

## Article

# Canonical Correlation for the Analysis of Lifestyle Behaviors versus Cardiovascular Risk Factors and the Prediction of Cardiovascular Mortality: A Population Study

Alessandro Menotti <sup>1</sup> and Paolo Emilio Puddu <sup>2,\*</sup><sup>1</sup> Association for Cardiac Research, 00182 Rome, Italy; amenotti2@gmail.com<sup>2</sup> EA 4650, Signalisation, Électrophysiologie et Imagerie des Lésions D'ischémie Reperfusion Myocardique, UNICAEN, 14000 Caen, France

\* Correspondence: puddu.pe@gmail.com; Tel.: +39-06-5404875 or +39-340-9323920

**Abstract:** Objectives: To assess the overall association of lifestyle behaviors with multiple cardiovascular risk factors and mortality. Material and Methods: In the Italian Rural Areas of the Seven Countries Study, involving 1712 middle-aged men (40–59 years) enrolled in 1960, smoking habits, physical activity, dietary habits, marital status, and socioeconomic status (SES) were studied as possible determinants of 15 measurable risk factors (body mass index, tricipital and subscapular skinfold, arm circumference, systolic and diastolic blood pressure, heart rate, double product (systolic blood pressure × heart rate), vital capacity, forced expiratory volume, serum cholesterol, urine protein, urine glucose, corneal arcus and xanthelasma) using canonical correlation (CC). Results: The first CC had a value of 0.54 ( $R^2$  0.29,  $p < 0.0001$ ). The role of marital status was marginal; that of a high SES was contrary to expectations. The strongest behaviors based on standardized CC coefficients were dietary habits and physical activity. The risk factors mostly associated with overall lifestyle behaviors were some anthropometric and cardiovascular measurements. The mean levels of risk factors distributed in tertile classes of the CC variate score of lifestyle behaviors were largely associated in a coherent and graded way with the expected relationship of behaviors versus risk factors. In a large series of Cox models, the CC variate scores were significantly associated with 50-year coronary heart disease (CHD) mortality and much less with stroke and other heart diseases of uncertain etiology. Conclusions: Lifestyle behaviors correlate well with cardiovascular risk factors associated with CHD mortality, and CC is a useful method of analysis to detect long-term impacting characteristics.

**Keywords:** lifestyle behaviors; risk factors; cardiovascular mortality; canonical correlation; residential cohort; men; middle age



**Citation:** Menotti, A.; Puddu, P.E. Canonical Correlation for the Analysis of Lifestyle Behaviors versus Cardiovascular Risk Factors and the Prediction of Cardiovascular Mortality: A Population Study. *Hearts* **2024**, *5*, 29–44. <https://doi.org/10.3390/hearts5010003>

Academic Editor: Matthias Thielmann

Received: 8 November 2023

Revised: 20 December 2023

Accepted: 22 December 2023

Published: 3 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It is well documented that some lifestyle behaviors may play a great role in health and disease. It is enough to peruse a few long reviews [1–3] to find the relationship of physical activity, smoking, and eating habits with disease and mortality, including the role of mediators between the behaviors and outcome, such as modifiable risk factors. In previous analyses, we showed the role of personal characteristics in predicting long-term mortality, heart diseases, age at death, and longevity in middle-aged men followed up to 50 years [4–13]. Among the predictors, there were lifestyle behaviors and measurable risk factors, which were frequently included in the same predictive models. In the latter case, both types of characteristics were significantly associated with mortality rates and the length of survival.

We thus posed the question of quantifying the role of lifestyle behaviors regarding the levels of measurable risk factors. The purpose of the present analysis was to study the relationship of physical activity, smoking and dietary habits with a series of cardiovascular risk factors commonly measured in population studies using canonical correlation, an

uncommon method adopted quite rarely (see Appendix A) that we use here for the first time to detect long-term impacting characteristics in a residential cohort.

Moreover, we planned to test canonical variables obtained from this analysis as possible predictors of cardiovascular events that occurred over 50 years in a residential cohort of middle-aged men.

## 2. Material and Methods

### 2.1. Study Population and Measurements

The data were derived from the Italian Rural Areas of the Seven Countries Study of Cardiovascular Disease located in two villages in Northern and Central Italy, first examined in 1960. The analysis uses baseline data from an entry field examination in a cohort of 1712 men aged 40–59 years, where the participation rate was 98.7%. The mean age of the participants was 49.8 years (SD = 5.1). More details can be found elsewhere [14].

The personal characteristics used in the present analysis are linked to some lifestyle behaviors and some measurable risk factors that in theory could be modulated by lifestyle behaviors. Among the lifestyle behaviors, we considered as possible causes, and thus called X variables, smoking habits, physical activity, and dietary scores, and, in addition, marital status and socioeconomic status (SES), which may play at least an indirect role in modulating risk factor levels. As possible effects, we considered 15 risk factors, thus called Y variables, including anthropometric (not skeletal), biophysical, biochemical and clinical measurements. All X and Y variables are listed in Table 1, with definitions, units of measurement, bibliographic references and other details [15–20].

In a separate analysis, we used mortality from coronary heart disease (CHD), stroke (STROKE) and other heart diseases of uncertain etiology (HDUEs) as the endpoint in multivariate Cox models of different types.

Collection of mortality data was complete after 50 years. Coding was based on the availability of causes of death with the addition of other information from repeated field examinations, hospital clinical records, other medical documents and interviews with hospitals, family doctors and relatives of the deceased. The final cause of death was assigned by a single coder following the rules of the Seven Countries Study and using the 8th Revision of the World Health Organization International Classification of Diseases (ICD-8) [21]. In cases of multiple causes of death and uncertainties about the choice of the first cause, a decreasing hierarchical rank was adopted with violence, cancer, CHD, STROKE and others in sequence.

Cardiovascular mortality endpoints were chosen as follows: (1) CHD including cases of myocardial infarction, acute ischemic heart attacks and sudden coronary death, after the exclusion of other possible causes (ICD-8 codes 410, 411, 412, 413 and 795); cases with only a mention or evidence of chronic coronary heart diseases (part of code 412) were not included in this group for reasons given in other contributions [10,12,14], while healed myocardial infarction was retained in this group. (2) STROKE including any type of cerebrovascular disease (ICD-8 codes 430–438). (3) HDUEs including a pool of symptomatic heart diseases (ICD-8 code 427 corresponding to heart failure, arrhythmia, blocks), ill-defined hypertensive heart disease (usually in the absence of documented left ventricular hypertrophy) (ICD-8 codes 402–404) and cases classified as chronic or other types of coronary heart disease in the absence of typical coronary syndromes (ICD-8 part of code 412 and 414), usually manifesting in heart failure, arrhythmia and blocks. The reasons for segregating CHD mortality from that of HDUEs are linked to the repeated documentation we presented about their differences at least for risk factors (mainly serum cholesterol higher for CHD) and age at death (definitely higher for HDUEs) [10,12,14].

After 50 years of follow-up, there were 1669 deaths (97.5%), while nobody (among the enrolled 1712 individuals) was lost to follow-up. All cardiovascular disease mortality covered 45.7% of all causes, while 705 belonged to the three major groups described above, covering 92.4% of all cardiovascular fatal events after the exclusion of other groups with well-defined or very rare etiology.

**Table 1.** Lifestyle behaviors (X variables) and measurable risk factors (Y variables): definitions, units of measurement, mean levels and selections.

Risk Factor	Definition and Details	Unit of Measurement	Mean (and SD) or Proportion (%) (and SE)	Type of Variable: X or Y	References	Notes
Cigarette smoking (*)	Derived from a questionnaire and classified as:			X		
	Smokers	Code 1	61.0 (1.2)			
	Ex-smokers	Code 2	13.6 (0.8)			
	Never smoked	Code 3	25.4 (1.1)			
Physical activity (*)	Job-related; derived from questions matched with reported occupations. Classified as:			X	[15,16]	Validated through ergonomic measurements and energy intake
	Sedentary	Code 1	9.7 (0.7)			
	Moderate	Code 2	22.1 (1.0)			
	Vigorous	Code 3	68.2 (1.1)			
Dietary habits (*)	Derived from dietary history. Classified as:			X	[16,17]	Factor score derived from factor analysis of 18 food groups
	Non-Mediterranean diet	Code 1	33.4 (1.1)			Arbitrary denomination and selection
	Intermediate diet	Code 2	33.2 (1.1)			Arbitrary denomination and selection
	Mediterranean diet	Code 3	33.4 (1.1)	X		Arbitrary denomination and selection
Marital status (*)	Currently married	0 = no 1 = yes	90.5 (0.7)	X		From questionnaire
High socioeconomic status (*)	Professional, business, public administrators, foremen and high-ranking clerical workers	0 = no 1 = yes	7.8 (0.6)	X		From questionnaire
Body mass index	Weight/height squared	kg/m <sup>2</sup>	25.2 (3.7)	Y	[18]	
Tricipital skinfold	Right arm	mm	9.4 (5.4)	Y	[18]	
Subscapular skinfold	Right side	mm	11.8 (5.8)	Y	[18]	

Table 1. Cont.

Risk Factor	Definition and Details	Unit of Measurement	Mean (and SD) or Proportion (%) (and SE)	Type of Variable: X or Y	References	Notes
Midarm circumference	Right arm; Tricipital skinfold was mathematically subtracted from midarm circumference for estimating only the muscular mass	mm	268.6 (23.6)	Y	[18,19]	
Systolic blood pressure	Supine; average of two measurements	mmHg	143.6 (21.0)	Y	[18]	
Diastolic blood pressure	Supine; average of two measurements	mmHg		Y	[18]	Fifth phase
Heart rate	From ECG; average rate in lead I and V <sub>6</sub>	beats/min	71.3 (12.9)	Y		
Double product	Systolic blood pressure times heart rate	score	10,328 (2858)	Y		Indicator of oxygen consumption
Vital capacity	Best of two tests; adjusted (divided by height <sup>2</sup> )	L/m <sup>2</sup>	1.65 (0.24)	Y	[18]	
Forced expiratory volume in ¾ s	Best of two tests; adjusted (divided by height <sup>2</sup> )	L/m <sup>2</sup>	1.08 (0.24)	Y	[18]	
Serum cholesterol	Abell–Kendall method modified by Anderson and Keys; casual blood sample	mg/dL	201.6 (40.8)	Y	[20]	
Urine protein (*)	Spot urine sample; semiquantitative method by stix; definite present	0 = absent 1 = present	7.8 (0.65)	Y		
Urine glucose (*)	Spot urine sample; semiquantitative method by stix; definite present.	0 = absent 1 = present	4.5 (0.50)	Y		
Corneal arcus (*)	Clinical judgement	0 = no 1 = yes	13.9 (0.83)	Y		
Xanthelasma (*)	Clinical judgement	0 = no 1 = yes	1.5 (0.30)	Y		

(\*) variable expressed in %.

## 2.2. Statistical Analysis

The analysis was separated into two different parts.

**Part 1. Canonical analysis.** For the purpose of relating lifestyle behaviors with measurable risk factors, we used canonical correlation (see Appendix A), which is an extension of multiple linear regression and correlation, where the dependent variables are more than one [22,23]. This statistical approach allows us to find the relationship of several independent variables, usually called Xs (and arbitrarily considered as possible causes), with several dependent variables, usually called Ys (and arbitrarily considered as possible effects).

The analytical process computes one or more variate Xs and one or more variate Ys that are the linear combinations (weighted averages of the original variables) capable of maximizing the correlation between variate X and variate Y. The estimates of variate Xs and variate Ys are based on Z-transformed variables, since their original values are converted into a mean of around 0 with a standard deviation of around 1. To be conceptually simple, variates X and Y are the pool of possible causes and the pool of possible effects, respectively, computed following the procedure mentioned above. The factor loadings are the correlations of variates with the original variables and reflect their role in the production of variates.

The coefficients of the linear combinations are usually reported as standardized (that is, multiplied by the standard deviation of the variables) and can be applied to individual subjects obtaining a variate score for X and Y separately. Again, to simplify, variate scores X and Y are the numerical characteristics of each individual derived from the pool of the possible causes and of the possible effects, respectively.

Both factor loadings and standardized coefficients are used to evaluate the role of the original variables in the production of variates, but the judgement is not univocal. In fact, when the group of variables is uncorrelated, the canonical loadings are similar to the coefficients and highly correlated. The reverse happens when the original variables are highly correlated, since in this case, factor loadings and standardized coefficients are rather different and uncorrelated.

Canonical correlation is the linear correlation between variate X and variate Y and can be interpreted as the usual R (simple linear correlation), including the  $R^2$  that represents the proportion of variance explained by the correlation. The final R and  $R^2$  represent the highest possible correlation between a linear combination of Xs and a linear combination of Ys.

After the first canonical correlation, others can be computed, but usually their levels are smaller and their interpretation is tricky since they are computed by the residuals of previous correlations. In this way, only the first correlation maximizes the association of the original Xs with Y variables and, for this reason, only one correlation has been considered in this analysis. The final outcome allows us also to invert the interpretation that assigns the role of dependent variables to Xs and the role of independent variables to Ys, an approach that is common in psycho-social sciences. In our case, the only reasonable approach was to consider the X variables as possible causes of the Y variables and the Y variables as possible effects.

In general, it is recognized that sometimes the interpretation of findings of canonical correlation is not easy. Another interesting feature of canonical correlation is that it compacts several variables into a kind of score that is independent from the judgement of the investigator while dependent on it in the so-called a priori scores. However, such compacting is conditioned by the search for the best correlation between the two groups of original variables (X and Y).

The output of the computer program [23] included a correlation matrix among all the variables, a canonical correlation with its  $p$  value, standardized coefficients and the factor loadings of variates X and Y; the last two reflect the rank of their power. To provide an indication of the connection of variate score X with risk factor levels, a table was filled with the levels of risk factors in tertile classes of variate score X levels. Moreover, we computed

the distribution of some risk factor levels in three classes of physical activity and three classes of dietary habits.

**Part 2. Cox proportional hazards predictive models.** In another part of the analysis, a series of Cox models was analyzed to test the role of canonical analysis findings in the prediction of the mortality of three major groups of cardiovascular disease occurring over 50 years:

- (a) Cox models including variate X and variate Y with a continuous shape predicting mortality from CHD, STROKE and HDUEs separately;
- (b) Similar Cox models to above including variate X and variate Y in three tertile classes, using tertile 1 (lowest) as a reference;
- (c) A Cox model with CHD mortality only and the five behavioral characteristics expressed as variable X in the canonical analysis as covariates (plus age);
- (d) The same Cox model as above with the addition of variate Y (continuous shape);
- (e) A Cox model with CHD mortality only and risk factors expressed as Y variables in the canonical analysis as covariates;
- (f) The same Cox model as above with the addition of variate X (continuous shape).

All behaviors were used in this analysis, while for the risk factors, we selected only those which could exclude multicollinearity problems, resulting in only 11 risk factors as follows: body mass index, subscapular skinfold thickness, arm circumference, systolic blood pressure, heart rate, vital capacity, serum cholesterol, urine protein, urine glucose, corneal arcus, xanthelasma (plus age).

The informativeness of the added variable (prediction ability) was estimated by the procedure proposed by Peto [24], as directly proportional to the chi square (twice the change of the model likelihood). This indicator was used to compare model (c) with model (d) and model (e) with model (f) using the loglikelihoods produced by the Cox solutions. The purpose of this was to see whether the addition of variate Y score in the first couple and variate X score in the second couple improved the goodness of fit (in two couples of nested models).

Additional tests were the Akaike information criterion and the AUCs of the models.

### 3. Results

#### 3.1. Canonical Analysis

The correlation matrix across all variables is not reported in detail as it is too bulky. Among 190 comparisons, only 29 (19%) had R values (correlation coefficient) equal or greater than 0.23, a level that explains about 5% or more of variance between two variables. The majority of these relatively high levels were related to anthropometric risk factors, and somewhat less to cardio-circulatory risk factors. There were some correlation coefficients of great interest, such as vigorous physical activity versus heart rate ( $-0.21$ ); sedentary physical activity versus heart rate ( $0.15$ ); non-Mediterranean diet versus body mass index ( $+0.28$ ), tricipital skinfold ( $+0.32$ ), subscapular skinfold ( $+0.20$ ) and systolic blood pressure ( $+0.29$ ); and Mediterranean diet versus BMI ( $-0.29$ ), tricipital skinfold ( $-0.33$ ), subscapular skinfold ( $-0.26$ ), systolic blood pressure ( $-0.24$ ) and heart rate ( $-0.21$ ).

The standardized canonical coefficients of the X and Y variables for the first canonical correlation (Table 2) allow us to compare their magnitude. The group of Xs showed a dominant role in dietary habits, followed by physical activity and smoking habits. The coefficients of diet and physical activity were positive, while those of smoking habits and a high SES were negative, suggesting an opposite effect on risk factor levels. In fact, the highest level for diet corresponds to a Mediterranean diet and the highest level for physical activity corresponds to vigorous activity. The reverse occurred for smoking habits, which overall were associated with higher levels of risk factors (recalling that the coding of smoking habits was 1 = smokers; 2 = ex-smokers; 3 = never). In the group of Ys, large coefficients were found for some anthropometric and cardiovascular measurements. All risk factors had negative coefficients, except arm circumference, double product, vital capacity, forced expiratory volume and urine glucose. When interpreting (see Appendix A),

one should consider the algebraic signs, looking at the same time at the coefficients of the Xs. For example, the coefficient of physical activity (X variable) was large and positive, while those of Y variables such as BMI, tricipital skinfold, systolic blood pressure and heart rate were negative, suggesting an inverse relationship between this X variable with these Y variables. In any case, the algebraic sign of the coefficients should not be interpreted as related to the possible predictive power of the variables in relation to possible events, but only to the direction of their association with the various risk factors.

**Table 2.** Standardized canonical coefficients and canonical loadings of X and Y variables.

X Variables	Standardized Canonical Coefficients	Rank	Canonical Loadings	Rank
Dietary habits	0.8378	1	0.8671	1
Physical activity	0.2934	2	0.4499	2
Smoking habits	−0.2885	3	−0.2497	4
High SES	−0.2016	4	−0.3343	3
Marital status	0.0338	5	0.0626	5
Y Variables	Standardized Canonical Coefficients	Rank	Canonical loadings	Rank
Heart rate	−0.8253	1	−0.5668	6
Double product	0.7576	2	−0.6833	4
Systolic blood pressure	−0.5960	3	−0.5718	5
Tricipital skinfold	−0.3990	4	−0.7928	1
Body mass index	−0.3220	5	−0.7058	3
Arm circumference	0.1107	6	−0.1001	11
Subscapular skinfold	−0.1062	7	−0.7558	2
Diastolic blood pressure	−0.1006	8	−0.5593	7
Urine protein	−0.0957	9	−0.1941	10
Corneal arcus	−0.0640	10	−0.0765	15
Vital capacity	0.0424	11	0.2744	8
Urine glucose	0.0419	12	−0.0877	13
Xanthelasma	−0.0373	13	−0.0975	12
Forced expiratory volume	0.0179	14	0.0859	14
Serum cholesterol	−0.0129	15	−0.2721	9

Units of measurement from Table 1.

A similar evaluation can be conducted by inspecting the canonical loadings reported in Table 2. For X variables, the rank order is sufficiently similar to that of standardized coefficients, and this is supported by the fact that the overall correlation in the matrix of the original variables is rather small (standardized Cronbach’s Alpha test for correlation matrix = 0.0622) and by the high correlation between factor loadings and standardized coefficients (R = 0.97). The situation is rather different for Y variables, where the overall correlation in the matrix of the original variables is relatively high (standardized Cronbach’s Alpha test for correlation matrix = 0.65) and consequently the correlation between factor loadings and standardized coefficient is lower (R = 0.38). The consequence is that the top five ranking variables for standardized coefficients are heart rate, double product, systolic blood pressure, tricipital skinfold and body mass index, while the corresponding five for canonical loadings are tricipital skinfold, subscapular skinfold, body mass index, double product and systolic blood pressure.

The first canonical correlation had a value of 0.54 ( $p < 0.0001$ ), corresponding to an  $R^2$  of 0.29, thus explaining a sizeable proportion of the relationships of the two groups of variables.

To provide an indication of the relationship of the characteristics of people located in the three tertiles of the variate X score, their mean levels are reported in Table 3. For all risk factors (except arm circumference), there was a trend from tertile 1 to tertile 3, ascending for vital capacity and forced expiratory volume, descending for all the others except arm circumference, urine glucose, urine protein, corneal arcus and xanthelasma, whose trends were irregular and not significant. The observed trends suggest that those located in tertile 3 (the highest) enjoy the high values of the ascending risk factors and the low values of the descending risk factors, although this situation is not yet bound by their predictive power. The  $p$  value of the ANOVA across the three tertiles was highly significant for most of the risk factors except for arm circumference, forced expiratory volume and three of the five extremely rare conditions expressed as proportions. Overall, the risk profile was definitely more favorable for tertile 3 than for tertile 1. For example, comparing these two extremes for two major risk factors, a difference of 14 mmHg in systolic blood pressure and 11 mg/dL in serum cholesterol corresponded to a difference of about 2 years in life expectancy according to a model available in a previous analysis on the same population and the same follow-up duration [13].

**Table 3.** Mean levels of risk factors distributed in tertile classes of variate X scores.

Risk Factors	Tertile 1		Tertile 2		Tertile 3		$p$ of ANOVA
	Mean	SD	Mean	SD	Mean	SD	
Body mass index	26.7	3.96	25.19	3.57	23.69	2.83	<0.0001
Tricipital skinfold	12.0	6.0	9.4	4.9	6.9	3.8	<0.0001
Subscapular skinfold	14.4	6.4	11.7	5.5	9.3	4.0	<0.0001
Arm circumference	270.0	25.7	256.7	23.6	267.0	21.3	0.0814
Systolic blood pressure	151.4	22.7	142.1	19.5	137.4	17.9	<0.0001
Diastolic blood pressure	89.6	11.9	84.6	10.4	82.3	10.0	<0.0001
Heart rate	75.7	14.0	71.0	11.9	67.2	11.1	<0.0001
Double product	11,556	3222	10,145	2529	9285	2265	<0.0001
Vital capacity	1.60	0.25	1.65	0.24	1.69	0.22	<0.0001
Forced expiratory volume	1.07	0.24	1.09	0.26	1.10	0.23	0.1787
Serum cholesterol	206.9	43.5	202.4	39.7	195.5	38.3	<0.0001
Urine protein *	0.12	0.01	0.06	0.01	0.49	0.009	<0.0001
Urine glucose *	0.05	0.01	0.05	0.01	0.03	0.01	0.1498
Corneal arcus *	0.15	0.02	0.15	0.02	0.12	0.01	0.1853
Xanthelasma *	0.021	0.144	0.016	0.125	0.037	0.073	0.2346

Unit of measurements as from Table 1. (\*) Proportions and standard error.

Table 4 provides a few selected examples that quantify the effect on levels of single risk factors across the three levels of each lifestyle behavior. We decided to choose physical activity and dietary scores since they were the most powerful as defined by the standardized canonical coefficients, while the “dependent risk factors” were selected as those more likely influenced by the correspondent behavior (see Table 2). All of them were coherent and graded following the expectation on their relationships. In fact, it was expected that the arm circumference and forced expiratory volume would increase and that double product would decrease with increasing levels of physical activity, and similarly that systolic blood pressure, BMI and serum cholesterol would decrease moving from a non-Mediterranean diet to a Mediterranean diet.

**Table 4.** Some examples of mean levels of risk factors in classes of physical activity and dietary score.

Physical Activity	Arm Circumference mm		Forced Expiratory Volume L/m <sup>2</sup>		Double Product SBP × HR	
	Mean	SD	Mean	SD	Mean	SD
Sedentary	259	26.7	1.043	0.26	11,580	3307
Moderate	268	25.6	1.062	0.25	10,828	3104
Vigorous	270	22.1	1.097	0.24	9988	2627
<i>p</i> of ANOVA	<0.001		0.004		<0.001	
Diet Score	Systolic Blood Pressure mmHg		Body Mass Index Kg/m <sup>2</sup>		Serum Cholesterol mg/dL	
	Mean	SD	Mean	SD	Mean	SD
Non-Mediterranean	152.2	23.0	26.6	26.7	206.1	40.6
Intermediate	143.2	18.8	25.2	25.3	202.2	48.7
Mediterranean	136.4	17.6	23.7	23.6	196.4	37.5
<i>p</i> of ANOVA	<0.001		<0.001		<0.001	

3.2. Cox Models Prediction

Tables 5–7 provide findings of Cox models using variate X and Y scores in both continuous and discrete shapes for mortality from CHD, STROKE and HDUE separately. For CHD models (Table 5), all coefficients were negative, meaning that the levels of covariates were inversely related to events and all were significant, except tertile 2 of the variate Y score. In general, coefficients and hazard ratios were larger for variate X score (behaviors) than for variate Y score. Findings were entirely different for STROKE models, as only variate Y score (continuous) was negative and significant (Table 6). A slightly better situation was found for HDUE models, where tertiles 2 and 3 of variate X scores were negative and significant (behaviors) (Table 7).

**Table 5.** Cox proportional hazards model of variate scores (X and Y) as covariates expressed in continuous and discrete shape predicting 50-year mortality from CHD (n = 278) as a dependent variable.

Cox Model Predicting CHD Mortality with X and Y Variates in a Continuous Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	<i>p</i> of Coefficient
X variate score continuous	−0.2573	0.77	0.67–0.89	0.0003
Y variate score continuous	−0.1582	0.85	0.74–0.98	0.0255
Cox Model Predicting CHD Mortality with X and Y Variates in a Discrete Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	<i>p</i> of Coefficient
Tertile 1 X variate score	Reference	—	—	—
Tertile 2 X variate score	−0.2362	0.59	0.44–0.79	0.0004
Tertile 3 X variate score	−0.6980	0.68	0.49–0.94	0.0188
Tertile 1 Y variate score	Reference	—	—	—
Tertile 2 Y variate score	−0.5321	0.79	0.59–1.05	0.1047
Tertile 3 Y variate score	−0.3871	0.50	0.36–0.70	0.0001

CI: confidence intervals.

**Table 6.** Cox proportional hazards model of variate scores (X and Y) as covariates expressed in continuous and discrete shape predicting 50-year mortality from STROKE (n = 225) as dependent variable.

Cox Model Predicting STROKE Mortality with X and Y Variates in a Continuous Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	p of Coefficient
X variate score continuous	−0.0386	0.96	0.82–1.13	0.6305
Y variate score continuous	−0.2185	0.80	0.69–0.94	0.0065
Cox Model Predicting STROKE Mortality with X and Y Variates in a Discrete Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	p of Coefficient
Tertile 1 X variate score	Reference	—	—	—
Tertile 2 X variate score	−0.1454	0.86	0.62–1.21	0.3917
Tertile 3 X variate score	−0.2424	0.78	0.54–1.13	0.1991
Tertile 1 Y variate score	Reference	—	—	—
Tertile 2 Y variate score	0.0099	1.01	0.73–1.41	0.9530
Tertile 3 Y variate score	−0.2703	0.76	0.53–1.11	0.1551

CI: confidence interval.

**Table 7.** Cox proportional hazards model of variate scores (X and Y) as covariates expressed in continuous and discrete shape predicting 50-year mortality from HDUE (n = 202) as a dependent variable.

Cox Model Predicting HDUE Mortality with X and Y Variates in a Continuous Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	p of Coefficient
X variate score continuous	−0.1149	0.89	0.75–1.05	0.1799
Y variate score continuous	0.0244	1.02	0.86–1.22	0.7845
Cox Model Predicting HDUE Mortality with X and Y Variates in a Discrete Shape				
Covariates	Coefficient	Hazard Ratio	95% CI	p of Coefficient
Tertile 1 X variate score	Reference	—	—	—
Tertile 2 X variate score	−0.4502	0.64	0.44–0.92	0.0147
Tertile 3 X variate score	−0.5327	0.59	0.40–0.86	0.0070
Tertile 1 Y variate score	Reference	—	—	—
Tertile 2 Y variate score	0.1216	1.13	0.77–1.65	0.5286
Tertile 3 Y variate score	0.2913	1.33	0.90–1.98	0.1470

CI: confidence intervals.

The last test was limited to the CHD mortality as an endpoint since this was the one with the best outcome in the previous analyses. In this case (Table 8), we found that adding the variate Y score (representing the pool of “dependent” risk factors) to a basic model (run with the single five behaviors plus age) provided a significant improvement in the model loglikelihood, which is an indicator of goodness of fit based on the informativeness procedure [24]. Similarly, when the basic model included the risk factors, the addition of the variate X score representing the behaviors again made the loglikelihood improve significantly, but to a lesser extent. The improvement in the models was confirmed by the computation of the Akaike information criterion and the AUCs. In particular, these last two tests were coherent with each other, indicating a better performance of the couple of Models 1 and 2 than of the couple of Models 3 and 4 (models are here numbered as in Table 8). In any case, these findings suggest that the pools of both behaviors (possible causes) and risk factors (possible effect of behaviors) carry important and significant information that improve prediction.

**Table 8.** Cox proportional hazards models predicting CHD mortality in 50 years ( $n = 278$ ) in four shapes including five behaviors alone plus variate Y score, and separately eleven risk factors alone plus variate X score.

	<b>Model 1</b> Five Behaviors Plus Age	<b>Model 2</b> Behaviors Plus Variate Y Score	<b>Model 3</b> Eleven Risk Factors Plus Age	<b>Model 4</b> Eleven Risk Factors Plus Variate X Score
Loglikelihood	−1821	−1812	−1799	−1796
Chi squared Informativeness	Model 1 versus model 2 17.8 ( $p < 0.0001$ )		Model 3 versus Model 4 5.2 ( $p = 0.0226$ )	
Akaike information criterion	−3.01	−1.00	9.01	11.00
AUC	0.524	0.547	0.552	0.557
$p$ of difference	Model 1 versus Model 2 0.0477		Model 3 versus Model 4 0.2206	

We also ran another separate Cox model, with behaviors and risk factors fed in a traditional way, excluding (like in previous ones) a few risk factors that could produce multicollinearity problems (not reported in detail). In this case, a significant predictive role was found for three (out of five) behaviors (physical activity, dietary score, smoking habits) and for three out of eleven risk factors (systolic blood pressure, serum cholesterol, vital capacity). However, the outcome could have been conditioned by the saturation effect linked to the use of too many covariates.

#### 4. Discussion

This analysis tends to confirm that some classical lifestyle behaviors such as smoking habits, physical activity and eating habits play a role in health at least partly through the modulation of some measurable risk factors. Overall, these lifestyle behaviors, together with marital status and a high SES, explain almost 30% of the variance in a group of 15 risk factors that are usually good predictors of cardiovascular diseases, all-cause mortality, age at death and life expectancy during long follow-up periods.

However, the role of marital status proved to be almost negligible, while that of SES was not major and apparently played a role against the expectation. In fact, its canonical coefficient was negative, like that of smoking habits, and the correlation matrix showed a direct association with some risk factors carrying adverse levels such as subscapular skinfold and sedentary physical activity. This may mean that in the middle of last century, people with a high SES were not healthier than others as apparently is the case nowadays, suggesting that SES is a variable whose meaning, in terms of public health, is changing on the basis of location and time.

The risk factors mostly associated with those lifestyle habits seem to be some anthropometric and cardiovascular measurements. Behaviors carrying the largest canonical coefficients were dietary habits and physical activity, both with three levels going from a non-Mediterranean diet to a Mediterranean diet and from sedentary to vigorous physical activity. These are the same behaviors that have a strong influence on blood pressure, BMI and cholesterol (for diet), and on arm circumference, forced expiratory volume and double product (for physical activity). On the other hand, systolic blood pressure, heart rate and double product are those that have the largest standardized canonical coefficients, followed by some anthropometric measurements.

An important aspect of these findings is the graded and significant coherence between the levels of lifestyle behaviors and those of the risk factors, as clearly shown in Tables 4 and 5. The fact is that when these lifestyle behaviors are fed into multivariate models as predictors of cardiovascular mortality, all-cause mortality or age at death, together with the usual measurable risk factors, both types of predictors still play a significant predictive role [13]. The consequent hypothesis is that these lifestyle behaviors play an extra

role, acting also on risk factors not available in our data or that are simply still unknown, probably including genetic markers.

However, looking to the models dedicated to the three cardiovascular endpoints, it appears that the predictive role of canonical variate scores is almost always significant for CHD, which is not the case for STROKE and HDUEs. This perhaps suggests that the three conditions have different risk factors and/or different relationships with the available risk factors, a fact that has been demonstrated in the same epidemiological material using traditional statistical approaches [4,9], forcing us to conclude that we are facing different diseases, at least in terms of their relationship with serum cholesterol (a significant predictor of CHD but not of STROKE or HDUEs) and different ages at death (shorter for CHD, longest for HDUEs, intermediate for STROKE) [11]. Moreover, the set of behaviors and risk factors used in this analysis is more suitably bound to CHD than to the other conditions. All this suggests the possible existence of a situation of competing risks, but this problem has already been tackled and substantially solved in a dedicated analysis of the same material [11], where the existence of a competition between CHD and HDUEs and STROKE was shown, mainly mediated by different relationships with serum cholesterol.

From a strictly technical point of view, it is interesting that adding the variate Y score to the predictive model initially solved with the single behaviors provided a significant increase in the informativeness, that is, the predictive ability of the added covariate. The same was found when the initial model included only risk factors and the additional covariate was the variate X score. This was clearly shown for the endpoint used for this part of the analysis, that is, CHD mortality. We can hypothesize that compacting, say, a block of risk factors or a block of behaviors into single covariates might be helpful to limit the excess of covariates in multivariate predictive models. However, much more work is needed to reach this goal and thus produce valuable operative procedures. The limits of this analysis are linked to the relatively small size of the population sample, partially compensated by the long follow-up; the absence of women in the cohort; and the limited number of risk factors, although in the literature it is rare to find analyses including many more risk factors.

The attempt to use variate scores for the prediction of events suggests that the canonical regression procedure, at least for these combinations of X and Y variables, somewhat helps in selecting subgroups of favorable versus unfavorable risk profiles, opening another path in this area. This does not mean that canonical correlation should substitute other well-established predictive models, but it is worth noting that there is a coherence between lifestyle behaviors, risk factor levels and the predictive power of these variables.

The contribution of this analysis probably has value due to its systematic and comprehensive approach including a defined population sample, the use of canonical correlation with more than one X variable (five in this case) and many Y variables (fifteen in this case). Moreover, the study of the relationships and the influence of lifestyle behaviors could exploit the availability, in the same population, of long-term (50 years of follow-up, close to extinction) mortality data that became the endpoints of the possible predictive power of the canonical variate score related to three major cardiovascular mortality groups. Finally, the coexistence in the same model of behaviors (possible causes) and risk factors (possible effects) as predictors showed that the former have some extra contributions as determinants of events beyond their influence on risk factors.

A review of the literature highlights that contributions can be divided into two groups, that is, the study of the relationship of behavior with risk factors and the use of canonical correlation for the same or other purposes in the cardiological domain. In the first group, the majority of the quoted studies dealt with one lifestyle behavior and a limited number of risk factors [25–35]. We found four studies that used physical activity, variously defined and measured, as a possible determinant of measurable risk factors, mainly blood lipids and components of metabolic syndrome [27,29,31,33], with findings tending to associate high levels of physical activity with a low BMI, low levels of triglycerides and LDL and total cholesterol and high levels of HDL cholesterol. Moreover, in an Australian study, healthy

behaviors of sleeping, physical activity and dietary habits were associated with lower levels of LDL cholesterol [34]. An a posteriori dietary score derived from a population study involving 2298 adults in France, Belgium and Luxembourg showed lower levels of cardiovascular risk factors when the diet was rich in fruit, nuts, vegetables oils and tea [28]. In a Polish study, the interactions of various lifestyle behaviors have been considered, but the upper tertile of the nutrition score (the healthy habit) was not associated with vigorous physical activity, while in this group, there were more people reading books and watching television [32]. Finally, in Canada, a survey on children aged 10–11 years showed that parental/peer smoking and drinking and low self-esteem were associated with multiple risk factors linked to behaviors [26]. The nutrition score and the physical activity score were associated in different ways with triglycerides, waist circumference and systolic blood pressure depending on the levels of BMI [30], suggesting that exercise had a better effect than nutrition in the subgroup of people with a normal BMI. Finally, an interesting survey conducted in the USA assessed the relationship of lifestyle behavior in couples in a study on employees. A high concordance was found for non-healthy behaviors and risk factor levels, but the relationship between the two types of variables was not considered [35].

Even rarer were contributions dealing with canonical correlation versus cardiovascular diseases. On the contrary, it is quite common in psycho-social sciences, neurology, genetics and biochemistry. Also, our experience in this field was limited to a single paper from 1994, where canonical correlation was used to compare a few cardiovascular risk factors with a few causes of death in an ecological analysis of the 25-year follow-up of the 16 cohorts of the Seven Countries Study [36].

In the output of PubMed, when searching for canonical correlation and cardiovascular diseases, less than 300 references can be found, many of which actually do not include the use of canonical correlation. Among the others, the majority deal with definitely clinical problems, either diagnostic, prognostic, therapeutic or even autoptic. At the end, we selected only six papers with at least some indirect connection with the content of our contribution [37–42]. A paper from Romania had a promising title, but the abstract did not provide anything precise and the full text was not available [37]. A study on 407 healthy Taipei Chinese adults successfully used canonical correlation to show the strict relationship of central adiposity with major cardiovascular risk factors [38]. A similar, much larger study conducted in Canada reached the same conclusions, with significant canonical correlations of 0.58 for adult men and 0.61 for adult women [39]. In the Framingham Heart Study, a variation of canonical correlation was used to identify the association of multiple repeatedly measured characteristics with single-nucleotide polymorphism data [40]. In another Chinese study [41] using canonical correlation in a residential cohort, physical activity was related to anthropometric parameters and blood lipids, obtaining a correlation of 0.44 ( $p < 0.0001$ ). However, curiously enough, anthropometric parameters were used as behaviors and not as a possible consequence of behaviors. Finally, the association of road traffic noise and air quality was related to more than 700,000 Scottish hypertensive patients from different areas, and the canonical correlation was 0.342, with 89% of the variance explained by the canonical independent variables [42].

This short review of the literature is disappointing, since the quoted papers generically deal with the relationships of lifestyle behaviors with risk factors and are characterized by the use of only one or few behaviors, and the case was the same for the consequent risk factors. In general, the analyses were very different from the one presented in this paper. Similarly, the contributions using canonical correlation in the field of cardiovascular diseases also had limited horizons and were very different from the systematic approach we used to tackle our data derived from a long-term population study. Although, in general, other investigations agree with our findings, none of them were fully comparable with this analysis in terms of methodology. In our case, using canonical correlation, a good association was found between some lifestyle behaviors and a small series of risk factors bound to cardiovascular diseases. Moreover, the canonical X and Y variate scores were capable of predicting CHD mortality in a satisfactory way. Thus, more analyses are needed

in this field, hopefully exploiting larger population samples and larger numbers of both behaviors and risk factors to obtain a better picture of the problem.

## 5. Conclusions

In a long follow-up population study, the use of canonical correlation proved to be useful in identifying the relationships of major lifestyle behaviors with a number of established cardiovascular risk factors. Moreover, the canonical variates and variate scores showed their ability to predict the mortality of major CVDs in multivariate models, thus providing a tool that, by compacting predictive variables, contributes to limiting their number and reducing the need to force them into these models.

**Author Contributions:** Conceptualization, A.M. and P.E.P.; Methodology and Analysis, A.M.; Writing—Original Draft, A.M. and P.E.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The board of directors of the institution involved in data collections were de facto playing the role of ethical committee, approving the execution of the study on the basis of the local existing legislation by the date this investigation started. The study was conducted in accordance with the Declaration of Helsinki, and since it was started before this Declaration, it was not formally approved by an Institutional Review Board (or Ethics Committee).

**Informed Consent Statement:** Baseline measurements were taken before the era of the Helsinki Declaration and approval was implied in participation, while verbal or written consent was obtained for the collection of follow-up data.

**Data Availability Statement:** The original data are not publicly available. However, research projects are evaluated centrally by an ad hoc committee.

**Acknowledgments:** The authors acknowledge the contribution of Giovina Catasta for help obtaining recent follow-up data of the participants in the Italian areas.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

**Technical Note:** Canonical correlation is a little known and little used procedure, as is its relative complexity. Below is a short list of terms that are specific to this statistical approach:

- X and Y variables: two groups of original variables that are treated as opposite;
- Variate X and variate Y: the linear combinations of X and Y variables, respectively (weighted averages of the original variables), that maximize the correlation between variate X and variate Y;
- Variate X and Y coefficients: the regression coefficients of each variable within variate X and variate Y;
- Canonical correlation: the linear correlation between variate X and variate Y;
- Variate X score and variate Y score: values obtained by applying the coefficients of the linear combinations of variate X and Y to each individual separately;
- Canonical coefficients: coefficients of the linear combination of the original variables in the construction of variates. They are an indicator of the influence of variables in the construction of variates;
- Canonical loadings: the correlation coefficients of variates with the original variables. They are another indicator of the influence of variables in the construction of variates.

## References

1. Prabhat, J.P. The hazards of smoking and the benefits of cessation: A critical summation of the epidemiological evidence in high-income countries. *eLife* **2020**, *9*, e49979.
2. Physical Activity and Health. *A Report of the Surgeon General*; US Department of Health and Human Services, Center for Disease Control and Prevention: Atlanta, GA, USA, 1996; pp. 1–278.

3. The Mediterranean Diet. *An Evidence-Based Approach*, 2nd ed.; Preedy, V.R., Watson, R.R., Eds.; Academic Press: London, UK, 2020; pp. 1–588.
4. Menotti, A.; Lanti, M.; Maiani, G.; Kromhout, D. Forty-year mortality from cardiovascular diseases and their risk factors in men of the Italian rural areas of the Seven Countries Study. *Acta Cardiol.* **2005**, *60*, 521–531. [[CrossRef](#)] [[PubMed](#)]
5. Menotti, A.; Lanti, M.; Maiani, G.; Kromhout, D. Determinants of longevity and all-cause mortality among middle-aged men. Role of 48 risk factors in a 40-year follow-up of Italian rural areas in the Seven Countries Study. *Aging Clin. Exp. Res.* **2006**, *18*, 394–406. [[CrossRef](#)] [[PubMed](#)]
6. Puddu, P.E.; Menotti, A.; Tolonen, H.; Nedeljkovic, S.; Kafatos, A.G. Determinants of 40-year all-cause mortality in the European cohorts of the Seven Countries Study. *Eur. J. Epidemiol.* **2011**, *26*, 595–608. [[CrossRef](#)] [[PubMed](#)]
7. Puddu, P.E.; Menotti, A. Artificial neural networks versus proportional hazards Cox models to predict 45-year all-cause mortality in the Italian rural areas of the Seven Countries Study. *BMC Med. Res. Methodol.* **2012**, *12*, 100. [[CrossRef](#)] [[PubMed](#)]
8. Menotti, A.; Puddu, P.E.; Lanti, M.; Maiani, G.; Fidanza, F. Cardiovascular risk factors predict survival in middle-aged men during 50 years. *Eur. J. Intern. Med.* **2013**, *24*, 67–74. [[CrossRef](#)] [[PubMed](#)]
9. Menotti, A.; Puddu, P.E.; Lanti, M.; Maiani, G.; Catasta, G.; Alberti Fidanza, A. Lifestyle habits and mortality from all and specific causes of death: 40-year follow-up in the Italian rural areas of the Seven Countries Study. *J. Nutr. Health Aging* **2014**, *18*, 314–321. [[CrossRef](#)]
10. Menotti, A.; Puddu, P.E. Lifetime prediction of coronary heart disease and heart disease of uncertain etiology in a 50-year follow-up population study. *Int. J. Cardiol.* **2015**, *196*, 55–60. [[CrossRef](#)]
11. Puddu, P.E.; Piras, P.; Menotti, A. Lifetime competing risks between coronary heart disease mortality and other causes of death during 50 years of follow-up. *Int. J. Cardiol.* **2017**, *228*, 359–363. [[CrossRef](#)]
12. Menotti, A.; Puddu, P.E.; Maiani, G.; Catasta, G. Lifestyle behavior and lifetime incidence of heart diseases. *Int. J. Cardiol.* **2015**, *201*, 293–299. [[CrossRef](#)]
13. Menotti, A.; Puddu, P.E.; Maiani, G.; Catasta, G. Age at death as a useful indicator of healthy aging at population level: A 50-year follow-up of the Italian rural areas of the Seven Countries Study. *Aging Exp. Clin. Res.* **2018**, *30*, 901–911. [[CrossRef](#)] [[PubMed](#)]
14. Menotti, A.; Puddu, P.E. How the Seven Countries Study contributed to the launch and development of cardiovascular epidemiology in Italy. A historical perspective. *Nutr. Metab. Cardiovasc. Dis.* **2020**, *30*, 368–383. [[CrossRef](#)] [[PubMed](#)]
15. Menotti, A.; Puddu, P.E. Ten-year mortality from coronary heart disease among 172,000 men classified by occupational physical activity. *Scand. J. Work. Environ. Health* **1979**, *5*, 100–108. [[CrossRef](#)] [[PubMed](#)]
16. Alberti Fidanza, A.; Seccareccia, F.; Torsello, S.; Fidanza, F. Diet of two rural population groups of middle-aged men in Italy. *Intern. J. Vit. Nutr. Res.* **1988**, *58*, 442–451.
17. Menotti, A.; Alberti-Fidanza, A.; Fidanza, F.; Lanti, M.; Fruttini, D. Factor analysis in the identification of dietary patterns and their predictive role in morbid and fatal events. *Public Health Nutr.* **2012**, *15*, 1232–1239. [[CrossRef](#)] [[PubMed](#)]
18. Rose, G.; Blackburn, H. *Cardiovasc. Survey Methods*; World Health Organization: Geneva, Switzerland, 1968; pp. 1–188.
19. Hemsfield, S.B.; MacManus, C.; Smith, J.; Stevens, V.; Nixon, D.W. Anthropometric measurement of muscle mass: Revised equations for calculating bone-free arm muscle area. *Am. J. Clin. Nutr.* **1982**, *36*, 680–690. [[CrossRef](#)] [[PubMed](#)]
20. Anderson, J.T.; Keys, A. Cholesterol in serum and lipoprotein fractions: Its measurement and stability. *Clin. Chem.* **1956**, *2*, 145–159. [[CrossRef](#)]
21. World Health Organization. *International Classification of Diseases and Causes of Death*, 8th ed.; World Health Organization: Geneva, Switzerland, 1965; pp. 1–671.
22. Afifi, A.A.; Clark, V. *Computer-Aided Multivariate Analysis*, 2nd ed.; Van Nostrand Reinhold: New York, NY, USA, 1990; pp. 1–505.
23. *NCSS-11 Statistical Software*; NCSS, LLC: Kaysville, UT, USA, 2016. Available online: <https://www.ness.com/software/ncss> (accessed on 8 November 2023).
24. Prospective Studies Collaboration. Blood cholesterol and vascular mortality by age, sex and blood pressure: A meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet* **2007**, *370*, 1829–1839. [[CrossRef](#)]
25. Van Oort, S.; Beulens, J.W.J.; van Ballengooijen, J.; Burgess, S.; Larsson, S.C. Cardiovascular risk factors and lifestyle behaviors in relation to longevity: A Mendelian randomization study. *J. Intern. Med.* **2021**, *289*, 232–242. [[CrossRef](#)]
26. Alamian, A.; Paradis, G. Individual and social determinants of multiple chronic disease behavioral risk factors among youth. *BMC Public Health* **2012**, *12*, 224. [[CrossRef](#)]
27. Honda, T.; Chen, S.; Kishimoto, H.; Narazaki, K.; Kumagai, S. Identifying associations between sedentary time and cardiometabolic risk factors in working adults using objective and subjective measures: A cross sectional analysis. *BMC Public Health* **2014**, *14*, 1307. [[CrossRef](#)] [[PubMed](#)]
28. Crichton, G.; Alkerwi, A. Physical activity, sedentary behavior time and lipid levels in the observation of cardiovascular risk factors in Luxembourg study. *Lipids Health Dis.* **2015**, *14*, 87. [[CrossRef](#)] [[PubMed](#)]
29. Rao, D.P.; Orpana, H.; Krewski, D. Physical activity and non-movement behaviors: Their independent and combined associations with metabolic syndrome. *Int. J. Behav. Nutr. Phys. Act.* **2016**, *13*, 26. [[CrossRef](#)] [[PubMed](#)]
30. Sauvageot, N.; Leite, S.; Alkerwi, A.; Sisanni, L.; Zannad, F.; Saverio, S. Association of empirically derived dietary patterns with cardiovascular risk factors. A comparison of PCA with RRR methods. *PLoS ONE* **2016**, *11*, e0161298. [[CrossRef](#)] [[PubMed](#)]
31. Silfee, V.; Lemon, S.; Lora, V.; Rosal, M. Sedentary behavior and cardiovascular disease risk factors among Latino adults. *J. Health Care Poor Underserved* **2017**, *28*, 798–811. [[CrossRef](#)] [[PubMed](#)]

32. Jezewska-Zychowicz, M.; Gebski, J.; Guzek, D.; Swiatkowska, M.; Stangierska, D.; Pitchta, M. The association between dietary pattern and sedentary behaviors in Polish adults (Lifestyle Study). *Nutrients* **2018**, *10*, 1004. [[CrossRef](#)] [[PubMed](#)]
33. Gubelmann, C.; Antiochos, P.; Vollerweider, P.; Marques-Vidal, P. Association of activity behaviors and patterns with cardiovascular risk factors in Swiss middle-age adults. The Colaus study. *Prev. Med. Rep.* **2018**, *11*, 31–36. [[CrossRef](#)]
34. Dash, S.R.; Hoare, E.; Varsamis, P.; Jennings, G.L.R.; Kingwill, B.A. Sex-specific lifestyle and biomedical risk factors for chronic disease among early-middle, middle and older aged Australian adults. *Int. J. Environ. Res. Public Health* **2019**, *16*, 224. [[CrossRef](#)]
35. Shiffman, D.; Louie, J.Z.; Devlin, J.J.; Rowlan, C.M.; Mora, S. Concordance of cardiovascular risk factor and behaviors in a multiethnic US nationwide cohorts of married couples and domestic partners. *JAMA Netw. Open* **2020**, *3*, e2022119. [[CrossRef](#)]
36. Menotti, A.; Seccareccia, F. Risk factors and mortality patterns in the Seven Countries Study. In *Lessons for Science from the Seven Countries Study*; Toshima, H., Loga, Y., Blackburn, H., Keys, A., Eds.; Springer: Tokyo, Japan, 1994; pp. 17–33.
37. Băldescu, R.; Macarie, E.; Schioiu-Costache, L.; Suciuc, A. Canonical correlation analysis as a special method for the study of the structural relations of risk factors in cardiovascular diseases. *Rom. J. Intern. Med.* **1991**, *29*, 133–138.
38. Lyu, L.C.; Shieh, M.J.; Bailey, G.E.; Carrasco, W.I.; Ordivas, J.M.; Lichtenstein, A.H.; Schaefer, J. Relationship of body fat distribution with cardiovascular risk factors in healthy Chinese. *Ann. Epidemiol.* **1994**, *4*, 434–444. [[CrossRef](#)] [[PubMed](#)]
39. Reeder, B.A.; Senthilselvan, A.; Despres, J.P.; Angel, A.; Liu, L.; Wang, H.; Rabkin, S.W. The association of cardiovascular disease risk factors with abdominal obesity in Canada. Canadian Heart Health Surveys Research Group. *CMAJ* **1997**, *157* (Suppl. S1), S39–S45. [[PubMed](#)]
40. Waaijenborg, S.; Zwindermman, A.H. Associating multiple longitudinal traits with high-dimensional single-nucleotide polymorphism data: Application to the Framingham Heart Study. *BMC Proc.* **2009**, *15* (Suppl. S7), S47. [[CrossRef](#)] [[PubMed](#)]
41. Yu, N.; Zhang, Q.; Zhang, L.; He, T.; Liu, Q.; Zhang, S. Canonical correlation analysis (CCA) of anthropometric parameters and physical activities and blood lipids. *Lipids Health Dis.* **2017**, *16*, 236. [[CrossRef](#)]
42. Adza, W.K.; Hursthouse, A.S.; Miller, J.; Boakye, D. Exploring the Joint Association of Road Traffic Noise and Air Quality with Hypertension Using QGIS. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2238. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.