



**HAL**  
open science

## **Towards precision cardiometabolic prevention: results from a machine learning, semi-supervised clustering approach in the nationwide population-based ORISCAV-LUX 2 study**

Guy Fagherazzi, Lu Zhang, Gloria Aguayo, Jessica Pastore, Catherine Goetzinger, Aurélie Fischer, Laurent Malisoux, Hanen Samouda, Torsten Bohn, Maria Ruiz-Castell, et al.

### ► **To cite this version:**

Guy Fagherazzi, Lu Zhang, Gloria Aguayo, Jessica Pastore, Catherine Goetzinger, et al.. Towards precision cardiometabolic prevention: results from a machine learning, semi-supervised clustering approach in the nationwide population-based ORISCAV-LUX 2 study. *Scientific Reports*, 2021, 11 (1), pp.16056. 10.1038/s41598-021-95487-5 . inserm-03831020

**HAL Id: inserm-03831020**

**<https://inserm.hal.science/inserm-03831020>**

Submitted on 26 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

## Towards precision cardiometabolic prevention: results from a machine learning, semi-supervised clustering approach in the nationwide population-based ORISCAV-LUX 2 study

Guy Fagherazzi<sup>1,2✉</sup>, Lu Zhang<sup>3</sup>, Gloria Aguayo<sup>1</sup>, Jessica Pastore<sup>1</sup>, Catherine Goetzinger<sup>1,4</sup>, Aurélie Fischer<sup>1</sup>, Laurent Malisoux<sup>1</sup>, Hanan Samouda<sup>1</sup>, Torsten Bohn<sup>1</sup>, Maria Ruiz-Castell<sup>1</sup> & Laetitia Huiart<sup>1,4</sup>

Given the rapid increase in the incidence of cardiometabolic conditions, there is an urgent need for better approaches to prevent as many cases as possible and move from a one-size-fits-all approach to a precision cardiometabolic prevention strategy in the general population. We used data from ORISCAV-LUX 2, a nationwide, cross-sectional, population-based study. On the 1356 participants, we used a machine learning semi-supervised cluster method guided by body mass index (BMI) and glycated hemoglobin (HbA1c), and a set of 29 cardiometabolic variables, to identify subgroups of interest for cardiometabolic health. Cluster stability was assessed with the Jaccard similarity index. We have observed 4 clusters with a very high stability (ranging between 92 and 100%). Based on distinctive features that deviate from the overall population distribution, we have labeled Cluster 1 (N = 729, 53.76%) as “Healthy”, Cluster 2 (N = 508, 37.46%) as “Family history—Overweight—High Cholesterol”, Cluster 3 (N = 91, 6.71%) as “Severe Obesity—Prediabetes—Inflammation” and Cluster 4 (N = 28, 2.06%) as “Diabetes—Hypertension—Poor CV Health”. Our work provides an in-depth characterization and thus, a better understanding of cardiometabolic health in the general population. Our data suggest that such a clustering approach could now be used to define more targeted and tailored strategies for the prevention of cardiometabolic diseases at a population level. This study provides a first step towards precision cardiometabolic prevention and should be externally validated in other contexts.

Globally, the epidemic of cardiometabolic diseases, such as type 2 diabetes and hypertension, is rising, thus there is an urgent need for better tools to manage the crisis and prevent as many cases as possible<sup>1</sup>. Primary prevention has been shown to be possible; lifestyle intervention, medication or bariatric surgery strategies have shown to be efficient to reduce the incidence of type 2 diabetes or hypertension in at-risk individuals<sup>2–4</sup>. However, these strategies may be sub-optimal and do not rely on a complete understanding of the detailed cardiometabolic profiles of the general population<sup>5</sup>. Most of the screening and prevention strategies are simply based on a few factors such as age, body mass index, metabolic syndrome, hyperglycemia or risk score such as Findris<sup>6</sup> to identify eligible people. We tend to omit a potentially high variability in individuals at a given level of risk, for instance in terms of genetic profiles<sup>7</sup>, inflammation, oxidative stress<sup>8</sup>, insulin resistance<sup>9</sup> and hepatic gluconeogenesis<sup>10</sup>, that could open a window of opportunity for more relevant strategies.

<sup>1</sup>Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, 1A-B, rue Thomas Edison, 1445 Strassen, Luxembourg. <sup>2</sup>Center of Epidemiology and Population Health UMR 1018, Inserm, Gustave Roussy Institute, Paris South - Paris Saclay University, Villejuif, France. <sup>3</sup>Quantitative Biology Unit, Luxembourg Institute of Health, 1A-B, rue Thomas Edison, 1445 Strassen, Luxembourg. <sup>4</sup>University of Luxembourg, 2, avenue de l'Université, 4365 Esch-sur-Alzette, Luxembourg. ✉email: guy.fagherazzi@lih.lu

Cluster analyses are useful approaches to identify subgroups with different cardiometabolic profiles. Such an approach has recently been developed among people with diabetes, the analysis revealing 5 subgroups with different clinical profiles and risks of diabetes-related complications<sup>11</sup>, but has never been investigated in the general population at large scale<sup>12</sup>. Besides, clustering approaches used in the literature so far were mostly unsupervised where it is assumed that there is no outcome variable nor is anything known about the relationships between the observations in the data set, which is not a reliable hypothesis with respect to cardiometabolic prevention. Semi-supervised clustering techniques may therefore be more adapted to derive meaningful groups<sup>13</sup>, similarly to what has been recently suggested in people with type 1 diabetes<sup>14</sup>, to redefine the way we consider, prevent and treat cardiometabolic diseases in the general population, not as independent entities but rather with a more comprehensive, patient-centered, approach.

Therefore, based on the unique set of cardiometabolic data available in the nationwide population-based ORISCAV-LUX 2 study, our objective was to stratify the general population in terms of cardiometabolic profiles with a high level of granularity, guided by two key factors to assess cardiometabolic health, namely (1) body mass index (BMI), the most frequently used indicator to evaluate adiposity in large populations and an established risk factor of numerous cardiometabolic disorders, highly correlated with various cardiometabolic and cardiovascular risk factor and (2) glycated hemoglobin (HbA1c), a reliable and documented biomarker of glycemic control that is also correlated with many cardiometabolic conditions and surrogate markers<sup>15–18</sup>. This new clustering will help to have a better understanding of the cardiometabolic health of the general population and might eventually help to tailor and target early prevention strategies to people who would benefit the most, thereby representing a first step towards precision prevention for cardiometabolic diseases.

## Materials and methods

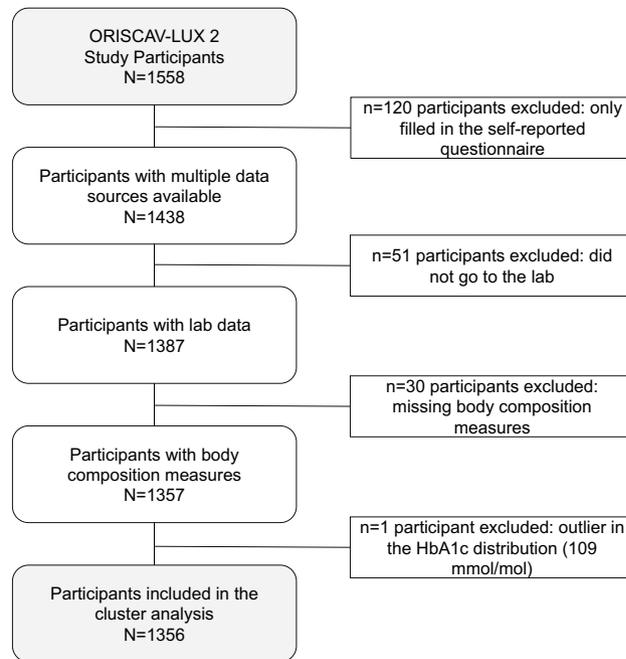
**ORISCAV-LUX 2 study.** The “Observation of cardiovascular risk factors in Luxembourg” (ORISCAV-LUX) 2 is the second wave of the nationwide cross-sectional, population-based ORISCAV-LUX study. The ORISCAV-LUX 1 survey, conducted between November 2007 and January 2009, was the first nationwide cross-sectional survey of cardiovascular health monitoring in Luxembourg with the objective of describing baseline information on the prevalence of “traditional” cardiovascular risk factors, including obesity, hypertension, diabetes mellitus, lipid disorders, smoking and physical inactivity among the general adult population in Luxembourg<sup>19</sup>. The second wave of ORISCAV-LUX was initiated in 2016 to update and monitor the evolution of cardiometabolic parameters in the general population. An extended set of health indicators, new clinical examinations and self-reported information were then integrated in this second round of data collection. The data collection workflow has already been detailed extensively elsewhere<sup>20</sup>. Informed consent was obtained from all participants. The study design and information collected were approved by the National Research Ethics Committee (CNER, No 201,505/12) and the National Commission for Private Data Protection (CNPDP). All methods were carried out in accordance with the Declaration of Helsinki, 2008.

**Study population.** We included participants from the second wave of the ORISCAV-LUX study (2016–2018), where more detailed information on cardiometabolic health was available. We initially included 1558 participants, then excluded participants who only filled in the self-administrated questionnaire (n = 120), did not get a lab test (n = 51), with no body composition measures available (n = 30) and an outlier in the HbA1c distribution (HbA1c = 109 mmol/mol, n = 1). Therefore, we finally considered N = 1356 participants in the present analysis (see flow chart, Fig. 1).

**Clinical and laboratory data assessment.** HbA1c was measured on an HPLC analyser, Tosof G8™. Heart rate, pulse wave velocity, central pressure, arterial age, lying position blood pressure were measured with Complior™. ECG were read and interpreted by a cardiologist and then categorized as normal or abnormal. Bio-impedanciometry measures of body fat percentage in the trunk, muscle mass in the trunk, total fat and fat free mass in the trunk were assessed with a Tanita™ digital scale. Insulin was measured on Abbott immunology analyser (chemiluminescence technique). Insulin resistance was assessed with the HOMA-IR index, calculated as  $\text{Insulin (mIU/l)} \times \text{Glucose (mmol/l)} / 22.5$ . Insulin sensitivity was estimated with the Quicki index, calculated as  $1 / [\log (\text{Insulin, mIU/l}) + \log (\text{Glucose, mg/dl})]$ . Glomerular filtration Rate was estimated with the MDRD formula.

**Cluster analysis.** We performed a semi-supervised cluster analysis guided by BMI and HbA1c to identify subgroups of interest<sup>13</sup>. Five measures, i.e. the means and variances of BMI and HbA1c, as well as the covariance between BMI and HbA1c, were predicted for each individual using reinforcement learning trees (RLT), a type of tree-based machine learning technique<sup>21</sup>. The five clustering variables (RLT-predicted means and variances of BMI and HbA1c and their covariance) were standardized and a k-means clustering algorithm<sup>22</sup> with Euclidean distance was applied. Clustering was tested with and without taking the covariance between BMI and HbA1c into account.

A set of 51 cardiometabolic factors was available in ORISCAV-LUX 2. The factors of body fat and muscle mass from different body parts were highly correlated (pearson coefficient > 0.95), so we only kept the body fat and muscle mass from the trunk for further analysis to increase clustering stability. Overall, we used a subset of 31 factors to be included in the cluster analysis (the remaining factors were only used a posteriori for illustrative purposes, see Table 1). RLT prediction was performed based on the following set of cardiometabolic factors: demographic (age and sex), clinical (ECG interpretation, heart rate, carotid-femoral pulse wave velocity, central pressure, arterial age, defined as the average age for a given carotid-femoral pulse wave velocity<sup>23</sup>, lying position blood pressure), anthropometric (waist circumference, hip circumference, thigh circumference, waist to hip ratio,



**Figure 1.** Flow-chart. Total participants included in the study (ORISCAV-LUX 2 study, N = 1356).

anthropometrically predicted visceral adiposity<sup>22</sup>, body fat percentage in the trunk, muscle mass in the trunk, total fat and fat free mass in the trunk), and laboratory (insulin, insulin resistance, insulin sensitivity, glomerular filtration rate, creatinine, total cholesterol, LDL cholesterol, HDL-cholesterol, triglycerides, CRP) measures. A missing at random mechanism was assumed and missing values were imputed using multiple imputation by chained equations (mice R package<sup>24</sup>).

Clustering stability was assessed using clusterboot function from the fpc R package. The data is resampled 100 times using bootstrap and the Jaccard similarities<sup>25</sup> of the original clusters to the most similar clusters in the resampled data are computed. The mean over these similarities is used as an index of the stability of a cluster. The assessment was applied to the clustering with the number of clusters from 3 and 8. We chose the clustering with the highest mean Jaccard similarity index of the clusters and the smallest cluster greater than 20 participants. Clusters were ordered by increasing HbA1c median. Each cluster was then described according to the variables used for the clustering, but also with additional illustrative variables: lifestyle factors (physical activity assessed with the International Physical Activity (IPAQ) questionnaire, time spent in seated position and smoking status categorized into never, former and current smoker), equivalised disposable income, sedentary occupation and other health factors such as self-perceived health (five categories from excellent to poor), family history of diabetes, hypertension, hypercholesterolemia and personal history of diabetes, cancer and hypertension.

Data are presented in Table 1 as n [%] and median [min, max] for categorical and continuous variables, respectively in the entire population. In Table 2, study participants' characteristics are displayed according to their clusters. In Table 2, we also computed the average 10-year cardiovascular risk [%] per cluster, based on either the SCORE<sup>26</sup> (validated for people < 70 years and no previous cardiovascular disease or type 2 diabetes mellitus) or the ADVANCE<sup>27</sup> (validated for people with type 2 diabetes) risk score, whichever was most appropriate. We used the median values of the continuous variables, and considered that the binary variables were present if more than 50% of the cluster were concerned. In Fig. 2, a scatter plot of body mass index and HbA1c distribution was computed and stratified by cluster group. In Fig. 3, we have plotted the distribution of the clusters in radar diagrams according to 35 key characteristics grouped in 5 themes (Diabetes-related factors, Anthropometry, Lipids & Biomarkers, Cardiovascular Health, Sociodemographic, Lifestyle and other Health Factors). For each feature, we computed the relative difference, expressed in percentage, between the median value (or frequency for categorical variables) in the cluster and the median value (or frequency for categorical variables) in the overall population.

## Results

**Population study characteristics.** The RLT model without taking the covariance between BMI and HbA1C into account provided the most stable clusters. We tested iteratively clustering with  $k=3$  to 8 and we defined the final number of clusters as the one which maximized the stability index while ensuring a sufficient number of individuals in each group, with at least 20 individuals. Therefore, the optimal number of clusters appeared to be 4 and the analysis revealed a very high level of stability, with Jaccard similarity index values of 100%, 100%, 94% and 92% for clusters 1, 2, 3 and 4 respectively (Table 1). Based on the extensive description of characteristics of individuals in each cluster, Cluster 1 was labeled “Healthy”, Cluster 2 was labeled “Family his-

Label	Overall population
N (%)	1356 (100%)
Variables	Median [1st–3rd Quartiles] or n [%]
<b>Sociodemographic, lifestyle and other health factors</b>	
Sex (female, %)*	709 [52.29%]
Age (years)*	51.11 [41.94–60.15]
Equivalised disposable income (€/month)	3571.43 [2625.00–5000.00]
Sedentary occupation (% yes)	755 [55.68%]
Total physical activity (MET-minute/week)	3492.00 [1779.75–6084.00]
Time spent sitting (mn/day)	360.00 [210.00–480.00]
Smoking status (never smoker, %)	806 [59.44%]
Vigorous physical activity (MET-minute/week)	960.00 [0.00–2880.00]
Moderate physical activity (MET-minute/week)	720.00 [200.00–1680.00]
Walking (MET-minute/week)	990.00 [396.00–2079.00]
Personal history of cancer (%yes)	54 [3.98%]
Self-perceived health (1 = excellent, 5 = poor)	3.00 [2.00–3.00]
<b>Diabetes-related factors</b>	
HbA1c (mmol/mol)	36.00 [33.00–39.00]
Fasting blood glucose (mg/dl)*	89.00 [83.00–96.00]
HOMA-IR Index*	1.56 [1.10–2.34]
Insulin ( $\mu$ IU/ml)*	7.10 [5.10–9.90]
Quicki—insulin sensitivity index*	0.36 [0.34–0.38]
Family history of diabetes (%yes)	296 [21.83%]
Personal history of diabetes (%yes)	72 [5.31%]
<b>Cardiovascular health</b>	
Vascular age (years)*	47.00 [37.00–58.00]
Central pulse pressure (mmHG)*	39.00 [33.00–48.00]
Carotid-femoral pulse wave velocity (m/s)*	7.90 [6.90–9.20]
Resting heart rate (bpm)*	58.00 [52.00–64.00]
ECG reading (abnormal, %)*	192 [14.16%]
Systolic blood pressure in lying position (mmHg)*	124.50 [114.50–135.50]
Diastolic blood pressure in lying position (mmHg)*	77.00 [71.00–83.50]
Family history of hypertension (%yes)	575 [42.40%]
Personal history of hypertension (%yes)	468 [34.51%]
Family history of stroke before the age of 45 (%yes)	29 [2.14%]
Family history of myocardial infarction (%yes)	144 [10.62%]
<b>Lipids and biomarkers</b>	
Total cholesterol (mg/dl)*	202.00 [179.00–229.00]
HDL cholesterol (mg/dl)*	56.00 [47.00–66.00]
LDL cholesterol (mg/dl)*	124.00 [102.20–148.00]
Triglycerides (mg/dl)*	88.00 [66.00–123.00]
GFR—MDRD formula (ml/min/1.73 m <sup>2</sup> )*	83.31 [75.29–92.21]
CRP (mg/l)*	1.14 [1.00–2.39]
Family history of high cholesterol (%yes)	561 [41.37%]
Creatinine (mg/dl)*	0.82 [0.75–0.93]
Body mass index (kg/m <sup>2</sup> )	25.45 [22.94–28.78]
Waist to hip ratio*	0.89 [0.82–0.95]
Thigh circumference (cm)*	57.90 [54.40–62.00]
Hip circumference (cm)*	100.10 [95.20–106.20]
Waist circumference (cm)*	88.70 [80.60–98.00]
<b>Anthropometry</b>	
Anthropometrically predicted visceral adiposity (cm <sup>2</sup> )*	8.00 [5.00–11.00]
Total fat mass percentage (%)	27.90 [22.20–34.40]
Total fat mass (kg)	20.50 [15.47–26.42]
Total free fat mass (kg)	51.90 [44.08–64.00]
Fat mass in the left arm (kg)	1.00 [0.70–1.40]
Fat mass percentage in the left arm (%)	25.20 [19.50–33.30]
Continued	

Label	Overall population
Fat free mass in the left arm (kg)	2.80 [2.20–3.70]
Predicted muscle mass in the left arm (kg)	2.60 [2.10–3.50]
Fat mass in the left leg (kg)	3.60 [2.60–4.70]
Fat mass percentage in the left leg (%)	29.40 [20.28–38.50]
Fat free mass in the left leg (kg)	8.60 [7.20–10.60]
Predicted muscle mass in the left leg (kg)	8.10 [6.80–10.10]
Fat mass in the right arm (kg)	0.90 [0.70–1.30]
Fat mass percentage in the right arm (%)	24.20 [18.60–32.30]
Fat free mass in the right arm (kg)	2.80 [2.20–3.70]
Predicted muscle mass in the right arm (kg)	2.60 [2.10–3.50]
Fat mass in the right leg (kg)	3.60 [2.60–4.80]
Fat mass percentage in the right leg (%)	29.95 [20.30–38.32]
Fat free mass in the right leg (kg)	8.60 [7.30–10.80]
Predicted muscle mass in the right leg (kg)	8.20 [6.90–10.20]
Fat mass in the trunk (kg)*	11.50 [8.20–15.03]
Fat mass percentage in the trunk (%)*	28.20 [22.10–33.60]
Fat free mass in the trunk (kg)*	29.00 [25.00–35.20]
Predicted muscle mass in the trunk (kg)*	27.80 [23.98–33.90]

**Table 1.** Study characteristics in the overall population. (ORISCAV-LUX 2 study, N = 1356). \*Features used in the cluster analysis.

tory—Overweight—High Cholesterol”, Cluster 3 was labeled “Severe Obesity—Prediabetes—Inflammation” and Cluster 4 was labeled “Diabetes—Hypertension—Poor CV Health”.

Cluster 1 “Healthy” encompassed a total of N = 729 participants (53.76% of the total population). Compared to the overall population (Table 1), members of Cluster 1 were characterized by young individuals (median,  $m = 46.69$  years old) with a low median HbA1c level ( $m = 34.00$  mmol/mol) and low BMI ( $m = 23.36$  kg/m<sup>2</sup>) (Fig. 2). They also had the lowest values for anthropometric features such as waist-to-hip ratio ( $m = 0.85$ ), fat mass percentage ( $m = 24.30\%$ ) or predicted visceral adiposity ( $m = 6.00$  cm<sup>2</sup>). In terms of lipids and biomarkers, they had the highest level of HDL cholesterol ( $m = 60.00$  mg/dl), a high percentage of family history of hypercholesterolemia (42.39%) and the best renal function (GFR = 84.88 ml/min/1.73 m<sup>2</sup>). Regarding diabetes-related factors, Cluster 1 members had the lowest values for fasting blood glucose ( $m = 86.00$  mg/dl), diabetes diagnosis (1.10%) and HOMA-IR ( $m = 1.24$ ). Oppositely, they had the highest insulin sensitivity (Quicki index  $m = 0.37$ ). Cluster 1 can be considered as the healthiest cluster in terms of cardiovascular health, as they had the lowest values of vascular age ( $m = 43.00$  years), central pulse pressure ( $m = 38.00$  mmHg), pulse wave velocity ( $m = 7.50$  m/s), abnormal ECG reading (10.70%), and systolic blood pressure ( $m = 120.00$  mmHg). Finally, they were also more frequently non-smokers ( $m = 62.83\%$ ), had higher income (3750.00 €/month) and had a higher median time spent sitting ( $m = 360.00$  min/day) and sedentary occupation ( $m = 59.26\%$ ) (Table 1, Fig. 3). The average 10-year cardiovascular risk for Cluster 1 was 0%.

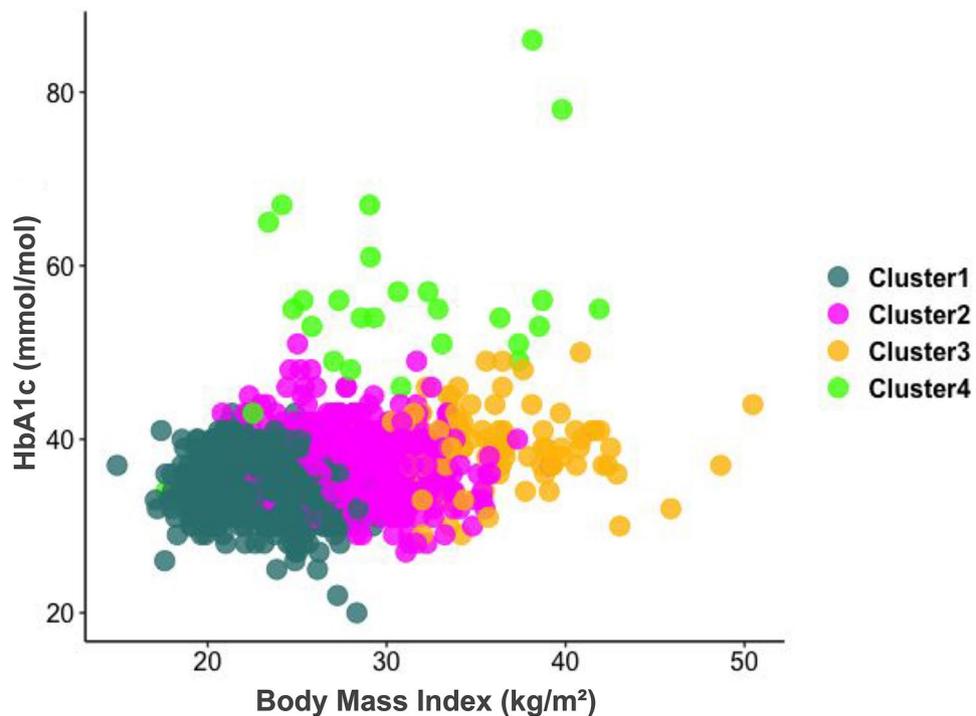
Cluster 2 “Family history—Overweight—High Cholesterol” encompassed N = 508 participants (37.46% of the total population). Members of Cluster 2 were in the vast majority overweight ( $m = 28.48$  kg/m<sup>2</sup>) with low values of HbA1c levels ( $m = 37.00$  mmol/mol). Overall, they had intermediate values for all considered anthropometric features. They were characterized by elevated total ( $m = 205.00$  mg/dl) and LDL cholesterol levels ( $m = 128.50$  mg/dl). They also had a high frequency of family history of diabetes (25.00%) and a high percentage of family history of high blood pressure (43.70%). The average 10-year cardiovascular risk for Cluster 2 was 2%.

Cluster 3 “Severe Obesity—Prediabetes—Inflammation” encompassed N = 91 participants (6.71% of the total population). Cluster 3 included individuals with obesity or severe obesity with a higher BMI ( $m = 35.69$  kg/m<sup>2</sup>) and a higher HbA1c level ( $m = 39.00$  mmol/mol) than those in Cluster 2. Cluster 3 was characterized by the highest values for all considered anthropometric features—except waist-to-hip ratio—with elevated waist circumference ( $m = 114.00$  cm), hip circumference ( $m = 118.30$  cm) or fat mass percentage ( $m = 42.20\%$ ). They had the highest level of inflammation, based on CRP levels ( $m = 3.03$  mg/l). Cluster 3 members had intermediate values for all diabetes-related factors and cardiovascular health factors. There was an over-representation of women in Cluster 3 (61.54%), with a relatively high level of physical activity (3558.00 MET-minutes/week). The average 10-year cardiovascular risk for Cluster 3 was 1%.

Cluster 4 “Diabetes—Hypertension—Poor Cardiovascular Health” encompassed N = 28 participants (2.06% of the population). Members of Cluster 4 were mainly overweight and individuals with obesity (BMI,  $m = 29.20$  kg/m<sup>2</sup>) with elevated HbA1c levels ( $m = 54.50$  mmol/mol). Cluster 4 is characterized by elevated Waist-to-Hip ratio ( $m = 1.00$ ). Members of Cluster 4 had the highest triglycerides levels of all ( $m = 149.00$  mg/dl). Regarding diabetes-related factors, most of Cluster 4 members had diabetes (89.29%), they had the highest levels of fasting blood glucose ( $m = 149.50$  mg/dl), insulin levels ( $m = 16.05$  µIU/ml) and insulin resistance (HOMA-IR,  $m = 6.54$ ). Most of them had hypertension (89.29%) and had the highest values for vascular age ( $m = 61.00$  years), central pulse pressure ( $m = 43.00$  mmHg), pulse wave velocity ( $m = 9.55$  m/s), systolic blood pressure ( $m = 136.50$  mmHg)

Label		Cluster 1 Healthy	Cluster 2 Family history - Overweight - High Cholesterol	Cluster 3 Severe Obesity - Prediabetes - Inflammation	Cluster 4 Diabetes - HTA - Poor CV Health
N (%)		729 (53.76%)	508 (37.46%)	91 (6.71%)	28 (2.06)
Cluster Stability (% of Recovery)		100%	100%	94%	92%
10 year Cardiovascular Risk #		0%	2%	1%	15%
Variables					
Sociodemographic, Lifestyle and other Health Factors	Sex (Female, %)*	414 [56.79%]	227 [44.69%]	56 [61.54%]	12 [42.86%]
	Age (years)*	46.69 [39.03-55.18]	55.94 [47.36-63.71]	56.05 [47.49-63.15]	63.24 [57.84-67.58]
	Equivalised Disposable Income (€/month)	3750.00 [2666.67-5000.00]	3571.43 [2625.00-5000.00]	3000.00 [2000.00-5000.00]	3004.81 [2171.47-5000.00]
	Sedentary occupation (% Yes)	432 [59.26%]	263 [51.77%]	47 [51.65%]	13 [46.43%]
	Total Physical Activity (MET-Minute/week)	3493.00 [1780.00-6078.00]	3492.00 [1737.00-6102.00]	3558.00 [2042.00-6064.00]	3246.00 [2173.50-4391.25]
	Time Spent Sitting (mn/day)	360.00 [240.00-480.00]	300.00 [180.00-480.00]	360.00 [180.00-480.00]	360.00 [262.50-480.00]
	Smoking Status (Never smoker, %)	458 [62.83%]	280 [55.12%]	55 [60.44%]	13 [46.43%]
	Vigorous Physical Activity (MET-Minute/week)	960.00 [160.00-2880.00]	960.00 [0.00-2800.00]	960.00 [0.00-2400.00]	700.00 [0.00-2220.00]
	Moderate Physical Activity (MET-Minute/week)	600.00 [160.00-1680.00]	720.00 [240.00-1680.00]	720.00 [240.00-2160.00]	690.00 [120.00-1200.00]
	Walking (MET-Minute/week)	990.00 [346.50-2079.00]	1039.50 [396.00-2079.00]	1188.00 [396.00-2772.00]	1014.75 [486.75-2376.00]
	Personal History of Cancer (% Yes)	20 [2.74%]	29 [5.71%]	5 [5.49%]	0 [0%]
	Self-Perceived Health (1=Excellent, 5=Poor)	3.00 [2.00-3.00]	3.00 [3.00-3.00]	3.00 [3.00-3.00]	3.00 [3.00-3.25]
Diabetes-related factors	HbA1c (mmol/mol)	34.00 [32.00-37.00]	37.00 [34.00-40.00]	39.00 [37.00-41.00]	54.50 [50.50-57.00]
	Fasting Blood Glucose (mg/dL)*	86.00 [81.00-91.00]	92.00 [87.00-99.00]	97.00 [90.00-104.00]	149.50 [134.00-164.50]
	HOMA-IR Index*	1.24 [0.95-1.68]	2.01 [1.42-2.84]	3.41 [2.44-4.86]	6.54 [4.29-13.34]
	Insulin (µU/mL)*	5.80 [4.50-7.50]	8.55 [6.40-11.80]	13.40 [10.10-18.75]	16.05 [10.95-34.47]
	Quicki - Insulin Sensitivity Index*	0.37 [0.35-0.39]	0.34 [0.33-0.36]	0.32 [0.30-0.33]	0.29 [0.26-0.31]
	Family History of Diabetes (% Yes)	140 [19.20%]	127 [25.00%]	22 [24.18%]	7 [25.00%]
	Personal History of Diabetes (% Yes)	8 [1.10%]	22 [4.33%]	17 [18.68%]	25 [89.29%]
	Cardiovascular Health	Vascular Age (years)*	43.00 [34.00-54.00]	52.00 [41.00-63.00]	52.00 [40.50-63.00]
Central Pulse Pressure (mmHG)*		38.00 [32.00-45.00]	40.00 [34.00-51.00]	39.00 [33.00-48.00]	43.00 [37.75-53.75]
Carotid-Femoral Pulse Wave Velocity (m/s)*		7.50 [6.80-8.60]	8.40 [7.30-9.83]	8.30 [7.35-10.05]	9.65 [8.55-10.70]
Resting Heart Rate (bpm)*		57.00 [51.00-62.00]	58.00 [52.00-65.00]	64.00 [57.00-71.00]	63.50 [57.75-67.25]
ECG Reading (Abnormal, %)*		78 [10.70%]	87 [17.13%]	18 [19.78%]	9 [32.14%]
Systolic Blood Pressure in lying position (mmHg)*		120.00 [111.50-129.50]	129.00 [119.50-139.50]	133.50 [123.00-143.00]	136.50 [127.50-141.75]
Diastolic Blood Pressure in lying position (mmHg) *		75.00 [69.00-81.50]	79.00 [73.50-85.50]	82.00 [77.00-87.25]	83.00 [79.38-86.62]
Family History of Hypertension (% Yes)		304 [41.70%]	222 [43.70%]	38 [41.76%]	11 [39.29%]
Personal History of Hypertension (% Yes)		150 [20.58%]	233 [45.87%]	60 [65.93%]	25 [89.29%]
Family History of Stroke Before the Age of 45 (% Yes)		14 [1.92%]	12 [2.36%]	3 [3.30%]	0 [0.00%]
Family History of Myocardial Infarction (% Yes)	62 [8.50%]	66 [12.99%]	13 [14.29%]	3 [10.71%]	
Lipids & Biomarkers	Total Cholesterol (mg/dL)*	201.00 [178.00-224.00]	205.00 [185.00-234.25]	196.00 [172.50-227.50]	190.00 [143.00-225.25]
	HDL Cholesterol (mg/dL)*	60.00 [51.00-69.00]	53.00 [44.00-62.00]	51.00 [42.50-61.50]	45.00 [41.00-54.25]
	LDL Cholesterol (mg/dL)*	122.00 [101.00-144.40]	128.50 [108.60-154.40]	123.40 [96.40-149.00]	102.40 [66.00-132.00]
	Triglycerides (mg/dL)*	76.00 [59.00-103.00]	101.00 [77.00-140.00]	113.00 [87.50-158.50]	149.00 [80.75-247.50]
	GFR - MDRD formula (mL/min/1.73m <sup>2</sup> )*	84.88 [77.47-93.22]	82.03 [73.95-91.61]	78.30 [71.61-89.10]	75.03 [67.34-80.73]
	CRP (mg/L)*	1.00 [1.00-1.55]	1.59 [1.00-2.82]	3.03 [1.48-5.52]	2.02 [1.25-4.36]
	Family History of High Cholesterol (% Yes)	309 [42.39%]	207 [40.75%]	35 [38.46%]	10 [35.71%]
	Creatinine (mg/dL)*	0.81 [0.74-0.92]	0.84 [0.76-0.95]	0.82 [0.76-0.92]	0.89 [0.80-1.01]
Anthropometry	Body Mass Index (kg/m <sup>2</sup> )	23.36 [21.44-24.86]	28.48 [26.70-30.50]	35.69 [33.40-38.86]	29.20 [25.70-36.59]
	Waist to Hip Ratio*	0.85 [0.79-0.90]	0.94 [0.87-1.00]	0.96 [0.87-1.03]	1.00 [0.92-1.07]
	Thigh Circumference (cm)*	56.00 [53.20-58.50]	61.00 [57.30-64.00]	68.00 [63.35-73.40]	58.50 [53.88-61.97]
	Hip Circumference (cm)*	96.50 [93.00-100.00]	104.60 [100.30-109.00]	118.30 [113.60-123.70]	103.80 [97.83-113.70]
	Waist Circumference (cm)*	82.00 [76.00-87.20]	97.00 [91.18-103.60]	114.00 [104.30-121.00]	103.30 [92.05-117.62]
	Anthropometrically Predicted Visceral Adiposity (cm <sup>2</sup> )*	6.00 [4.00-8.00]	11.00 [9.00-14.00]	14.00 [12.00-19.50]	13.50 [9.00-17.75]
	Total Fat Mass Percentage (%)	24.30 [18.70-30.30]	30.95 [25.80-37.52]	42.20 [35.25-45.85]	32.05 [26.32-38.18]
	Total Fat Mass (kg)	16.10 [12.60-19.70]	25.45 [22.10-29.60]	40.70 [36.10-47.55]	25.95 [20.38-37.45]
	Total Free Fat Mass (kg)	47.40 [43.10-60.40]	56.80 [45.70-66.95]	55.70 [51.25-72.25]	60.40 [45.58-68.38]
	Fat Mass in the Left Arm (kg)	0.80 [0.60-0.90]	1.30 [1.10-1.60]	2.50 [2.00-3.30]	1.20 [0.98-2.00]
	Fat Mass Percentage in the Left Arm (%)	21.40 [16.90-28.50]	28.45 [23.30-37.82]	45.00 [33.15-50.45]	28.70 [22.73-37.70]
	Fat Free Mass in the Left Arm (kg)	2.40 [2.10-3.40]	3.20 [2.40-4.00]	3.20 [2.80-4.40]	3.40 [2.50-4.05]
	Predicted Muscle Mass in the Left Arm (kg)	2.30 [2.00-3.20]	3.00 [2.20-3.73]	3.00 [2.60-4.10]	3.20 [2.38-3.78]
	Fat Mass in the Left Leg (kg)	2.90 [2.10-3.90]	4.20 [3.18-5.30]	7.10 [5.85-8.50]	4.40 [3.25-6.03]
	Fat Mass Percentage in the Left Leg (%)	28.40 [17.20-35.70]	27.75 [22.20-41.73]	45.20 [30.85-48.40]	30.15 [25.57-39.58]
	Fat Free Mass in the Left Leg (kg)	7.70 [7.00-9.90]	9.40 [7.60-11.30]	9.90 [8.60-12.55]	10.25 [7.30-12.00]
	Predicted Muscle Mass in the Left Leg (kg)	7.30 [6.60-9.40]	8.90 [7.10-10.70]	9.40 [8.10-11.90]	9.70 [6.90-11.40]
	Fat Mass in the Right Arm (kg)	1.00 [0.60-0.90]	1.20 [1.00-1.50]	2.30 [1.90-2.90]	1.25 [0.98-2.00]
	Fat Mass Percentage in the Right Arm (%)	20.40 [16.30-27.50]	26.90 [21.90-36.92]	43.90 [30.95-50.00]	29.95 [22.65-36.12]
	Fat Free Mass in the Right Arm (kg)	2.50 [2.20-3.40]	3.20 [2.30-4.00]	3.10 [2.60-4.40]	3.50 [2.40-4.00]
	Predicted Muscle Mass in the Right Arm (kg)	2.30 [2.00-3.20]	3.00 [2.20-3.80]	2.90 [2.50-4.10]	3.30 [2.20-3.80]
	Fat Mass in the Right Leg (kg)	3.00 [2.10-3.90]	4.20 [3.20-5.40]	7.30 [6.15-8.90]	4.75 [3.32-6.10]
	Fat Mass Percentage in the Right Leg (%)	28.30 [16.80-35.60]	28.65 [22.50-41.90]	45.40 [31.35-48.60]	30.15 [25.60-39.73]
	Fat Free Mass in the Right Leg (kg)	7.80 [7.10-10.20]	9.50 [7.70-11.33]	10.00 [8.70-12.70]	10.25 [7.38-11.53]
	Predicted Muscle Mass in the Right Leg (kg)	7.40 [6.70-9.70]	9.00 [7.27-10.72]	9.40 [8.20-12.00]	9.70 [6.97-10.93]
	Fat Mass in the Trunk (kg)*	8.70 [6.70-10.80]	14.60 [12.50-16.72]	22.40 [19.65-25.20]	15.25 [11.15-22.33]
Fat Mass Percentage in the Trunk (%)*	23.10 [18.30-28.30]	31.90 [28.28-35.80]	39.90 [36.95-44.25]	33.20 [25.27-38.00]	
Fat Free Mass in the Trunk (kg)*	27.00 [24.60-33.30]	31.60 [25.78-36.82]	31.20 [28.20-37.90]	33.45 [26.43-36.95]	
Predicted Muscle Mass in the Trunk (kg)*	25.90 [23.50-32.00]	30.40 [24.60-35.40]	29.80 [26.95-36.50]	32.20 [25.25-35.53]	

**Table 2.** Study characteristics by cluster. (ORISCAV-LUX 2 study, N = 1356). \*Features used in the cluster analysis. #Estimated with the SCORE risk score for Clusters 1-3 and with the ADVANCE risk score for Cluster 4.



**Figure 2.** Scatter plot of body mass index and HbA1c distribution (ORISCAV-LUX 2 study, N = 1356).

and percentage of abnormal ECG reading ( $m = 32.14\%$ ). When compared to the overall population, Cluster 4 members were the oldest participants ( $m = 63.24$  years) and had elevated time spent sitting ( $m = 360$  min/day) but the lowest frequency of sedentary occupation ( $m = 46.43\%$ ). The average 10-year cardiovascular risk for Cluster 4 was 15%.

## Discussion

In this large, nationwide population-based study, we have observed 4 stable clusters of individuals from the general population with diverse cardiometabolic health profiles. Our study suggests that this classification could help disentangle the heterogeneity in the general population in terms of cardiometabolic health and be used to tailor prevention strategies. Whereas a first group of more than 50% of the total population (Cluster 1 “Healthy”) was characterized with healthy cardiometabolic features and could benefit from a general prevention strategy, the other 3 groups (Clusters 2–4) may benefit from a more personalized and intensive approach to improve their health. Individuals in Cluster 2 “Family history—Overweight—High Cholesterol” may benefit from a more comprehensive strategy regarding overweight/obesity management and cholesterol with a personalized treatment (e.g. through diet, physical activity, psychology or pharmacological treatment) and starting from an early age for individuals with family history of cardiometabolic diseases. This could delay or prevent them from transitioning from Cluster 2 to Clusters 3 or 4<sup>28</sup>. People in the Cluster 3 “Severe Obesity—Prediabetes—Inflammation” may benefit from an intense lifestyle management strategy adapted to individuals with moderate obesity<sup>29,30</sup>, or bariatric surgery for those with severe obesity<sup>31,32</sup> with a close monitoring of the impact on low-grade inflammation levels and the reverse of prediabetes to a normoglycemic status<sup>33,34</sup>. Cluster 4 “Diabetes—Hypertension—Poor Cardiovascular Health” are often in a multimorbid state, with diabetes and hypertension simultaneously and for a third of them with an abnormal ECG reading or elevated triglyceride levels. Therefore, they could benefit from an intensive combined approach, personalized according to the socioeconomic profile and occupation, with nutritional/dietary<sup>35</sup> or lifestyle<sup>36</sup> interventions, smoking cessation<sup>37</sup>, medication or surgery strategies, targeting both high blood pressure and diabetes with the ultimate objective to reduce arterial stiffness and prevent the occurrence of cardiovascular disease and improve general health status<sup>38,39</sup>.

Overall, these groups may benefit from more efficient prevention and therapeutic strategies. If externally validated, general practitioners could one day rely on this profiling to have a better picture of a new patient when limited information is available and try to optimize several cardiometabolic parameters simultaneously. Some

**Figure 3.** Radar diagrams of the median values for each cluster, according to 35 key diabetes-related factors, anthropometry, lipids and biomarkers, cardiovascular health, sociodemographic, lifestyle and other health factors. *BMI* body mass index, *FMP* fat mass percentage, *VISC ADI* anthropometrically predicted visceral adiposity, *WC* waist circumference, *HC* hip circumference, *TC* thigh circumference, *WHR* waist-to-hip ratio, *CHOL* total cholesterol, *FAM HC* family history of hypercholesterolemia, *CRP* C-reactive protein, *GFR* glomerular filtration rate, *TRIG* triglycerides, *LDL* LDL cholesterol, *HDL* HDL cholesterol, *HbA1c* glycated hemoglobin, *DIABETES* diabetes diagnosis, *FAM DIABETES* family history of diabetes, *QUICKI* quantitative insulin sensitivity check index, *INSULIN* insulin, *HOMA-IR* homeostatic model assessment for insulin resistance, *FBG* fasting blood glucose, *VASC AGE* vascular age, *HTA* hypertension diagnosis, *FAM HBP* family history of high blood pressure, *SBP* systolic blood pressure, *ECG* electrocardiogram, *PWV* pulse wave velocity, *CPP* central pulse pressure, *SEX* sex, *NEVER SMOKER* never smoker, *SITTING* time spent sitting, *PA* physical activity, *INCOME* income, *AGE* age. For each feature, we computed the relative difference, expressed in percentage, between the median value (or frequency for categorical variables) in the cluster and the median value (or frequency for categorical variables) in the overall population.

previous attempts of defining metabolotypes<sup>40</sup>, i.e. metabolomic profiles or combinations of specific metabolites used for classification of individuals into groups have been proposed to advance cardiometabolic prevention<sup>41</sup>. These approaches, along with other recent technologies (big data analysis of gut microbiota, integration of real-time data from wearables), are still complex and not yet cost-effective to implement in practice<sup>42</sup> and our approach could help to fill the gap and help move towards precision cardiometabolic prevention.

These findings are also an opportunity to rethink the strategies that can be offered, for instance to people with obesity<sup>43</sup>, with new models developed according to a more refined definition of the targeted sub-population. Cardiometabolic health relies on complex, intricate, physiological relationships between all the considered parameters in this work. These results imply a move from a “one-size fits-all” vision to a precision cardiometabolic prevention approach to tackle cardiometabolic diseases according to the variety of phenotypes observed in the general population<sup>14</sup>.

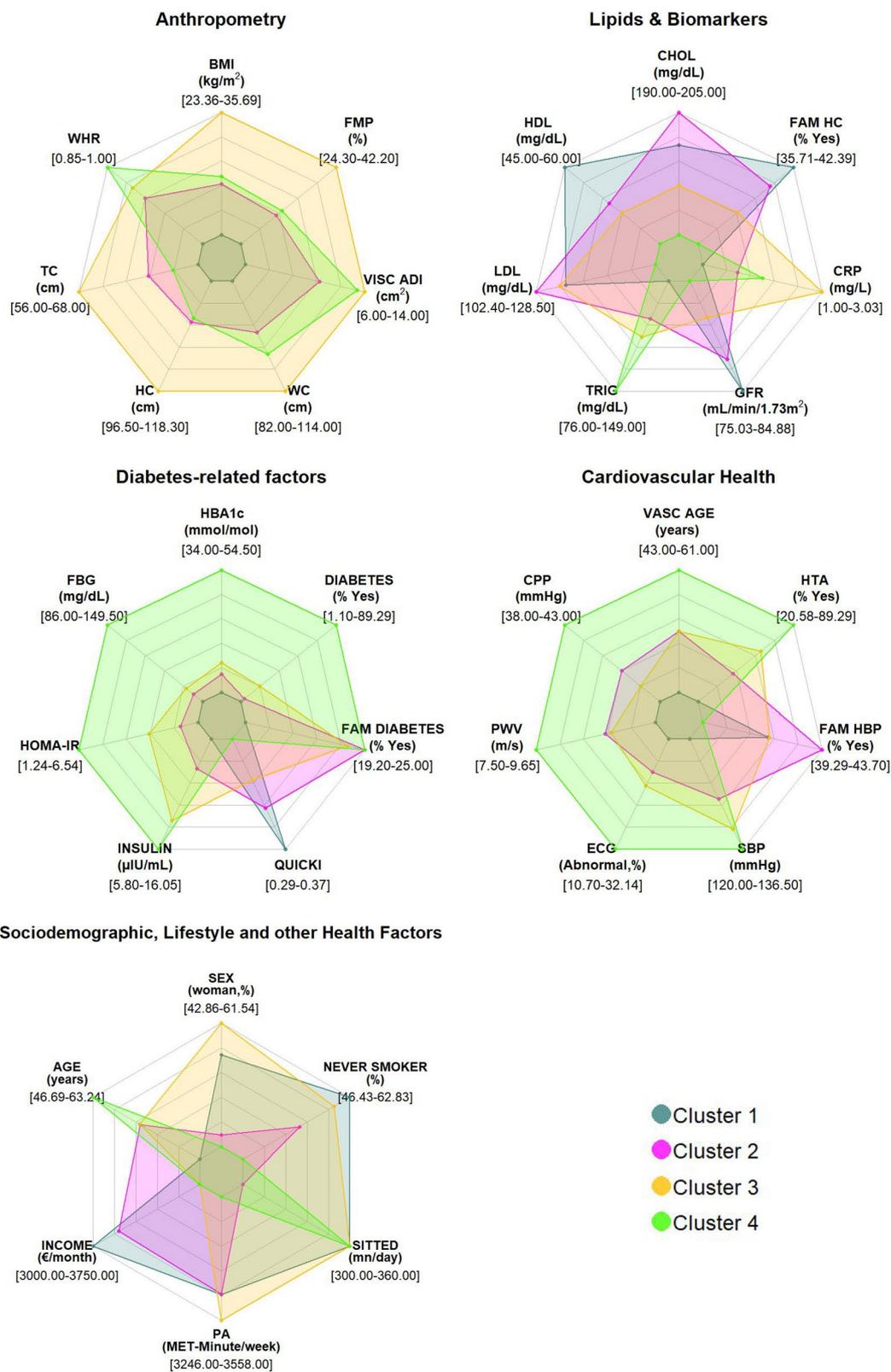
**Strengths and limitations.** This study has numerous strengths. First, the large population size, combined with a unique set of cardiometabolic features or lifestyle and demographic factors, enabled us to extensively and deeply phenotype the general population in terms of cardiometabolic health. It has been shown that the ORISCAV-LUX 2 population was representative of the Luxembourgish adult population in terms of geographical district, but not with respect to sex and age distribution, young and elderly individuals being slightly under-represented and women over-represented. Nonetheless, it has been demonstrated that ORISCAV-LUX 2 is a reliable tool for epidemiological research and for cardiometabolic health monitoring in the adult residents in Luxembourg<sup>20</sup>. We also used a semi-supervised clustering approach, guided by two main features for cardiometabolic health, which seems to be more adapted than totally unsupervised clustering to the reality of the knowledge of cardiometabolic health<sup>13</sup>.

This study also has some limitations. Cluster labelling is always subject to interpretation. We used, to the best of our ability, a systematic approach and relied on the most distinctive characteristics in each cluster to label them. Changing the choice of the key factors to guide the semi-supervised clustering (here BMI and HbA1c) could yield to different distributions, but they were chosen as they are frequently assessed in large populations and valid surrogate of the overall cardiometabolic health status<sup>15–18</sup>. The relatively low number of individuals in clusters 3 and 4 could limit the inference that can be made out of these groups.

Stability of the clusters has been evaluated internally but now there is a need to replicate this approach externally, in other large nationwide population-based studies to evaluate external validation of this grouping. Some factors used to describe the clusters, such as physical activity, are self-reported, and therefore could be reported differently in the clusters. Besides, no mental health nor sleep-related factors were included in the descriptive analysis. In future replication studies, wearable devices could be used to collect objective measures of physical activity and sleep quality, which may be valuable information to add in the cluster description.

## Conclusion

In conclusion, our work provides an in-depth characterization and thus, a better understanding of the general population in terms of cardiometabolic health. Our data suggest that such a clustering approach could now be used to define more targeted and tailored strategies for the prevention of cardiometabolic diseases at a population level. This study provides a first step towards precision cardiometabolic prevention and should be replicated in other contexts. Further studies evaluating the associations between these clusters and subsequent incidence of various cardiometabolic and cardiovascular diseases are warranted.



Received: 14 October 2020; Accepted: 27 July 2021  
 Published online: 06 August 2021

**References**

1. IDF Diabetes Atlas 9th edition 2019. (Accessed 1 July 2021); <https://www.diabetesatlas.org/en/>.
2. Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *Obstetr. Gynecol. Surv.* 58(3), 182–183 (2003).

3. Shubrook, J. H., Chen, W. & Lim, A. Evidence for the prevention of type 2 diabetes mellitus. *J. Am. Osteopath. Assoc.* **118**, 730–737 (2018).
4. Diaz, K. M. & Shimbo, D. Physical activity and the prevention of hypertension. *Curr. Hypertens. Rep.* **15**, 659–668 (2013).
5. Sánchez, A., Silvestre, C., Campo, N., Grandes, G. & PreDE Research Group. Type-2 diabetes primary prevention program implemented in routine primary care: A process evaluation study. *Trials* **17**, 254 (2016).
6. Kivelä, J. *et al.* Obtaining evidence base for the development of Feel4Diabetes intervention to prevent type 2 diabetes—A narrative literature review. *BMC Endocr. Disord.* **20**, 140 (2020).
7. Padilla-Martínez, F., Collin, F., Kwasniewski, M. & Kretowski, A. Systematic review of Polygenic risk scores for type 1 and type 2 diabetes. *Int. J. Mol. Sci.* **21**, 1703 (2020).
8. Cătoi, A. F. *et al.* Metabolically healthy versus unhealthy morbidly obese: Chronic inflammation, nitro-oxidative stress, and insulin resistance. *Nutrients* **10**, 1199 (2018).
9. Samocha-Bonet, D. *et al.* Metabolically healthy and unhealthy obese—The 2013 Stock Conference report. *Obes. Rev.* **15**, 697–708 (2014).
10. Roden, M. & Shulman, G. I. The integrative biology of type 2 diabetes. *Nature* **576**, 51–60 (2019).
11. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
12. Tzeng, C.-R. *et al.* Cluster analysis of cardiovascular and metabolic risk factors in women of reproductive age. *Fertil. Steril.* **101**, 1404–1410 (2014).
13. Bair, E. Semi-supervised clustering methods. *Wiley Interdiscip. Rev. Comput. Stat.* **5**, 349–361 (2013).
14. Kahkoska, A. R. *et al.* Characterizing the weight-glycemia phenotypes of type 1 diabetes in youth and young adulthood. *BMJ Open Diabetes Res. Care* **8**, e000886 (2020).
15. Zorena, K. *et al.* Association between vascular endothelial growth factor and hypertension in children and adolescents type 1 diabetes mellitus. *J. Hum. Hypertens.* **24**, 755–762 (2010).
16. Bower, J. K. *et al.* Glycated hemoglobin and risk of hypertension in the atherosclerosis risk in communities study. *Diabetes Care* **35**, 1031–1037 (2012).
17. Takao, T., Matsuyama, Y., Suka, M., Yanagisawa, H. & Iwamoto, Y. The combined effect of visit-to-visit variability in HbA1c and systolic blood pressure on the incidence of cardiovascular events in patients with type 2 diabetes. *BMJ Open Diabetes Res Care* **3**, e000129 (2015).
18. Huang, T. *et al.* A network analysis of biomarkers for type 2 diabetes. *Diabetes* **68**, 281–290 (2019).
19. Alkerwi, A., Pagny, S., Lair, M.-L., Delagardelle, C. & Beissel, J. Level of unawareness and management of diabetes, hypertension, and dyslipidemia among adults in Luxembourg: findings from ORISCAV-LUX study. *PLoS ONE* **8**, e57920 (2013).
20. Alkerwi, A. *et al.* Challenges and benefits of integrating diverse sampling strategies in the observation of cardiovascular risk factors (ORISCAV-LUX 2) study. *BMC Med. Res. Methodol.* **19**, 27 (2019).
21. Zhu, R., Zeng, D. & Kosorok, M. R. Reinforcement learning trees. *J. Am. Stat. Assoc.* **110**, 1770–1784 (2015).
22. Samouda, H. *et al.* VAT=TAAT-SAAT: Innovative anthropometric model to predict visceral adipose tissue without resort to CT-Scan or DXA. *Obesity* **21**, E41–E50 (2013).
23. Khoshdel, A. R., Thakkinian, A., Carney, S. L. & Attia, J. Estimation of an age-specific reference interval for pulse wave velocity: A meta-analysis. *J. Hypertens.* **24**, 1231–1237 (2006).
24. van Buuren, S. & Groothuis-Oudshoorn, K. Mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
25. Hancock, J. M. Jaccard distance (Jaccard Index, Jaccard similarity coefficient). *Dict. Bioinform. Comput. Biol.* <https://doi.org/10.1002/9780471650126.dob0956> (2004).
26. Motamed, N. *et al.* Comparison of cardiovascular risk assessment tools and their guidelines in evaluation of 10-year CVD risk and preventive recommendations: A population based study. *Int. J. Cardiol.* **228**, 52–57 (2017).
27. Woodward, M. *et al.* Prediction of 10-year vascular risk in patients with diabetes: The AD-ON risk score. *Diabetes Obes. Metab.* **18**, 289–294 (2016).
28. Mutie, P. M., Giordano, G. N. & Franks, P. W. Lifestyle precision medicine: The next generation in type 2 diabetes prevention? *BMC Med.* **15**, 171 (2017).
29. König, D., Hörmann, J., Predel, H.-G. & Berg, A. A 12-month lifestyle intervention program improves body composition and reduces the prevalence of prediabetes in obese patients. *Obes. Facts* **11**, 393–399 (2018).
30. Lv, N. *et al.* Behavioral lifestyle interventions for moderate and severe obesity: A systematic review. *Prev. Med.* **100**, 180–193 (2017).
31. Nguyen, N. T. & Varela, J. E. Bariatric surgery for obesity and metabolic disorders: State of the art. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 160–169 (2017).
32. Sheetz, K. H., Gerhardinger, L., Dimick, J. B. & Waits, S. A. Bariatric Surgery and long-term survival in patients with obesity and end-stage kidney disease. *JAMA Surg.* <https://doi.org/10.1001/jamasurg.2020.0829> (2020).
33. Kerrison, G. *et al.* The effectiveness of lifestyle adaptation for the prevention of prediabetes in adults: A systematic review. *J Diabetes Res.* **2017**, 8493145 (2017).
34. Moutzouri, E., Tsimihodimos, V., Rizos, E. & Elisaf, M. Prediabetes: To treat or not to treat? *Eur. J. Pharmacol.* **672**, 9–19 (2011).
35. Jenkins, D. J. A. *et al.* Effect of a dietary portfolio of cholesterol-lowering foods given at 2 levels of intensity of dietary advice on serum lipids in hyperlipidemia: A randomized controlled trial. *JAMA* **306**, 831–839 (2011).
36. Johansen, M. Y. *et al.* Effect of an intensive lifestyle intervention on glycemic control in patients with type 2 diabetes: A randomized clinical trial. *JAMA* **318**, 637–646 (2017).
37. Li, W. H. C. *et al.* Effectiveness of a brief self-determination theory-based smoking cessation intervention for smokers at Emergency Departments in Hong Kong: A randomized clinical trial. *JAMA Intern. Med.* <https://doi.org/10.1001/jamainternmed.2019.5176> (2019).
38. Ikramuddin, S. *et al.* Lifestyle Intervention and medical management with vs without Roux-en-Y Gastric bypass and control of hemoglobin A1c, LDL cholesterol, and systolic blood pressure at 5 years in the diabetes surgery study. *JAMA* **319**, 266–278 (2018).
39. Byrne, J. L. *et al.* Effectiveness of the ready to reduce risk (3R) complex intervention for the primary prevention of cardiovascular disease: A pragmatic randomised controlled trial. *BMC Med.* **18**, 198 (2020).
40. Riedl, A., Gieger, C., Hauner, H., Daniel, H. & Linseisen, J. Metabotyping and its application in targeted nutrition: An overview. *Br. J. Nutr.* **117**, 1631–1644 (2017).
41. Brennan, L. Use of metabotyping for optimal nutrition. *Curr. Opin. Biotechnol.* **44**, 35–38 (2017).
42. Wang, D. D. & Hu, F. B. Precision nutrition for prevention and management of type 2 diabetes. *Lancet Diabetes Endocrinol.* **6**, 416–426 (2018).
43. Kar, P. Partha Kar: Our approach to tackling obesity needs rethinking. *BMJ* **370**, m3034 (2020).

## Acknowledgements

The authors thank the participants in ORISCAV-LUX 2 study and the research nurses from the Luxembourg Institute of Health for the data collection. The authors are thankful to the ORISCAV-LUX Study group, which is composed of Ala'a Alkerwi, Stephanie Noppe, Charles Delagardelle, Jean Beissel, Anna Chiotti, Saverio Stranges, Jean-Claude Schmit, Marie-Lise Lair, Marylène D'Incau, Jessica Pastore, Gwenaëlle Le Coroller, Gloria Aguayo,

Brice Appenzeller, Laurent Malisoux, Sophie Couffignal, Manon Gantenbein, Yvan Devaux, Michel Vaillant, Laetitia Huiart, Dritan Bejko, Torsten Bohn, Hanen Samouda, Guy Fagherazzi, Magali Perquin, Maria Ruiz-Castell and Isabelle Ernens.

### Author contributions

G.F. and L.Z. wrote the main manuscript text. L.Z. prepared Fig. 2. G.F. prepared Figs. 1 and 3. All authors reviewed the manuscript.

### Funding

The ORISCAV-LUX 2 study has been funded by the Luxembourg Institute of Health. No role to be declared of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. GF has received consulting fees from Lilly, MSD, Roche Diabetes Care, AstraZeneca, Diabeloop and Danone Research.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to G.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021