Received: 18 May 2020 | Revised: 15 June 2020 | Accepted: 27 June 2020

DOI: 10.1111/jch.13984

ORIGINAL PAPER

WILEY

Unsupervised hierarchical clustering identifies a metabolically challenged subgroup of hypertensive individuals

Felix C. Vaura MA[1] | Veikko V. Salomaa MD, PhD[2] | Ilkka M. Kantola MD, PhD[3] | Risto Kaaja MD, PhD[1,3] | Leo Lahti DSc[4] | Teemu J. Niiranen MD, PhD[1,2,3]

[1]Department of Medicine, University of Turku, Turku, Finland

[2]Finnish Institute for Health and Welfare (THL), Helsinki, Finland

[3]Division of Medicine, Turku University Hospital, Turku, Finland

[4]Department of Future Technologies, University of Turku, Turku, Finland

Correspondence
Felix Vaura, MA, Department of Medicine, University of Turku, Kiinamyllynkatu 10, 20014 Turku, Finland.
Email: fechva@utu.fi

Funding information
V.V.S. was supported by the Finnish Foundation for Cardiovascular Research. L.L. was supported by the Academy of Finland (295741). T.J.N. was supported by the Academy of Finland (321351), Urmas Pekkala Foundation, the Paavo Nurmi Foundation, the Finnish Medical Foundation, and the Emil Aaltonen Foundation.

Abstract

The current classification of hypertension does not reflect the heterogeneity in characteristics or cardiovascular outcomes of hypertensive individuals. Our objective was to identify distinct phenotypes of hypertensive individuals with potentially different cardiovascular risk profiles using data-driven cluster analysis. We performed clustering, a procedure that identifies groups with similar characteristics, in 3726 individuals (mean age 59.4 years, 49% women) with grade 2 hypertension (blood pressure ≥160/100 mmHg or antihypertensive medication) selected from FINRISK 1997, 2002, and 2007 cohorts. We computed clusters based on eight factors associated with hypertension: mean arterial pressure, pulse pressure, non-high-density lipoprotein cholesterol, blood glucose, BMI, C-reactive protein, estimated glomerular filtration rate, and alcohol. After that, we used Cox regression models adjusted for age and sex to assess the relative risk of cardiovascular disease (CVD) outcomes between the clusters and a reference group of 11 020 individuals. We observed two comparable clusters in both men and women. The Metabolically Challenged (MC) cluster was characterized by high blood glucose (*Z*-score 4.4 ± 1.1 vs 0.2 ± 0.8, men; 3.5 ± 1.1 vs 0.0 ± 0.6, women) and elevated BMI (30.4 ± 4.1 vs 28.9 ± 4.3, men; 32.7 ± 4.9 vs 29.3 ± 5.5, women). Over a 10-year follow-up (1034 CVD events), MC had 1.6-fold (95% CI 1.1-2.4) CVD risk compared to non-MC and 2.5-fold (95% CI 1.7-3.7) CVD risk compared to the reference group ($P \leq .009$ for both). Using unsupervised hierarchical clustering, we found two phenotypically distinct hypertension subgroups with different risks of CVD complications. This substratification could be used to design studies that explore the differential effects of antihypertensive therapies among subgroups of hypertensive individuals.

1 | INTRODUCTION

Hypertension is the most important risk contributor to the global burden of disease,[1] accounting for up to 15% of global mortality.[2] Despite considerable advances in hypertension care over the past five decades,[3] the world's 1.4 billion hypertensive individuals are mainly treated under the general umbrella of primary hypertension with little consideration for personalized therapy.

Hypertension guidelines categorize hypertensive patients into subgroups based on a few risk factors that determine their cardiovascular disease (CVD) risk. These risk factors include blood pressure (BP) level and presence of chronic kidney disease, type 2 diabetes, or hypertension-mediated organ damage.[4,5] However, such classifications are based on arbitrary risk factor thresholds instead of the underlying cause of hypertension. Indeed, selecting the correct treatment for hypertensive patients is still commonly based on trial

TABLE 1 Baseline characteristics of the clustering sample stratified by sex (men vs women) and substratified by cluster (non-MC vs MC)

Variable	All men	non-MC	MC	All women	non-MC	MC
N (% of sample)	1910 (51.3)	1848 (49.6)	62 (1.7)	1816 (48.8)	1765 (47.4)	51 (1.4)
Age (y)	59.2 (9.5)	59.1 (9.5)	62.8 (7.3)	59.6 (9.0)	59.6 (9.1)	61.6 (7.8)
Glucose[a] (*Z*-score)	0.3 (1.1)	0.2 (0.8)	4.4 (1.1)	0.1 (0.9)	0.0 (0.6)	3.5 (1.1)
FPG[a] (mmol/L)	6.0 (1.3)	5.9 (0.9)	10.8 (1.3)	5.7 (1.0)	5.6 (0.8)	9.7 (1.3)
HbA1c[a] (mmol/L)	39.3 (7.4)	38.3 (5.2)	68.2 (7.4)	38.8 (6.6)	37.6 (3.8)	62.0 (8.3)
BMI (kg/m^2)	29.0 (4.3)	28.9 (4.3)	30.4 (4.1)	29.4 (5.5)	29.3 (5.5)	32.7 (4.9)
CRP (mg/L)	1.4 (2.3)	1.4 (2.3)	1.6 (3.2)	1.8 (2.8)	1.7 (2.8)	2.2 (2.5)
logCRP	0.4 (1.1)	0.4 (1.1)	0.5 (1.1)	0.6 (1.1)	0.6 (1.1)	0.8 (1.0)
non-HDLc (mmol/L)	4.3 (1.0)	4.3 (1.0)	4.0 (1.3)	4.1 (1.0)	4.1 (1.0)	3.8 (1.0)
MAP (mm Hg)	116.2 (10.7)	116.1 (10.7)	118.2 (10.7)	112.8 (10.6)	113.0 (10.6)	108.3 (9.4)
PP (mm Hg)	69.1 (19.1)	68.9 (19.1)	77.0 (18.9)	72.6 (18.5)	72.6 (18.6)	74.5 (15.9)
eGFR (mL/min/1.73 m^2)	85.8 (15.1)	85.7 (15.2)	86.3 (13.6)	81.7 (15.6)	81.6 (15.5)	85.8 (16.3)
Alcohol (g/wk)	53.8 (146.1)	55.6 (146.6)	17.6 (126.0)	8.0 (35.5)	8.4 (35.5)	2.7 (23.5)
rAlcohol	0.4 (1.0)	0.4 (0.9)	0.0 (1.1)	−0.3 (0.8)	−0.3 (0.8)	−0.4 (0.9)

Note: Summary statistics are given as mean (standard deviation), except for CRP and Alcohol, which are reported as median (interquartile range). The clustering sample consisted of 3726 individuals with grade 2 hypertension from FINRISK 1997 (N = 1098), 2002 (N = 1385), and 2007 (N = 1243) cohorts.

Abbreviations: BMI, body-mass index; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; FPG, fasting plasma glucose; Glucose, standardized *Z*-score of either glycated hemoglobin or fasting glucose; HbA1c, glycated hemoglobin; MAP, mean arterial pressure; MC, the metabolically challenged cluster with higher blood glucose and BMI; non-HDLc, non-high-density lipoprotein cholesterol; non-MC, the non-metabolically challenged cluster with lower blood glucose and BMI; PP, pulse pressure; rAlcohol, rank-normalized alcohol.

[a]For the calculation of Glucose *Z*-score, FPG was used in 2770 individuals, and HbA1c was used in 956 individuals.

and error, and an improved, objective classification of hypertension based on commonly measured clinical variables would benefit both patients and clinicians.

Unsupervised clustering is a machine learning technique that has the potential to classify individuals into subgroups that differ from each other.[6] Previous studies have shown that unsupervised clustering can divide individuals with hypertension into clinically meaningful subgroups. Guo et al[7] (N = 513) and Yang et al[8] (N = 9361) both applied unsupervised clustering on selected samples of individuals with hypertension and demonstrated subgroups with differing baseline characteristics and CVD risk profiles. These studies have laid the groundwork for an improved classification of hypertension but are limited either in sample size or lack of individuals with diabetes.

An improved, pathogenesis-driven classification of hypertension is needed. The aim of this study was to use unsupervised cluster analysis to objectively identify subgroups of hypertensives with different baseline characteristics in the general population and analyze their CVD risk profiles.

2 | METHODS

2.1 | Study sample

The study participants took part in the national FINRISK epidemiological surveys, which have been carried out since 1972 every 5 years to survey risk factors of chronic diseases in Finland.[9] For this study, we considered participants of the 1997 (N = 8444, aged 25-74 years), 2002 (N = 9485, aged 25-74 years), and 2007 (N = 7857, aged 25-74 years) cohorts (total N = 25 786) who were drawn from the population register using stratified random sampling. FINRISK's methods, measurements, and protocols have stayed nearly identical over the years and have been previously described in detail.[9] The study protocols were approved by the Ethical Committee for Epidemiology and Public Health of the Hospital District of Helsinki and Uusimaa or the Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa. All participants gave informed written consent.

We excluded participants with missing covariates (N = 10 184), prevalent CVD (N = 3583), and extreme outliers in any clustering variables (>5 standard deviations from the mean; N = 118). After these exclusions, we included 14 746 participants in the study sample. For clustering, we considered 3726 individuals (Table 1) with grade 2 hypertension, defined as: systolic BP ≥160 mm Hg, diastolic BP ≥100 mm Hg, or the use of antihypertensive medication.[4] We included individuals with grade 2, instead of grade 1, hypertension to increase the specificity of our clustering sample as the prevalence of grade 1 hypertension is approximately 40% in Finns aged >30 years.[10] In addition, the threshold for initiation of antihypertensive therapy was 160/100 mm Hg for most patients until 2014 in Finland,[11] therefore coinciding with the definition of hypertension we used in this study. We used the remaining 11 020 individuals as the reference group for the subsequent survival analysis.

2.2 | Baseline data

Trained nurses measured BP (average of 3 measurements using a mercury sphygmomanometer after at least 5 minutes of rest[9]), height, and weight on-site. We determined plasma creatinine, glycated hemoglobin, fasting blood glucose, C-reactive protein (CRP), total cholesterol, and high-density lipoprotein (HDL) cholesterol from collected blood samples. We assessed alcohol consumption (1-year average of grams of pure alcohol per week) and the use of antihypertensive medication from a self-reported survey. We imputed the BP of individuals with antihypertensive medication by adding 10 mm Hg to SBP and 5 mm Hg to DBP. We defined mean arterial pressure as[12] $1/3 \times \text{SBP} + 2/3 \times \text{DBP}$ and pulse pressure as SBP − DBP. We calculated the estimated glomerular filtration rate (eGFR) with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation[13] and body mass index (BMI) as weight (kg) divided by height (m) squared. We defined non-high-density lipoprotein (non-HDL) cholesterol as total cholesterol minus HDL cholesterol.

2.3 | Follow-up

In Finland, each permanent resident can be linked to nationwide electronic health records that cover all major health events and deaths. We obtained follow-up data for CVD events from the National Hospital Discharge Register and the National Causes of Death Register and used incident fatal and non-fatal CVD as the end point. The National Hospital Discharge Register and the National Causes of Death Register cover[14] all years since 1969. In order to make follow-up comparable between cohorts, we restricted the follow-up time to 10 years. We defined incident CVD events as those after the baseline examination date, and prevalent CVD events as those before or at the baseline examination date. We defined CVD as either coronary heart disease (including myocardial infarction) or stroke (excluding subarachnoid hemorrhage). For details of the definitions, see Appendix S1. The register data have been previously validated for these diagnoses.[15,16]

2.4 | Cluster analysis

We performed unsupervised hierarchical clustering (R function "hclust"; see below for details) on the participants with eight continuous, standardized (mean-centered with unit variance) variables: mean arterial pressure, pulse pressure, alcohol intake, eGFR, BMI, blood glucose, CRP, and non-HDL cholesterol. We chose these variables because they are clinically relevant, well-established risk factors for hypertension and CVD that are routinely measured in clinical practice. We calculated blood glucose as a standardized *Z*-score of either glycated hemoglobin or fasting glucose (when glycated hemoglobin was not available). To achieve balanced variable distributions for clustering, we took the natural logarithm of CRP before standardization to obtain log CRP and rank-normalized[17] (R package *RNOmni*[18]) alcohol intake. We clustered men (N = 1910) and women (N = 1816) separately to avoid stratification due to sex differences in the clustering variables. For clustering, we used Ward's method[19] (Ward2) with Euclidean distance. We determined the number of clusters by maximizing the minimum average silhouette width[20] (R package *cluster*[21]) of clusters and used the resulting clustering as the primary exposure variable for survival analysis. To visualize the high-dimensional clusters in two dimensions, we used principal component analysis (PCA[22], R function "prcomp") on the same standardized variables as in the clustering.

2.5 | Statistical analysis

To assess the association between hypertension clusters and CVD outcomes, we used Cox proportional hazards regression (R package *rms*[23]) with time-on-study as the time-scale. After observing that clusters with similar characteristics could be identified in both men and women, we merged the corresponding clusters from both sexes for the prospective analyses to increase statistical power. We used three models: an unadjusted model, a model adjusted for age and sex, and a model adjusted for age and sex with an interaction term between sex and hypertension cluster. We checked the proportional hazards assumptions using a correlation test based on Schoenfeld residuals[24,25] (R package *survival*[26]). We measured the association between glycated hemoglobin and fasting glucose with Pearson's *r* (R function "cor.test"). All statistical tests were two-sided, and we considered $P < .05$ statistically significant. We used R version 3.6.1 (R Core Team 2019) for all computations.

3 | RESULTS

The study sample characteristics are shown in Table 1. The optimal number of clusters was two for both sexes, and the smaller cluster was characterized by high blood glucose (*Z*-score 4.4 ± 1.1 vs 0.2 ± 0.8 in men; 3.5 ± 1.1 vs 0.0 ± 0.6 vs in women) and elevated BMI (30.4 ± 4.1 vs 28.9 ± 4.3 in men; 32.7 ± 4.9 vs 29.3 ± 5.5 in women). We thus termed the smaller cluster (N = 113) as the metabolically challenged (MC) and the larger cluster (N = 3613) as non-MC. The baseline characteristics of the two clusters were comparable between sexes (Figure 1). The average silhouette widths for non-MC and MC were 0.38 and 0.31, respectively, for men and 0.38 and 0.32, respectively, for women. A separation of the two clusters was visible in the PCA projections (Figure 2). Out of the 113 individuals in the MC cluster, 110 (97%) had metabolic syndrome as defined by the International Diabetes Federation.[27] Pearson's *r* between glycated hemoglobin and fasting glucose in this study was .58 (95% CI 0.53-0.63).

Because the cluster characteristics were comparable between sexes, we used merged clustering results from men and women to assess the relationship between hypertension cluster membership

(MC, non-MC, or reference group) and CVD outcomes. A total of 1034 CVD events occurred during a median follow-up time of 10 years: 30 in the MC cluster, 504 in the non-MC cluster, and 500 in the reference group. In the unadjusted model, membership in the MC cluster was associated with significantly increased CVD risk compared to the non-MC cluster (hazard ratio [HR] 2.0, 95% CI 1.4-2.9, $P < .001$) and to the reference group (HR 6.7, 95% CI 4.7-9.7, $P < .001$). In the age- and sex-adjusted model, the risk for incident CVD remained significantly higher for the MC cluster compared to the non-MC cluster (HR 1.6, 95% CI 1.1-2.4, $P = .009$) and to the reference group (HR 2.5, 95% CI 1.7-3.7, $P < .001$) (Figures 3 and 4). In the age- and sex-adjusted model with a sex × cluster interaction, there appeared to be additional CVD risk for women compared to men, associated with the MC cluster compared to the non-MC cluster (HR 1.3, 95% CI 0.6-2.8, $P = .47$) and to the reference group (HR 1.6, 95% CI 0.7-3.4, $P = .23$) but the effects were not statistically significant. We found no evidence of non-proportional hazards in any of the models (global P values .81, .67 and .78, respectively).

4 | DISCUSSION

The results of our study suggest that a metabolically challenged subgroup of individuals exists in the general hypertensive population, characterized by elevated blood glucose and increased BMI (Figures 1 and 2). This group has an elevated CVD risk (Figures 3 and 4).

Only two prior studies have assessed the unsupervised clustering of hypertension. In order to identify high-risk individuals with hypertension, Yang et al[8] performed unsupervised hierarchical two-step clustering on 9361 hypertensive participants of the randomized, controlled, open-label Systolic Blood Pressure Intervention Trial (SPRINT). The trial compared the effects of intensive (BP target <120 mm Hg) and standard (BP target <140 mm Hg) BP control. Notably, individuals with diabetes were excluded from the trial. According to the authors, clustering variables included all clinically relevant baseline variables in the SPRINT trial relevant to hypertension, including Framingham risk score (FRS) for 10-year CVD risk, BMI, total cholesterol, HDL, SBP and DBP, eGFR, and age. The authors observed four clusters: Cluster 1 had relatively healthy individuals, cluster 2 individuals with slightly decreased eGFR, cluster 3 individuals with the highest BMI, and cluster 4 individuals with the highest FRS for 10-year CVD risk. Yang et al then compared the cumulative incidence of CVD outcomes between the four clusters with the Kaplan-Meier estimator. They defined the primary CVD outcome as a combination of myocardial infarction, acute coronary syndrome, stroke, heart failure, and death from CVD causes. The results showed that cluster 4 with the highest FRS for 10-year CVD risk also had the highest incidence of CVD outcomes, while the CVD risk of the other clusters did not differ from one another. Since all individuals with diabetes were excluded

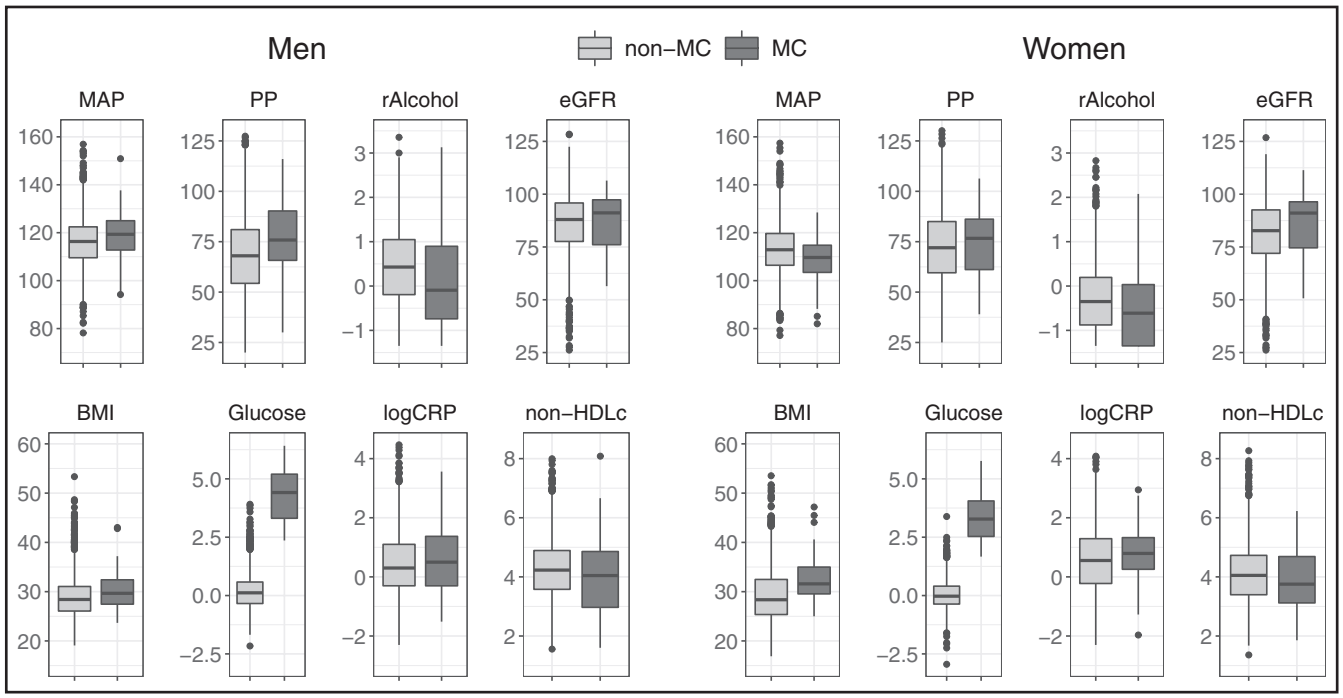

FIGURE 1 Baseline characteristics of MC and non-MC clusters stratified by sex. MC and non-MC clusters consisted of 3726 individuals with grade 2 hypertension from FINRISK 1997, 2002, and 2007 cohorts. BMI, body-mass index (kg/m^2); eGFR, estimated glomerular filtration rate (mL/min/1.73 m^2); Glucose, standardized Z-score of either glycated hemoglobin or fasting glucose; logCRP, natural logarithm of C-reactive protein concentration (mg/L); MAP, mean arterial pressure (mm Hg); MC, the metabolically challenged cluster with higher blood glucose and BMI; non-HDLc, non-high-density lipoprotein cholesterol (mmol/L); non-MC, the non-metabolically challenged cluster with lower blood glucose and BMI; PP, pulse pressure (mm Hg); rAlcohol, rank-normalized alcohol

from the SPRINT trial, it is challenging to compare the results of Yang et al with those from our study, as all our MC individuals had diabetes. In addition, inclusion of both the Framingham risk score and its components among the clustering variables could result in misclassification.

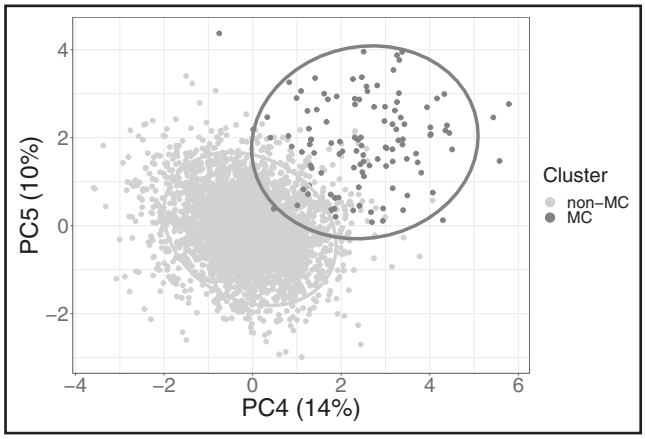

FIGURE 2 Principal component analysis (PCA). The detected hypertension clusters are indicated by color. Positions of 3726 individuals with grade 2 hypertension from FINRISK 1997, 2002, and 2007 cohorts are plotted in a plane defined by principal components PC4 and PC5. We chose to show these components as they produced the highest variance between the two clusters. The ellipses represent regions of highest density for each cluster. MC, the metabolically challenged cluster with higher blood glucose and body-mass index; non-MC, the non-metabolically challenged cluster with lower blood glucose and body-mass index; PC, principal component

Guo et al[7] recruited 513 patients with a mean age of 61 years to explore clinical phenotypes in patients with essential hypertension using cluster analysis. Baseline characteristics included age, sex, prevalence of coronary artery disease (CAD) and cerebral infarction, smoking, diabetes and fasting glucose, carotid plaque thickness, several variables derived from 24-hour ambulatory BP monitoring, and lipids. After reducing baseline data to two dimensions with PCA, Guo et al applied k-means clustering and observed four clusters. Cluster 1 (N = 172) included younger male smokers, cluster 2 (N = 70) older diabetic females, cluster 3 (N = 144) relatively healthy individuals, and cluster 4 (N = 127) all individuals with prevalent CAD. While our study and the study by Guo et al used a somewhat different set of variables, our clustering results also have similarities. The prevalence of diabetes was 100% in our MC cluster and 100% in cluster 2 of Guo et al. However, the MC cluster was 3.0% of our clustering sample, while cluster 2 of Guo et al was 14% of their clustering sample. The differences in study samples could explain some of the discrepancies between the current study and the one by Guo et al: They excluded hypertensives who were under antihypertensive treatment, while 63% of our clustering sample were on antihypertensive medication. Furthermore, 41% of their study sample had had a stroke. In addition, it remains unclear why exactly four clusters were chosen as the optimal number of clusters.

Existing research on the pathophysiology of vascular alterations supports the classification of MC as a hypertension subtype. Hyperglycemia leads to low-grade vascular inflammation that further causes endothelial dysfunction and vascular stiffening.[28] Therefore, the vascular stiffening caused by hypertension itself gets compounded, and the risk for the development and progression of CVD increases. In the Framingham Heart Study,[29] having diabetes

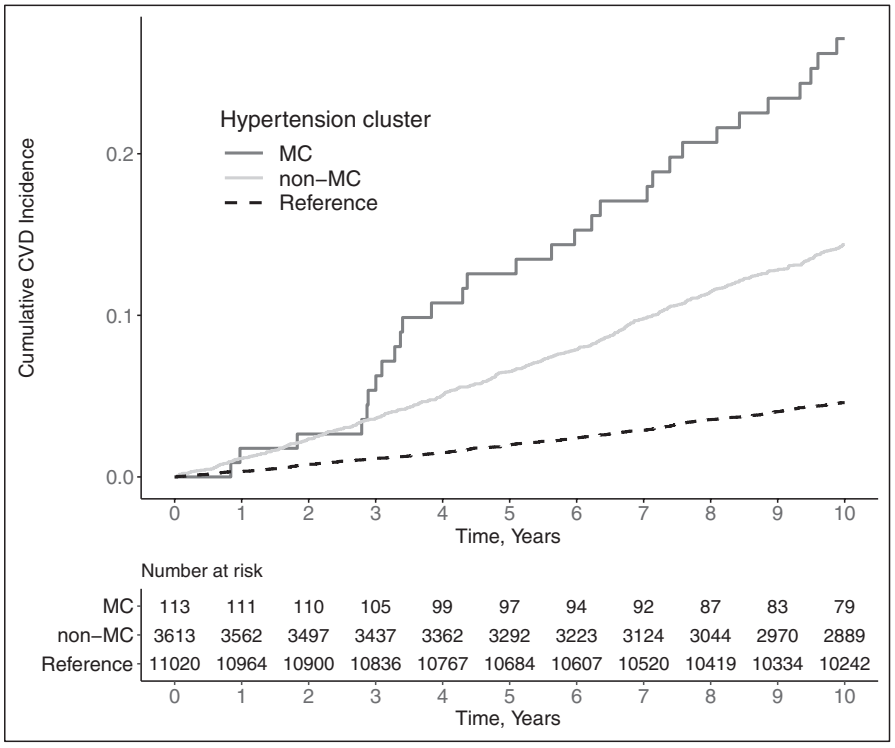

FIGURE 3 Kaplan-Meier plot for cumulative CVD incidence plotted against time in years. *P* value for the difference in survival between clusters was <.0001 (log-rank test). The study sample consisted of 3726 individuals with grade 2 hypertension and 11 020 reference individuals from FINRISK 1997, 2002, and 2007 cohorts. CVD, cardiovascular disease; MC, the metabolically challenged cluster with higher blood glucose and body-mass index; non-MC, the non-metabolically challenged cluster with lower blood glucose and body-mass index

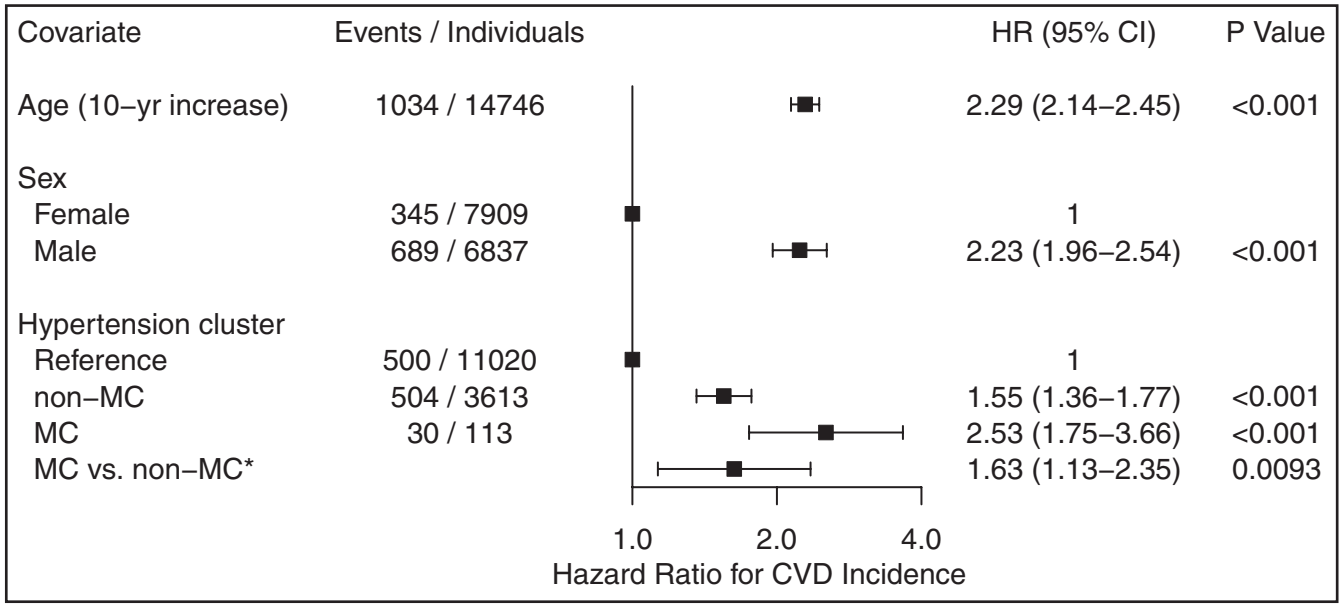

FIGURE 4 Hazard ratios with 95% confidence intervals for CVD incidence associated with age, sex and cluster in the full Cox regression model. The study sample consisted of 3726 individuals with grade 2 hypertension and 11 020 reference individuals from FINRISK 1997, 2002, and 2007 cohorts. CVD, cardiovascular disease; HR, hazard ratio; MC, the metabolically challenged cluster with higher blood glucose and body-mass index; non-MC, the non-metabolically challenged cluster with lower blood glucose and body-mass index. *Here the non-MC cluster was used as the reference group

was related to almost 10-fold lower odds for maintaining healthy vasculature in old age. MC individuals therefore potentially represent a distinct group of high-risk hypertensive individuals that could benefit from aggressive and tailored drug and lifestyle therapies.

The characteristics of the MC cluster resemble those of the metabolic syndrome (MetS). While MetS has many similar definitions,[27,30-32] they all include hyperglycemia, obesity, and elevated blood pressure, all of which are present in the MC cluster. In the USA alone, over a third of the population[33] is estimated to have MetS as defined by the International Diabetes Federation, and it is associated with numerous cardiovascular complications.[34] The 97% prevalence of MetS in the MC cluster indicates that it is likely a strict subset of MetS. The established research on MetS, along with the phenotypic separation of the MC cluster from the rest of the population (Figure 2), further supports the classification of the MC cluster as a hypertension subtype.

Although the stratified random sampling and the quantification of cluster quality are the strengths of our study, it also has several limitations. First, our sample size, although large, is still somewhat limited. A direct consequence of this can be seen in Figure 3, in which the Kaplan-Meier curves fluctuate between years 2 and 4. This is most likely explained by the low number of CVD events in the MC cluster during the 10-year follow-up, which renders the initial follow-up years particularly susceptible to random variation. Second, combining *Z*-scores from glycated hemoglobin and fasting glucose to quantify blood glucose is not ideal, but their moderate[35] correlation of $r = .58$ supports this choice. Third, self-reported alcohol intake might often underestimate true alcohol use. However, self-report is usually the only feasible option for measuring alcohol use and usually ranks the individuals correctly. Similarly, a self-reported survey can misestimate the proportion of individuals on antihypertensive medication. However, we believe self-report to be a good compromise when estimating the true medication use as prescribed antihypertensive drugs are often not purchased.[36] Fourth, while two clusters were optimal, the silhouette widths of <0.5 do not indicate particularly strong clustering,[37] and the separation seen in the PCA projection (Figure 2) is not complete. Fifth, we recognize that in hierarchical clustering the choice of metric (Euclidean), linkage (Ward2), and variable transformations (normalization and standardization) greatly affect the results. However, we find our choices to be the most appropriate for a noisy dataset with continuous variables. Finally, hypertension-specific phenotyping—including direct measurements of the cardiovascular system such as echocardiography, ECG, or assessment of arterial stiffness—could have improved cluster quality.

In conclusion, we used unsupervised hierarchical clustering to uncover a metabolically challenged subgroup of hypertensive individuals in the general population. The subgroup has a high risk for incident CVD and is characterized by high blood glucose and BMI. Stratification of hypertensive patients more strictly by metabolic status could help tailor and target early treatment to patients who would benefit most from it, thereby allowing for a more precision medicine approach. More studies are therefore needed to examine if hypertension care in MC patients should markedly differ from their counterparts in terms of treatment targets and modalities. In addition, more studies focused on hypertension clustering with data on hypertension-specific phenotyping of the cardiovascular

system, and discovering optimal therapies for these clusters, are needed.

ACKNOWLEDGMENTS

We thank the participants of The National FINRISK Study for invaluable contributions to this work.

CONFLICT OF INTEREST

VS has received honoraria from Novo Nordisk and Sanofi for consultations. He also has ongoing research collaboration with Bayer Ltd. (All unrelated to the present study).

AUTHOR CONTRIBUTIONS

Felix C. Vaura involved in design of cluster analysis, design of risk analysis, data analysis, and main drafting of the work. Veikko V. Salomaa involved in design of the work and critical revision of the work. Ilkka M. Kantola, Risto Kaaja, Leo Lahti involved in design of the work and critical revision of the work. Teemu J. Niiranen involved in conception and main design of the work, main design of risk analysis, and drafting and main critical revision of the work.

ORCID

Felix C. Vaura https://orcid.org/0000-0002-6036-889X
Veikko V. Salomaa https://orcid.org/0000-0001-7563-5324
Teemu J. Niiranen https://orcid.org/0000-0002-7394-7487

REFERENCES

1. Forouzanfar MH, Afshin A, Alexander LT, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1659-1724.
2. World Health Organization. A global brief on hypertension: silent killer, global public health crisis: World Health Day 2013; 2013. https://apps.who.int/iris/handle/10665/79059. Accessed April 6, 2020.
3. Saklayen MG, Deshpande NV. Timeline of history of hypertension treatment. *Front Cardiovasc Med*. 2016;3:3.
4. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension. *Eur Heart J*. 2018;39:3021-3104.
5. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *J Am Coll Cardiol*. 2014;63:2935-2959.
6. Rodriguez MZ, Comin CH, Casanova D, et al. Clustering algorithms: a comparative approach. *PLOS ONE*. 2019;14:e0210236.
7. Guo Q, Lu X, Gao Y, et al. Cluster analysis: a new approach for identification of underlying risk factors for coronary artery disease in essential hypertensive patients. *Sci Rep*. 2017;7:43965.
8. Yang D, Nie Z, Liao L, et al. Phenomapping of subgroups in hypertensive patients using unsupervised data-driven cluster analysis: an exploratory study of the SPRINT trial. *Eur J Prev Cardiol*. 2019;26:1693-1706.
9. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort profile: the National FINRISK Study. *Int J Epidemiol*. 2018;47:696-696i.
10. Kastarinen M, Antikainen R, Peltonen M, et al. Prevalence, awareness and treatment of hypertension in Finland during 1982–2007. *J Hypertens*. 2009;27:1552-1559.
11. Jula A, Kantola I, Lehto S, et al. Working group set up by the Finnish Medical Society Duodecim and the Finnish Hypertension Society. Update on current care guidelines: hypertension. Current care guidelines. *Duodecim* 2010;126(6):673-674.
12. Razminia M, Trivedi A, Molnar J, et al. Validation of a new formula for mean arterial pressure calculation: the new formula is superior to the standard formula. *Catheter Cardiovasc Interv*. 2004;63:419-425.
13. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604-612.
14. Gissler M, Haukka J. Finnish health and social welfare registers in epidemiological research. *Nor Epidemiol*. 2009;14(1):113–120.
15. Pajunen P, Koukkunen H, Ketonen M, et al. The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil*. 2005;12:132-137.
16. Tolonen H, Salomaa V, Torppa J, et al. The validation of the Finnish Hospital Discharge Register and Causes of Death Register data on stroke diagnoses. *Eur J Cardiovasc Prev Rehabil*. 2007;14:380-385.
17. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav Genet*. 2009;39:580-595.
18. McCaw Z. RNOmni: Rank Normal Transformation Omnibus Test. R package version 0.7.1. https://CRAN.R-project.org/package=RNOmni
19. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J Classif*. 2014;31:274-295.
20. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53-65.
21. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0. https://CRAN.R-project.org/package=cluster
22. Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag (Abingdon)*. 1901;2:559-572.
23. Harrell FE. rms: Regression Modeling Strategies. R package version 5.1-4. https://CRAN.R-project.org/package=rms
24. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*. 1982;69:239-241.
25. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515-526.
26. Therneau TM. survival: a package for survival analysis in R. R package version 3.1-11. https://CRAN.R-project.org/package=survival
27. Alberti KGMM, Zimmet P, Shaw J, IDF Epidemiology Task Force Consensus Group. The metabolic syndrome – a new worldwide definition. *Lancet*. 2005;366:1059-1062.
28. Volpe M, Battistoni A, Savoia C, Tocci G. Understanding and treating hypertension in diabetic populations. *Cardiovasc Diagn Ther*. 2015;5:11.
29. Niiranen TJ, Lyass A, Larson MG, et al. Prevalence, correlates, and prognosis of healthy vascular aging in a western community-dwelling cohort: the Framingham Heart Study. *Hypertension*. 2017;70:267-274.
30. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med J Br Diabet Assoc*. 1998;15:539-553.
31. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III) final report. *Circulation*. 2002;106:3143-3421.
32. Grundy SM, Cleeman JI, Daniels SR, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/

National Heart, Lung, and Blood Institute Scientific Statement. *Circulation*. 2005;112:2735-2752.
33. Ford ES. Prevalence of the metabolic syndrome defined by the International Diabetes Federation among adults in the U.S. *Diabetes Care*. 2005;28:2745-2749.
34. Tune JD, Goodwill AG, Sassoon DJ, Mather KJ. Cardiovascular consequences of metabolic syndrome. *Transl Res J Lab Clin Med*. 2017;183:57-70.
35. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18:91-93.
36. Fischer MA, Choudhry NK, Brill G, et al. Trouble getting started: predictors of primary medication nonadherence. *Am J Med*. 2011;124(11):1081.e9–1081.e22.
37. Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons; 1990.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Vaura FC, Salomaa VV, Kantola IM, Kaaja R, Lahti L, Niiranen TJ. Unsupervised hierarchical clustering identifies a metabolically challenged subgroup of hypertensive individuals. *J Clin Hypertens*. 2020;22:1546–1553. https://doi.org/10.1111/jch.13984