, even if it shouldn't logically influence the ranking of the remaining candidates [32]. However, the impact of IIA violations can appear less significant in larger datasets. As the number of units increases, the likelihood of ties (where multiple alternatives receive the same number of points) also increases, and the relative ranking between alternatives becomes less sensitive to small changes in individual rankings [32].

Analysing the indicator

A descriptive analysis of the level of HLCI and its four dimensions was conducted for several demographic and socio-economic characteristics. The descriptive analysis of the four dimensions is important, because the same level of HLCI may be the result of different combinations of the four dimensions. Since the distributions do not conform to the assumptions of normality and homoscedasticity, the non-parametric Wilcoxon test is employed for comparing two groups (e.g. males vs females), while the non-parametric Kruskal-Wallis test is selected in the case of three or more groups (e.g. educational level). Conducting a large number of tests places our analysis within the context of multiple testing, in which the probability that at least one rejected hypothesis is a false positive is much higher than the nominal significance level. This probability is called the family wise error rate (FWER). To control it, we apply the Wilcoxon rank-sum test with continuity correction using the Bonferroni adjustment method.

After that, we proceeded with a multivariate analysis to determine which demographic and socio-economic characteristics significantly affect HLCI. Given that we are interested in capturing the effect of the demographic and socio-economic variables on the individuals who have a particularly low index, we adopted a quantile regression [37], in which we model the first decile of the index distribution.

The quantile regression model is:

$$y_i = x'_i \beta_p + \varepsilon_i$$

where:

- y_i indicates the HLCI for the i -th respondent;

- $x'_i = \{x_{i1}, \dots, x_{is}\}$ indicates the values of the set of S explanatory variables for the i -th respondent;
- $\beta_p = \{\beta_{p0}, \beta_{p1}, \dots, \beta_{ps}\}$ indicates the $(S+1)$ regression parameters for p -th quantile (the 10th quantile in our case).

As a consequence, the p -th quantile is given by:

$$p = P(y_i \leq x'_i \beta_p \mid x_i)$$

and the p -th quantile of the conditional distribution of y_i given x_i is:

$$Q_y(p) = x'_i \beta_p.$$

The quantile regression can be applied to the data regardless of the distribution of the variable of interest, thus making unnecessary any hypothesis about it.

Results

Sample characteristics

The initial sample size of the AVQ 2022 survey is 42,022, from which 6,106 individuals under the age of 18 are subtracted. The sample consisting only of adults has a sample size of 35,916 individuals, representative of the Italian adult population, which amounts to approximately 47.5 million individuals. In the data quality check, about 3,316 units are identified to have at least one missing value for one of the key variables for defining the dimensions described above. These units are removed, resulting in a final sample size of 32,600 units. Table 3 shows the main descriptive characteristics of the sample. All the analyses are performed using R 4.3.0 software [38].

As it can be observed in Table 3, the sample is fairly evenly split between males and females, with a slight majority of females (52.25%). The largest age group is 45–54 years (18.99%), followed by those 75 and older (15.16%) and 65–74 (14.63%). Married individuals form the largest group (48.72%), followed by singles (31.59%). As regards educational level, high school diploma holders form the largest group (39.54%), followed by those with middle school diplomas (27.41%). The sample shows a diverse geographical representation across Italy, with the South having the largest share (26.88%).

The estimated population size after applying sampling weights can be found in Table S3 in the Appendix. As can be observed by comparing the distribution of characteristics in the sample and in the population (last two columns of Table S3), after removing observations with missing data, the sample's characteristics remain largely consistent with those of the population, with all differences falling within a $\pm 5\%$ range.

Descriptive analysis of HLCI and its dimensions

The overall structure of the indicator, inclusive of weights, is depicted in Figure 1. The dimension that receives the highest weight is related to tobacco consumption (0.387), followed by diet at 0.249, physical activity at 0.189 and alcohol consumption at 0.174.

Table 3: Characteristics of the sample ($n = 32,600$) by demographic and social traits

Variable	Levels	%
Gender ^a	Female	52.25
	Male	47.75
Age	18–19	2.35
	20–24	5.92
	25–34	11.24
	35–44	13.37
	45–54	18.99
	55–59	9.75
	60–64	8.60
	65–74	14.63
Marital Status	75+	15.16
	Single	31.59
	Married	48.72
	Separated, Divorced	9.12
	Widowed	9.61
Educational Level	Missing	0.95
	None/Elementary School	14.13
	Middle School Diploma	27.41
	High School Diploma	39.54
	Graduate/Postgraduate	17.42
Geographical Region	Missing	1.50
	North-east	21.32
	North-west	22.94
	Central	19.08
	South	26.88
	Islands	9.69
	Missing	0.09

^aWith regard to the categories of the variable gender, it should be noted that the questionnaire only provides for the response options male and female, and no other response options are provided. There are no non-responses/missing data for this variable.

The average score for the Healthy Lifestyle Composite Index (HLCI) is 61.77 (Table 4), exhibiting a left-skewed distribution (Figure 2). The analysis reveals that 61 individuals (0.19%) exhibit the minimum value of the composite index (0), indicating they adopt the least healthy behaviour across all four dimensions, while 9 individuals (0.03%) achieve the maximum value, demonstrating the healthiest score in all four dimensions (100).

Turning to the descriptive analysis of HLCI dimension scores, it is important to note that these scores do not involve weights, which are only used to aggregate the dimensions into the final index. Among these unweighted dimensions, the one with the highest average score is related to tobacco consumption (mean = 88.37), followed by alcohol consumption (mean = 72.03). These high scores are attributable to the relatively low number of individuals in the sample who smoke and consume alcohol daily (80.5% of individuals do not smoke, 82.0% do not consume alcohol or consume it very slightly). The lower scores in the physical activity (36.20) and diet (32.60) dimensions indicate areas where there is more room

Figure 1: Hierarchical structure of the Health Lifestyle Composite Index (HLCI) based on formative approach

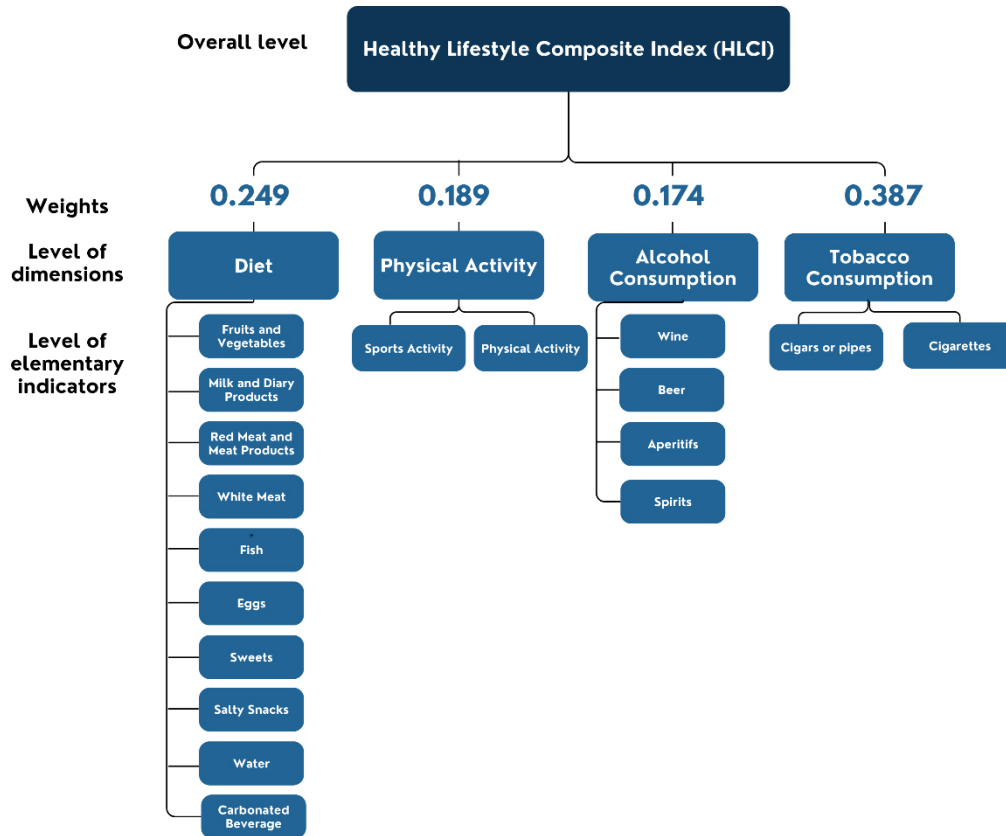


Table 4: Descriptive statistics of HLCI and its dimensions (higher is better)

	Mean	SD	Q1	Median	Q3
HLCI	61.77	15.90	55.06	65.28	72.25
Diet	36.20	26.28	20.0	40.0	60.0
Physical Activity	32.60	32.35	0.0	40.0	60.0
Alcohol Consumption	72.03	30.14	70.0	70.0	100.0
Tobacco Consumption	88.37	26.36	100.0	100.0	100.0

for improvement in the population's health behaviours, and at the same point potentially less awareness of their importance.

In the formative approach, it is important to consider the correlations between dimensions. These should not be too high to avoid considering the same concept in several dimensions and thus overestimating it. To this aim, the Goodman and Kruskal's Gamma index, which measures the covariation between two ordinal variables, was calculated (Table 5). Due to the large sample size in this study, even weak correlations emerged as statistically significant, highlighting the high statistical power of the analysis rather than necessarily indicating strong or practically meaningful relationships between variables. Covariations between diet and alcohol consumption, diet and physical activity, and tobacco consumption and physical activity can be considered as negligible. Covariations between diet and tobacco consumption, and alcohol consumption and physical activity, although higher, are weak. A moderate Gamma covariation is between tobacco consumption and alcohol consumption (0.232), indicating a positive monotonic relationship between the two variables. This is largely due to the high proportion

of people who neither smoke nor drink. On the contrary, an inverse relationship is observed between alcohol consumption and physical activity. This means that high levels of physical activity are accompanied by low alcohol consumption, and vice versa. Interestingly, the same inverse relationship is not observed between physical activity and smoking habits.

HLCI and individual characteristics

Table 6 illustrates how the HLCI and its dimensions are distributed among individuals according to certain demographic and socioeconomic characteristics. The results of the Wilcoxon rank-sum test with continuity correction using the Bonferroni adjustment method are provided in the Supplementary Appendix 1 from Table S4 to Table S23.

HLCI is significantly higher for females compared to males (65.54 vs. 59.95, $p < 0.001$). Except physical activity, all the dimensions and in particular alcohol consumption exhibit higher scores for women, meaning lower consumption. The HLCI mean scores across age groups show a significant yet subtle U-shaped trend, with highest values in the 18-19

Figure 2: Distribution of the Healthy Lifestyle Composite Index (HLCI) scores

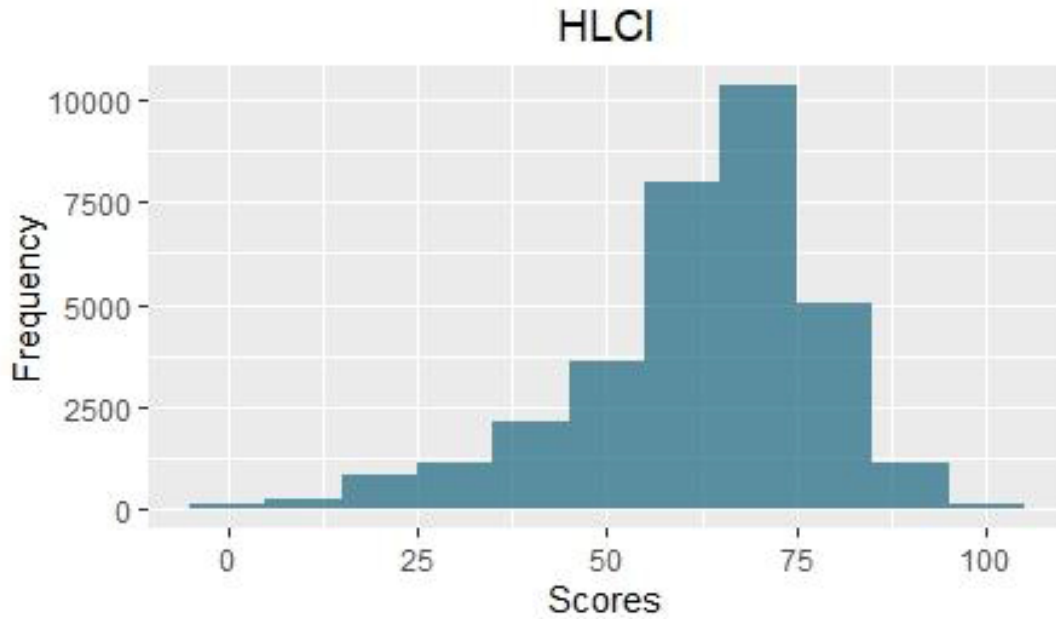


Table 5: Covariations between Dimensions measured by Goodman and Kruskal's Gamma index (***)p-value<0.001)

	Diet	Physical activity	Alcohol consumption	Tobacco consumption
Diet	—	0.060 ^a	0.057 ^a	0.159 ^b
Physical Activity	—	—	−0.145 ^b	0.063 ^a
Alcohol Consumption	—	—	—	0.232 ^c
Tobacco Consumption	—	—	—	—

Note: All cells $p < 0.001$ due to sample size. Following the rule of thumb from Rea and Parker [39], we consider the following ranges and interpretations of gamma: a. $0.00 < 0.10$ – Negligible, b. $0.10 < 0.20$ – Weak, c. $0.20 < 0.40$ – Moderate, d. $0.40 < 0.60$ – Relatively strong, e. $0.60 < 0.80$ – Strong, f. $0.80 < 1.00$ – Very strong.

age group, decreasing until the 35-44 age group, and then increasing again. This non-linear trend is also evident from the results of the pairwise comparison tests (Table S4). The diet dimension scores increase with age, demonstrating the adoption of a healthier diet as age advances, except for a slight decline in the oldest age group (75+). All age groups are significantly different from their subsequent groups, with the sole exceptions of the 18-19 age group not differing from the 20-24 group, and no significant differences are observed among the age groups above 60 years old (Table S5). Physical activity decreases with increasing age, with this decline being gradual, as demonstrated by the pairwise comparison tests not being significant for adjacent age groups (Table S6); the only age group characterised by markedly lower physical activity levels, significantly different from all others, is the 75+ age group. Alcohol consumption appears more prevalent among adults and the older adults, with a drop for the 75+ age group. Also in this case the pairwise comparison tests results to be not significant for some adjacent age groups. Smoking is concentrated in the middle-age groups (Table S7).

Single individuals score lower on the diet dimension, while married individuals do not exhibit significant differences compared to those who are separated or divorced (Table S10).

Widowed individuals have lower scores in physical activity, possibly due to age-related factors. Married individuals, analogously to separated or divorced (Table S13), demonstrate the lowest scores in alcohol consumption, while separated or divorced individuals and singles show the lowest scores in smoking (Table S12).

Individuals with a university degree or higher education exhibit the healthiest lifestyle, as reflected in both the composite index and individual dimensions, particularly in physical activity. Conversely, those with a middle school education show the lowest scores on the composite index and in the dimensions of diet, alcohol consumption, and tobacco use. Notably, the physical activity dimension demonstrates a clear and increasing trend corresponding to higher levels of education. Regarding the diet dimension, the differences are not significant between those with no education or elementary education and individuals with a university degree or postgraduate qualification, as well as between those with a middle school diploma and high school graduates (Table S15). The latter comparison is also not significant for alcohol consumption (Table S17).

Finally, regarding the geographical distribution, the South and Islands regions show the lowest scores, primarily due to

Table 6: Descriptive statistics of HLCI and its dimensions (higher is better) – Means (sd)

Variable	HLCI	Diet	Physical Activity	Alcohol Consumption	Tobacco Consumption
Gender					
Male	58.67 (17.29)	34.05 (25.89)	35.40 (32.67)	62.20 (33.11)	85.06 (29.83)
Female	64.59 (13.92)	38.17 (26.47)	30.03 (31.84)	81.02 (23.84)	91.40 (22.31)
	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001
Age					
18-19	64.73 (13.49)	23.45 (25.22)	45.86 (32.75)	83.27 (20.54)	92.44 (19.65)
20-24	63.43 (15.12)	26.58 (25.77)	47.14 (32.00)	78.64 (23.66)	88.57 (23.36)
25-34	61.60 (16.70)	31.22 (26.62)	42.10 (31.81)	74.15 (26.06)	85.49 (27.97)
35-44	59.93 (17.23)	33.25 (26.84)	35.17 (31.72)	72.41 (27.84)	84.00 (29.47)
45-54	60.69 (17.13)	35.21 (26.54)	34.41 (31.55)	71.06 (29.91)	85.74 (28.97)
55-59	61.10 (17.04)	38.06 (25.99)	32.85 (31.67)	70.71 (31.28)	85.89 (29.49)
60-64	61.58 (16.69)	40.34 (25.96)	32.07 (32.07)	67.83 (32.57)	87.48 (28.02)
65-74	62.77 (15.46)	41.72 (24.67)	30.16 (32.44)	68.47 (33.88)	90.35 (25.13)
75+	63.32 (11.34)	40.63 (24.53)	15.77 (27.11)	73.71 (31.57)	97.15 (13.51)
	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001
Marital Status					
Single	61.41 (16.85)	31.86 (26.71)	40.72 (32.54)	73.23 (27.91)	85.66 (27.92)
Married	61.92 (15.50)	38.06 (25.89)	30.33 (31.54)	69.76 (31.52)	89.81 (25.31)
Separated, Divorced	60.31 (17.63)	37.80 (26.17)	34.46 (32.22)	70.43 (30.69)	83.40 (31.03)
Widowed	63.57 (12.35)	39.96 (25.02)	15.76 (27.38)	80.40 (27.93)	94.96 (18.11)
	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001
Educational Level					
None/Elementary Sch	61.78 (13.11)	38.91 (24.56)	13.08 (25.52)	76.57 (31.22)	94.16 (19.98)
Middle Sch Diploma	58.73 (17.14)	34.45 (25.95)	26.76 (31.17)	71.48 (32.39)	84.77 (30.24)
High Sch Diploma	62.06 (16.03)	35.00 (26.52)	38.52 (32.18)	70.96 (29.25)	87.50 (26.66)
Graduate/Postgrad	65.91 (14.57)	39.77 (26.98)	44.53 (30.47)	71.10 (27.16)	91.38 (22.28)
	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001
Geographical Region					
North-east	63.32 (15.29)	33.73 (26.18)	42.27 (31.57)	70.62 (29.34)	89.90 (24.26)
North-west	62.09 (15.91)	35.01 (26.57)	37.17 (31.69)	70.00 (30.70)	88.73 (25.90)
Central	62.36 (16.29)	40.01 (26.07)	34.22 (32.56)	69.90 (31.00)	87.69 (26.69)
South	60.10 (15.81)	35.96 (25.86)	23.13 (30.57)	74.19 (29.91)	87.83 (27.32)
Islands	61.00 (16.28)	37.77 (26.40)	23.58 (30.81)	78.03 (28.26)	86.99 (28.38)
	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001	<i>p</i> -value<0.001

low values in physical activity, despite having high scores in alcohol consumption. Conversely, the North-East presents the highest healthy lifestyle index score, mainly thanks to its performance in physical activity and tobacco consumption, in which it exhibits significant differences compared to all other areas except the North-west (Tables S21 and S22).

Table 7 shows the results of the quantile regression, in which we model the first decile of the index distribution.

The quantile regression on the first decile shows interesting results that differ from what emerges in the descriptive analysis. We start from an HLCI value at the first decile of 54.95, corresponding to a female subject, aged 18-19, single, with at most a primary school education and residing in the north-east.

Firstly, the difference between males and females, already evident in Table 6, is even more pronounced here, with males having a first decile value 11 points lower than females. This means that, at the lowest levels of the indicator, the difference

between females and males is even greater. The difference between young people and adults is also much more marked, with first decile values more than 12 points lower between 25 and 34 years (even 16.21 between 35 and 44 years) compared to young people aged 18-19. The effect of higher education is also noteworthy: a graduate subject has a first decile of HLCI 9.48 points higher than those with at most a primary school education.

The 'worst' profile on the first decile of the distribution is given by a male subject, between 35 and 44 years old, separated or divorced, with a middle school education and residing in South Italy. His first decile HLCI is 30.77.

Stability and robustness of HLCI

Stability was tested by splitting the sample initially into two and then into three random subsamples of equal size. The procedures were repeated fifty times, and the mean of the

Table 7: Effects of the explanatory variables on the first decile of the Healthy Lifestyle Composite Index. Coefficients (std.err), t-values and p-values

Variables		Estimate (std.err)	t value	p-value
	Intercept	54.95 (0.88)	62.54	<0.0001
Gender (ref: Female)	Male	−11.11 (0.39)	−28.58	<0.0001
Age (ref: 18–19)	20–24	−6.73 (0.81)	−8.31	<0.0001
	25–34	−12.84 (0.87)	−14.68	<0.0001
	35–44	−16.21 (0.95)	−17.07	<0.0001
	45–54	−14.17 (0.96)	−14.72	<0.0001
	55–59	−13.57 (1.23)	−10.99	<0.0001
	60–64	−13.21 (1.15)	−11.50	<0.0001
	65–74	−8.95 (1.05)	−8.51	<0.0001
	75+	−2.22 (0.86)	−2.57	0.01015
Marital Status (ref: Single)	Married	3.73 (0.64)	5.79	<0.0001
	Separated, Divorced	−2.69 (0.91)	−2.96	0.00311
	Widowed	1.09 (0.82)	1.34	0.180
Education Level (ref: Elementary School or None)	Middle School Diploma	−2.46 (0.59)	−4.20	<0.0001
	High School Diploma	2.94 (0.64)	4.60	<0.0001
	Bachelor's/Postgraduate	9.48 (0.65)	14.60	<0.0001
Geographical Region (ref: Northeast)	Northwest	−1.02 (0.57)	−1.78	0.075
	Central	−2.76 (0.67)	−4.13	<0.0001
	South	−2.82 (0.62)	−4.53	<0.0001
	Islands	−1.74 (0.62)	−2.80	0.005

HLCI and its dimensions for each sample was calculated. The Wilcoxon signed-rank test (for two subsamples) and the Kruskal-Wallis test (for three subsamples) were employed to determine the frequency of significant differences among the means of the HLCI and its dimensions among the subsamples. Regarding HLCI stability, just three of the 50 trials based on two subsamples shows statistically significant differences between the HLCI means while, in the test with three subsamples, it was not observed that the means of HLCI differed significantly. As for the stability of the HLCI dimensions, the test with two subsamples provided the following results: one instance of significant differences in diet means (error rate = 2%), three cases of significant difference in physical activity and alcohol consumption (error rate = 6%), and two significant differences for tobacco consumption (error rate=4%). When the sample was divided into three subsamples, the means of the diet, alcohol consumption and tobacco consumption dimensions differ significantly in four cases (error rate = 8%) and for the physical activity in three cases (error rate=6%). In conclusion, we retain that such error rates are acceptable and that the validation results demonstrate good stability of the HLCI.

Application of the HLCI to 2023 data

To further illustrate the potential of the Healthy Lifestyle Composite Index (HLCI) for temporal monitoring, we applied the same methodology to the most recent wave of the Aspects of Daily Life (AVQ) survey, collected in 2023². The 2023 AVQ survey involved 35,681 individuals aged 18 years and

over (after excluding respondents with missing values on key variables, the final analytical sample consists of 32,785 individuals). The survey design remains unchanged compared to 2022, ensuring comparability of the results across years.

As expected, the results obtained for 2023 are very similar to those of 2022 (Table 8). This is consistent with the fact that the two survey waves are consecutive and no major changes in population behaviours are typically observed within such a short time frame. The main interest in applying the HLCI longitudinally lies in the possibility of comparing results over longer periods, which would allow the identification of relevant lifestyle trends and shifts. Such an in-depth temporal comparison, however, goes beyond the scope of the present paper, whose main focus is on the construction and illustration of the index.

Discussion

Research has consistently shown that lifestyle choices play a crucial role in determining an individual's vulnerability to various health conditions [3–6]. Consequently, people have the capacity to substantially improve their overall well-being by actively managing and modifying key factors that influence their health. However, it's not just an individual responsibility. The healthcare system and public health organisations also play a crucial role: they need to encourage and promote healthy lifestyles among the population. So, creating healthy habits is a task that involves both individuals and institutions. Health literature consistently identifies four key factors contributing to a healthy lifestyle: a balanced, nutrient-rich diet that supports overall health and disease prevention; regular physical activity

²Data were released during the review process of the paper.

Table 8: Descriptive statistics of HLCI and its dimensions (higher is better) for 2023

	Mean	SD	Q1	Median	Q3
HLCI	61.88	15.75	55.60	65.28	72.25
Diet	34.81	26.08	20.0	40.0	60.0
Physical Activity	33.18	32.29	0.0	40.0	60.0
Alcohol Consumption	72.65	29.59	70.0	70.0	100.0
Tobacco Consumption	88.99	25.75	100.0	100.0	100.0

that strengthens the body, improves cardiovascular health, and enhances mental well-being; moderate or no alcohol consumption to reduce the risk of liver disease and certain cancers; and avoiding or quitting smoking to protect lung health and lower the risk of heart disease and various cancers.

The primary aim of this article was to propose a composite index for measuring the adoption of healthy lifestyles (HLCI) among the Italian population. This index is derived from the aggregation of four dimensions: diet, physical activity, alcohol consumption, and tobacco use, using a weighted version of the Borda rule. It utilises data from the Aspect of Daily Life Survey, carried out every year by the National Institute of Statistics, which provides reliable and regularly updated information. Compared to existing proposals in the literature, the composite index we propose presents two distinctive features: the dimensions are structured as ordinal variables, aiming to represent the adoption of healthy habits in a gradual manner, and furthermore, it assumes that the dimensions do not carry equal weights.

We felt it was important to consult experts to assess whether the four dimensions were equally important. It turned out that different value judgements were made. Using experts in this area was undoubtedly the correct approach, as specific expertise was needed; however, adopting the Saaty method also required the experts to take a quantitative approach to assigning weights. Three of the experts we consulted work in public health and teach epidemiology, so they understood their task well. However, one expert had more experience in clinical practice than public health and found it difficult to make quantitative assessments. Therefore, the quantitative expertise of the experts needs to be taken into account, and eventually simpler approaches for assigning weights should be used.

The AVQ Survey is nationally and regionally representative. It collects a great deal of information on individuals, making it possible to measure the HLCI annually and compare results over time, even many years later, and between regions. It is also possible to compare different individual profiles and identify those at greater risk. However, as it is cross-sectional in nature, it generally does not allow the HLCI to be analysed as an explanatory factor for outcomes such as health status. Longitudinal surveys are necessary for this purpose. The SHARE survey on health and ageing in Europe is the data source that comes closest to achieving this [40], but its sample size is much smaller and does not guarantee adequate representativeness for specific policy purposes.

The analysis of HLCI reveals that the greatest differences are observed in relation to gender (females perform better than males), age (better lifestyles at younger and older ages) and educational qualifications (better lifestyles for people with

a bachelor or higher degree). At the lower levels of the distribution, these differences are even more pronounced.

Analysing the index is useful because it provides a general overview of the phenomenon, making it possible to visualise trends and differences over time and space. This highlights particularly relevant changes and categories of individuals at particular risk. However, jointly analysing its dimensions provides further important information for intervention purposes.

For example, we can see that the mean value for the dimension related to tobacco consumption is very high (88.4) due to the high proportion of non-smokers, particularly among females. In contrast, the mean values of the diet and physical activity dimensions are low (36.2 and 32.6, respectively). The results show that the population still underestimates the importance of adopting a healthy diet and reducing alcohol consumption, and that these issues deserve more attention from health policy. The harm caused by smoking is now well known, and literature shows that anti-smoking campaigns have been effective [41]. Our results demonstrate the importance of reducing the adoption of unhealthy lifestyles in other areas as well, particularly with regard to diet, which our experts have identified as the second most significant factor. Scientific literature increasingly demonstrates the detrimental effects of poor diet on health and premature mortality [42]. These findings are often reported in the media, but there are currently no food-related awareness campaigns in Italy. Several projects target school-age children and aim to reduce salt consumption in relation to cardiovascular risks, but there are no large-scale campaigns.

The population under consideration exhibits significantly different levels of HLCI by gender, age, and educational level. Therefore, it should be considered whether it is effective to adopt the same methodology and campaign for the entire population, or whether it would be better to take a more targeted approach. For example, men are at greater risk than women, and lower educational attainment leads to lower HLCI, so more targeted or intensive actions are needed for these groups. Education is therefore the main vehicle, and efforts should be made to reduce the knowledge and awareness gap between the more and less educated. While education is a proxy for socioeconomic level, which is not easily modifiable, the clear evidence of significant differences between groups with different levels of education highlights the need for diversified and targeted actions. A differential approach would lead to a more effective allocation of healthcare resources. It is necessary to encourage more physical activity among adults and women, and those with a low level of education; similarly, more attention needs to be paid to alcohol consumption among men. Understanding the categories at higher risk of unhealthy

behaviors, and the types of risky behaviors, allows for more effective targeting of interventions and resources.

Conclusion

This study proposes a novel Healthy Lifestyle Composite Index (HLCI) that provides a comprehensive measure of healthy behaviors among the Italian population. Constructed using reliable data from an annually updated national survey, the HLCI allows for monitoring lifestyle dynamics across different demographic groups and geographic regions. The findings highlight specific segments of the population that may benefit from targeted interventions promoting healthier diets, increased physical activity, and reduced alcohol and tobacco consumption. By incorporating this composite index into public health policies and programs, policymakers can make data-driven decisions to enhance the overall well-being of Italian citizens.

Acknowledgements

This research has received no external funding.

Statement of conflict of interests

None declared.

Ethics statement

Approval from an ethics committee was not required for this study, as it relied solely on pre-existing anonymised survey data collected by ISTAT under established ethical protocols.

Data availability statement

This study was conducted using data from the Aspects of Daily Life (AVQ) survey carried out by the Italian National Statistical Institute (ISTAT). Data are available at: <https://www.istat.it/en/microdata/multipurpose-survey-on-households-aspects-of-daily-life/>.

References

- World Health Organization. Constitution of the World Health Organization [Internet]. Geneva: World Health Organization; 1948. Available at: <https://www.who.int/about/governance/constitution>.
- World Health Organization. Social determinants of health [Internet]. Geneva: World Health Organization; 2021. Available at: <https://www.who.int/health-topics/social-determinants-of-health>.
- The US Burden of Disease Collaborators. The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. *JAMA*. 2018;319(14):1444–1472. <https://doi.org/10.1001/jama.2018.0158>
- Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. *Int J Epidemiol*. 2019;49(1): 113–130. <https://doi.org/10.1093/ije/dyz078>
- Li Y, Schoufour J, Wang DD, Dhana K, Pan A, Liu X, Song M, Liu G, Shin JH, Sun Q, Al-Shaar L, Wang M, Rimm EB, Hertzmark E, Stampfer MJ, Willett WC, Franco OH, Hu FB. Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study. *Bmj*. 2020; 368. <https://doi.org/10.1136/bmj.l6669>
- Van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. *Bmj*. 2008; 337. <https://doi.org/10.1136/bmj.a1440>
- Ding D, Rogers K, Ploeg HVD, Stamatakis E, Bauman AE. Traditional and emerging lifestyle risk behaviors and all-cause mortality in middle-aged and older adults: Evidence from a large population-based Australian cohort. *PLoS Med*. 2015; 12(12): e1001917. <https://doi.org/10.1371/journal.pmed.1001917>
- McKenzie F, Biessy C, Ferrari P, Freisling H, Rinaldi S, Chajès V, Dahm CC, Overvad K, Dossus L, Lagiou P, Trichopoulos D, Trichopoulou A, Bueno-de-Mesquita HB, May A, Peeters PH, Weiderpass E, Sanchez MJ, Navarro C, Ardanaz E, Ericson U, Wirfält E, Travis RC, Romieu I. Healthy lifestyle and risk of cancer in the European prospective investigation into cancer and nutrition cohort study. *Medicine*. 2016; 95(16): 2850 <https://doi.org/10.1097/MD.0000000000002850>
- Chen SLF, Braaten T, Borch KB, Ferrari P, Sandanger TM, Nøst TH. Combined lifestyle behaviors and the incidence of common cancer types in the Norwegian women and cancer study (NOWAC). *Clin Epidemiol*. 2021; 13: 721–734 <https://doi.org/10.2147/CLEP.S312864>
- Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, Michos ED, Miedema MD, Muñoz D, Smith SC, Virani SS, Williams KA, Yeboah J, Ziaeian B. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Jacc*. 2019; 74(10): 1376–1414. <https://doi.org/10.1161/CIR.0000000000000677>
- Sarki AM, Nduka CU, Stranges S, Kandala NB, Uthman OA. Prevalence of hypertension in low- and middle-income countries: A systematic review and meta-analysis. *Medicine*. 2015; 94(50): 1959. <https://doi.org/10.1097/MD.0000000000001959>
- Patrão AL, Almeida MdCC, Alvim S, Chor D, Aquino EML. Health behavior-related indicator of lifestyle: application in the ELSA-brasil study.

- Glob Health Promod. 2019; 26(4): 62–69. <https://doi.org/10.1177/1757975918763148>
13. Spencer C, Jamrozik K, Norman P, Lawrencebrown M. A simple lifestyle score predicts survival in healthy elderly men. *Prev Med*. 2005; 40(6): 712–717. <https://doi.org/10.1016/j.ypmed.2004.09.012>
14. Adjibade M, Lemogne C, Julia C, Hercberg S, Galan P, Assmann KE, Kesse-Guyot E. Prospective association between combined healthy lifestyles and risk of depressive symptoms in the French NutriNet-santé cohort. *J Affect Disord*. 2018; 238: 554–562. <https://doi.org/10.1016/j.jad.2018.05.038>
15. Meer R, Van De Pol J, Van Den Brandt PA, Schouten LJ. The association of healthy lifestyle index score and the risk of renal cell cancer in the Netherlands cohort study. *BMC Cancer*. 2023; 23(1): 156. <https://doi.org/10.1186/s12885-023-10627-6>
16. Fukunaga A, Inoue Y, Chandraratne N, Yamaguchi M, Kuwahara K, Indrawansa S, Gunawardena N, Mizoue T, Samarasinghe D. Healthy lifestyle index and its association with hypertension among community adults in Sri Lanka: A cross-sectional study. *PLoS One*. 2020; 15(1), 20226773. <https://doi.org/10.1371/journal.pone.0226773>
17. Istituto nazionale di Statistica. Benessere equo e sostenibile (BES) [Internet]. Roma: Istituto Nazionale di Statistica; 2023. Available at: [https://www.istat.it/it/benessere-e-sostenibilita/la-misurazione-del-benessere-\(bes\)](https://www.istat.it/it/benessere-e-sostenibilita/la-misurazione-del-benessere-(bes))
18. Diamantopoulos A, Winklhofer HM. Index construction with formative indicators: An alternative to scale development. *J Mark Res*. 2001; 38(2): 269–277. <https://doi.org/10.1509/jmkr.38.2.269.188>
19. Saisana M, Tarantola S. State-of-the-art report on current methodologies and practices for composite indicator development. Ispra: European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit. 2002. Available at: <https://op.europa.eu/en/publication-detail/-/publication/9253d939-b47b-4428-b792-619e6b6c8645>.
20. Gnerre P, Frediani R, Risicato R, Rondana M, Nardi R. Alimentazione, salute e longevità nei pazienti adulti. *Ital J Med* 2019; 7(7): 1–138. <https://doi.org/10.4081/itjm.q.2019.7>
21. Rossi L, Berni Canani S, Censi L, Gennaro L, Leclercq C, Scognamiglio U, Sette S, Ghiselli A. The 2018 revision of Italian dietary guidelines: development process, novelties, main recommendations, and policy implications. *Front Nutr* 2022; 9: 861526. <https://doi.org/10.3389/fnut.2022.861526>
22. Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria. Linee guida per una sana alimentazione. Revisione 2018 [Internet]. Roma: Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria; 2018. Available at: <https://www.crea.gov.it/web/alimenti-e-nutrizione/-/linee-guida-per-una-sana-alimentazione-2018>.
23. World Health Organization. WHO guidelines on physical activity and sedentary behavior [Internet]. Geneva: World Health Organization; 2020. Available at: <https://www.who.int/publications/i/item/9789240015128>.
24. Lordan G, Pakrashi D. Make time for physical activity or you may spend more time sick!. *Soc Indic Res* 2014; 119: 1379–1391. <https://doi.org/10.1007/s11205-013-0545-y>
25. Mukamal KJ, Rimm EB. Alcohol consumption: risks and benefits. *Curr Atheroscler Rep* 2008; 10(6): 536–543. <https://doi.org/10.1007/s11883-008-0083-2>
26. World Health Organization. Tobacco [Internet]. Geneva: World Health Organization; 2023. Available at: <https://www.who.int/news-room/fact-sheets/detail/tobacco>.
27. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2017; 392(10159): 1923–1994. [https://doi.org/10.1016/S0140-6736\(18\)32225-6](https://doi.org/10.1016/S0140-6736(18)32225-6)
28. Decancq K, Lugo MA. Weights in multidimensional indices of wellbeing: An overview. *Econom Rev* 2013, 32(1), 7–34. <https://doi.org/10.1080/07474938.2012.690641>
29. Saaty RW. The analytic hierarchy process—what it is and how it is used. *Math Model* 1987, 9(3–5): 161–176. [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8)
30. Saaty TL, Özdemir MS. How Many Judges Should There Be in a Group? *Ann Data Sci* 2014; 1: 359–368. <https://doi.org/10.1007/s40745-014-0026-4>
31. Saaty TL, Vargas LG. Possibility of group choice: pairwise comparisons and merging functions. *Soc Choice Welf* 2012; 38(3):481–496. <https://doi.org/10.1007/s00355-011-0541-6>
32. Brams SJ, Fishburn PC. Voting procedures. In: Arrow KJ, Sen AK, Suzumura K, editors. *Handbook of social choice and welfare*. Amsterdam: Elsevier; 2002. p 173–236. [https://doi.org/10.1016/S1574-0110\(02\)80008-X](https://doi.org/10.1016/S1574-0110(02)80008-X)
33. Young H. An axiomatization of Borda's rule. *J Econ Theory* 1974; 9(1): 43–52. [https://doi.org/10.1016/0022-0531\(74\)90073-8](https://doi.org/10.1016/0022-0531(74)90073-8)
34. Saari DG. Mathematical structure of voting paradoxes I: pairwise votes. *Econ Theory* 2000; 15(1): 1–53. <https://doi.org/10.1007/s001990050001>

35. Saari DG. Mathematical structure of voting paradoxes II: positional voting. *Econ Theory* 2000; 15(1): 55–102. <https://doi.org/10.1007/s001990050002>
36. Munda G. Choosing aggregation rules for composite indicators. *Soc Indic Res* 2012; 109(3): 337–354. <https://doi.org/10.1007/s11205-011-9911-9>
37. Koenker R. Quantile regression. Cambridge: Cambridge University Press; 2005. <https://doi.org/10.1257/jep.15.4.143>
38. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2024. <https://www.R-project.org/>.
39. Rea LM, Parker RA. Designing and Conducting Survey Research: A Comprehensive Guide. San Francisco: Jossey-Bass; 2014.
40. Börsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, Schaan B, Stuck S, Zuber S, SHARE Central Coordination Team. Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol* 2013; 42(4):992–1001. <https://doi.org/10.1093/ije/dyt088>
41. Akter S, Rahman MM, Rouyard T, Nsashiya RS, Nakamura R. A systematic review and network meta-analysis of population-level interventions to tackle smoking behaviour. *Nat Hum Behav* 2024; 8:2367–2391. <https://doi.org/10.1038/s41562-024-02002-7>
42. Tasnim Onni A, Balakrishna R, Perillo M, Amato M, Javadi Arjmand E, Thomassen LM, Lorenzini A, Fadnes LT. Umbrella review of systematic reviews and meta-analyses on consumption of different food groups and risk of all-cause mortality. *Adv Nutr* 2025; 16(4): 100393. <https://doi.org/10.1016/j.advnut.2025.100393>

Abbreviations

HLCI:	Healthy Lifestyle Composite Index
AVQ:	Aspects of Daily Life
ISTAT:	Italian National Statistical Institute
WHO:	World Health Organization
AAA:	Abdominal Aortic Aneurysm
HLI:	Healthy Lifestyle Index
BMI:	Body Mass Index
HLIS:	Healthy Lifestyle Index Score
EPIC:	European Prospective Investigation into Cancer
BES:	Well-being and sustainability
CREA:	Council for Research in Agricultural and Analysis of Agricultural Economy
COPD:	Chronic Obstructive Pulmonary Disease
AHP:	Analytic Hierarchy Process



Supplementary appendix 1

Table S1: Encoding of variables related to dietary habits

Elementary indicator	Response modes	Quantitative daily frequency
Food Categories	More than once a day	2
	Once a day	1
	Several times a week	2.5/7
	Less than once a week	3/28
	Never	0
Non-Alcoholic Drinks	Over one litre per day	1
	From 1/2 litre to 1 litre per day	0.5
	Less than 1/2 litre	0.25
	Less frequently	0
	Seasonally only	0
	I do not consume	0

Table S2: Encoding of variables related to physical activity

Elementary indicator	Response categories	Frequency
Physical Activity	Once or more a week	1-2/week
	Once or more a month	2-3/month
	Less frequently	0
Sports Activity	Five or more times a week	5+/week
	Three or four times a week	3-4/week
	Twice a week	1-2/week
	Once a week	1-2/week
	Two or three times a month	2-3/month
	Once a month	1/month
	Occasionally during the year	1/month



Table S3: Characteristics of the sample ($n = 32,600$) and estimated population size (N) after applying sampling weights (n =sample size after missing removal, N =population size before missing removal)

Variable	Levels	%	N (%)
Gender	Female	52.25	25,596,995 (51.72)
	Male	47.75	23,899,117 (48.28)
Age	18-19	2.35	1,184,828 (2.39)
	20-24	5.92	2,949,957 (5.96)
	25-34	11.24	6,194,782 (12.52)
	35-44	13.37	7,185,013 (14.52)
	45-54	18.99	9,303,917 (18.80)
	55-59	9.75	4,687,998 (9.47)
	60-64	8.60	4,065,038 (8.21)
	65-74	14.63	6,843,355 (13.83)
	75+	15.16	7,081,222 (14.31)
Marital Status	Single	31.59	15,906,715 (32.14)
	Married	48.72	23,867,856 (48.22)
	Separated, Divorced	9.12	4,652,503 (9.40)
	Widowed	9.61	4,612,386 (9.32)
	Missing	0.95	456,652 (0.92)
Educational Level	None/Elementary School	14.13	6,717,273 (13.57)
	Middle School Diploma	27.41	13,878,665 (28.04)
	High School Diploma	39.54	19,445,349 (39.29)
	Graduate/Postgraduate	17.42	8,739,261 (17.66)
	Missing	1.50	715,563 (1.45)
Geographical Region	North-east	21.32	9,696,826 (19.59)
	North-west	22.94	13,285,006 (26.84)
	Central	19.08	9,878,040 (19.96)
	South	26.88	11,207,238 (22.64)
	Islands	9.69	5,334,052 (10.78)
	Missing	0.09	94,949 (0.19)



Table S4: Pairwise comparisons of HLCI among age-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-74
20-24	1.00000	-	-	-	-	-	-	-
25-34	0.00419	0.03597	-	-	-	-	-	-
35-44	6.4e-09	2.9e-10	0.00135	-	-	-	-	-
45-54	1.5e-06	5.5e-07	0.43261	1.00000	-	-	-	-
55-59	0.00048	0.00266	1.00000	0.08958	1.00000	-	-	-
60-64	0.00306	0.03594	1.00000	0.00820	1.00000	1.00000	-	-
65-74	0.19436	1.00000	0.95643	2.1e-10	2.7e-06	0.07857	0.78464	-
75+	0.00014	0.00350	1.00000	7.7e-06	0.03447	1.00000	1.00000	0.23542

Table S5: Pairwise comparisons of the diet dimension among age-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-74
20-24	0.0914	-	-	-	-	-	-	-
25-34	1.0e-12	8.3e-09	-	-	-	-	-	-
35-44	< 2e-16	< 2e-16	0.0265	-	-	-	-	-
45-54	< 2e-16	< 2e-16	5.1e-12	0.0026	-	-	-	-
55-59	< 2e-16	< 2e-16	< 2e-16	2.7e-14	2.1e-05	-	-	-
60-64	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	0.0159	-	-
65-74	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	2.5e-09	0.7589	-
75+	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	6.3e-05	1.0000	1.0000

Table S6: Pairwise comparisons of the physical activity dimension among age-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-74
20-24	1.0000	-	-	-	-	-	-	-
25-34	0.0202	1.4e-07	-	-	-	-	-	-
35-44	< 2e-16	< 2e-16	< 2e-16	-	-	-	-	-
45-54	< 2e-16	< 2e-16	< 2e-16	1.0000	-	-	-	-
55-59	< 2e-16	< 2e-16	< 2e-16	0.0337	0.8300	-	-	-
60-64	< 2e-16	< 2e-16	< 2e-16	0.0014	0.0550	1.0000	-	-
65-74	< 2e-16	< 2e-16	< 2e-16	3.0e-13	8.1e-11	0.0053	0.2954	-
75+	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16

Table S7: Pairwise comparisons of the alcohol consumption dimension among age-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-74
20-24	0.00011	-	-	-	-	-	-	-
25-34	< 2e-16	1.2e-08	-	-	-	-	-	-
35-44	< 2e-16	3.1e-14	1.00000	-	-	-	-	-
45-54	< 2e-16	< 2e-16	0.06396	1.00000	-	-	-	-
55-59	< 2e-16	8.6e-12	1.00000	1.00000	1.00000	-	-	-
60-64	< 2e-16	< 2e-16	3.5e-07	0.00186	0.02833	0.03696	-	-
65-74	< 2e-16	< 2e-16	0.00380	1.00000	1.00000	1.00000	1.00000	-
75+	1.1e-07	1.00000	0.00046	1.7e-09	8.7e-14	5.4e-08	< 2e-16	3.0e-14

Table S8: Pairwise comparisons of the tobacco consumption dimension among age-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-74
20-24	0.00043	-	-	-	-	-	-	-
25-34	1.4e-07	1.00000	-	-	-	-	-	-
35-44	4.8e-10	0.00966	1.00000	-	-	-	-	-
45-54	5.0e-05	1.00000	1.00000	0.00155	-	-	-	-
55-59	0.00288	1.00000	0.13370	0.00028	1.00000	-	-	-
60-64	0.38260	0.42003	1.5e-05	1.5e-09	0.01424	1.00000	-	-
65-74	1.00000	4.7e-13	< 2e-16	< 2e-16	< 2e-16	6.3e-13	1.6e-05	-
75+	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16

Table S9: Pairwise comparisons of the HLCI among Marital Status-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	Single	Married	Separated, Divorced
Married	1.000	-	-
Separated, Divorced	0.060	0.025	-
Widowed	0.032	0.010	3.9e-05

Table S10: Pairwise comparisons of the diet dimension among Marital Status-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	Single	Married	Separated, Divorced
Married	< 2e-16	-	-
Separated, Divorced	< 2e-16	1.00000	-
Widowed	< 2e-16	0.00074	0.00282

Table S11: Pairwise comparisons of the physical activity dimension among Marital Status-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	Single	Married	Separated, Divorced
Married	< 2e-16	-	-
Separated, Divorced	< 2e-16	8.2e-10	-
Widowed	< 2e-16	< 2e-16	< 2e-16

Table S12: Pairwise comparisons of the alcohol consumption dimension among Marital Status-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	Single	Married	Separated, Divorced
Married	1.7e-08	-	-
Separated, Divorced	0.0076	1.0000	-
Widowed	< 2e-16	< 2e-16	< 2e-16

Table S13: Pairwise comparisons of the tobacco consumption dimension among Marital Status-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	Single	Married	Separated, Divorced
Married	<2e-16	-	-
Separated, Divorced	0.35	<2e-16	-
Widowed	<2e-16	<2e-16	<2e-16

Table S14: Pairwise comparisons of the HLCl among educational level-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	None/Elementary School	Middle School Diploma	High School Diploma
Middle School Diploma	1.4e-12	-	-
High School Diploma	2.5e-06	< 2e-16	-
Graduate/Post Graduate	< 2e-16	< 2e-16	< 2e-16

Table S15: Pairwise comparisons of the diet dimension among educational level-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	None/Elementary School	Middle School Diploma	High School Diploma
Middle School Diploma	<2e-16	-	-
High School Diploma	<2e-16	1	-
Graduate/Post Graduate	1	<2e-16	<2e-16

Table S16: Pairwise comparisons of the physical activity dimension among educational level-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	None/Elementary School	Middle School Diploma	High School Diploma
Middle School Diploma	<2e-16	-	-
High School Diploma	<2e-16	<2e-16	-
Graduate/Post Graduate	<2e-16	<2e-16	<2e-16

Table S17: Pairwise comparisons of the alcohol consumption dimension among educational level-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	None/Elementary School	Middle School Diploma	High School Diploma
Middle School Diploma	< 2e-16	-	-
High School Diploma	< 2e-16	7.8e-09	-
Graduate/Post Graduate	< 2e-16	5.5e-11	0.27

P value adjustment method: bonferroni

Table S18: Pairwise comparisons of the tobacco consumption dimension among educational level-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	None/Elementary School	Middle School Diploma	High School Diploma
Middle School Diploma	< 2e-16	-	-
High School Diploma	< 2e-16	2.8e-05	-
Graduate/Post Graduate	< 2e-16	< 2e-16	< 2e-16

Table S19: Pairwise comparisons of the HLCl among geographical areas-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	North-east	North-west	Central	South
North-west	0.00010	-	-	-
Central	0.05635	1.00000	-	-
South	< 2e-16	< 2e-16	< 2e-16	-
Islands	4.7e-10	0.02156	0.00051	0.00160

Table S20: Pairwise comparisons of the diet dimension among geographical areas-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	North-east	North-west	Central	South
North-west	0.0822	-	-	-
Central	< 2e-16	< 2e-16	-	-
South	6.7e-07	0.0831	< 2e-16	-
Islands	7.7e-12	3.1e-06	0.0011	0.0088

Table S21: Pairwise comparisons of the physical activity dimension among geographical areas-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	North-east	North-west	Central	South
North-west	< 2e-16	-	-	-
Central	< 2e-16	4.7e-06	-	-
South	< 2e-16	< 2e-16	< 2e-16	-
Islands	< 2e-16	< 2e-16	< 2e-16	1

Table S22: Pairwise comparisons of the alcohol consumption dimension among geographical areas-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	North-east	North-west	Central	South
North-west	1	-	-	-
Central	1	1	-	-
South	< 2e-16	< 2e-16	< 2e-16	-
Islands	< 2e-16	< 2e-16	< 2e-16	8.2e-11

Table S23: Pairwise comparisons of the tobacco consumption dimension among geographical areas-groups using Wilcoxon rank sum test with continuity correction (p-values, adjustment method: Bonferroni)

	North-east	North-west	Central	South
North-west	0.437	-	-	-
Central	9.8e-05	0.119	-	-
South	0.038	1.000	0.785	-
Islands	0.010	0.833	1.000	1.000



Section S2 Pairwise comparison questions for experts

For each pair of factors considered (diet, physical or sports activity, tobacco consumption, and alcohol consumption), we will ask you which of the two factors you consider more important in pursuing a healthy lifestyle and how much more important this factor is compared to the other on a scale from 1 to 9.

Importance Definition

- 1 A and B are equally important
- 3 A is slightly more important than B
- 5 A is moderately more important than B
- 7 A is strongly more important than B
- 9 A is extremely more important than B
- 1) What do you consider more important between a proper diet and moderate alcohol consumption in pursuing a healthy lifestyle?
- 1.1) To what extent do you consider factor A more important than factor B?
- 2) What do you consider more important between a proper diet and not smoking in pursuing a healthy lifestyle?
- 2.1) To what extent do you consider factor A more important than factor B?

- 3) What do you consider more important between a proper diet and engaging in physical or sports activity in pursuing a healthy lifestyle?
- 3.1) To what extent do you consider factor A more important than factor B?
- 4) What do you consider more important between moderate alcohol consumption and not smoking in pursuing a healthy lifestyle?
- 4.1) To what extent do you consider factor A more important than factor B?
- 5) What do you consider more important between moderate alcohol consumption and engaging in physical or sports activity in pursuing a healthy lifestyle?
- 5.1) To what extent do you consider factor A more important than factor B?
- 6) What do you consider more important between not smoking and engaging in physical or sports activity in pursuing a healthy lifestyle?
- 6.1) To what extent do you consider factor A more important than factor B?
- 7) What do you consider more important between a healthy diet and moderate alcohol consumption in pursuing a healthy lifestyle?
- 7.1) To what extent do you consider factor A more important than factor B?

