Atherosclerosis 294 (2020) 10–15

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

# Data-driven multivariate population subgrouping via lipoprotein phenotypes *versus* apolipoprotein B in the risk assessment of coronary heart disease

Pauli Ohukainen[a,b,c], Sanna Kuusisto[a,b,c,d], Johannes Kettunen[a,b,c,e], Markus Perola[e,f,g], Marjo-Riitta Järvelin[b,c,h,i,j], Ville-Petteri Mäkinen[k,l], Mika Ala-Korpela[a,b,c,d,*]

[a] *Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland*
[b] *Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland*
[c] *Biocenter Oulu, University of Oulu, Oulu, Finland*
[d] *NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland*
[e] *National Institute for Health and Welfare, Helsinki, Finland*
[f] *Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland*
[g] *Estonian Genome Center, University of Tartu, Tartu, Estonia*
[h] *Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland*
[i] *Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK*
[j] *Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK*
[k] *Computational and Systems Biology Program, Precision Medicine Theme, South Australian Health and Medical Research Institute, Australia*
[l] *Hopwood Centre for Neurobiology, Lifelong Health Theme, SAHMRI, Australia*

## HIGHLIGHTS

- Data-driven subgrouping algorithm was trained by multivariate lipoprotein data.
- Four coherent subgroups were identified in two large-scale population-based cohorts.
- Subgroups had characteristic lipoprotein profiles and risk for CHD.
- Apolipoprotein B quartiles stratified CHD risk better than multivariate subgroups.
- Caution on multivariate data-driven subgrouping in risk assessment is warranted.

## ARTICLE INFO

*Keywords:*
Apolipoprotein B
Lipoproteins
CHD
Risk assessment
Population subgroups
Data-driven
Artificial intelligence

## ABSTRACT

*Background and aims:* Population subgrouping has been suggested as means to improve coronary heart disease (CHD) risk assessment. We explored here how unsupervised data-driven metabolic subgrouping, based on comprehensive lipoprotein subclass data, would work in large-scale population cohorts.
*Methods:* We applied a self-organizing map (SOM) artificial intelligence methodology to define subgroups based on detailed lipoprotein profiles in a population-based cohort (n = 5789) and utilised the trained SOM in an independent cohort (n = 7607). We identified four SOM-based subgroups of individuals with distinct lipoprotein profiles and CHD risk and compared those to univariate subgrouping by apolipoprotein B quartiles.
*Results:* The SOM-based subgroup with highest concentrations for non-HDL measures had the highest, and the subgroup with lowest concentrations, the lowest risk for CHD. However, apolipoprotein B quartiles produced better resolution of risk than the SOM-based subgroups and also striking dose-response behaviour.
*Conclusions:* These results suggest that the majority of lipoprotein-mediated CHD risk is explained by apolipoprotein B-containing lipoprotein particles. Therefore, even advanced multivariate subgrouping, with comprehensive data on lipoprotein metabolism, may not advance CHD risk assessment.

* Corresponding author. Computational Medicine, Faculty of Medicine University of Oulu, Oulu, Finland.
*E-mail address:* mika.ala-korpela@oulu.fi (M. Ala-Korpela).

https://doi.org/10.1016/j.atherosclerosis.2019.12.009
Received 30 October 2019; Received in revised form 2 December 2019; Accepted 12 December 2019
Available online 13 December 2019

## 1. Introduction

Increasing amounts of data available in epidemiology and medicine have generated interest in more detailed stratification of disease risk. Data-driven subgroup analyses have revealed new metabolic characteristics of complex diseases and uncovered subgroup-specific risk factors that could potentially improve risk assessment [1]. Recent studies have investigated this approach in type 1 [2] and type 2 diabetes [3] as well as in sepsis [4]. Various algorithms can be utilised for subgrouping [5] but the core principle is the same; to characterise a heterogeneous population so that individuals with shared metabolic, genetic and/or clinical characteristics are grouped together. In addition to identifying subgroup-specific risk, this approach can be useful in understanding complex multivariate phenotypes and in finding metabolically and maybe also genetically characteristic subgroups of individuals [1].

Here we applied a statistical artificial intelligence framework – a so-called self-organizing map (SOM) – that clusters individuals without explicit boundaries between groups or cut-off values for variables [5–7]. SOM analyses have a long-term track record in biomedical applications [1,2,6,7] and recently open-source software, aimed at large-scale epidemiological data, was published as an R library [5]. Conceptually, the SOM is a projection of multi-dimensional data onto a two-dimensional map [5,7]. For example, in this study each participant is assigned a location on the map based on a pre-selected set of lipoprotein-related variables: people within the same map area share a similar overall lipoprotein profile, while people far apart have different profiles. Therefore, comparisons between map areas are analogous to comparisons between subgroups of individuals.

Detailed quantitative molecular data are becoming increasingly common for large-scale studies in epidemiology via quantitative high-throughput metabolomics [8,9]. A nuclear magnetic resonance (NMR) spectroscopy-based platform has been broadly applied in epidemiology and genetics over the last few years; this platform is particularly advantageous in detailed lipoprotein profiling [8–12]. These metabolic data are typically continuous and do not instinctively represent subgroups but continuous, heavily overlapping distributions. However, intuitive thinking would be that elaborate utilisation of extensive multivariate data would not only lead to better understanding of the complexities but also to better translational opportunities and improved disease risk assessment.

We investigated here if unsupervised data-driven metabolic subgrouping, with SOM-based artificial intelligence and comprehensive NMR-based lipoprotein subclass data in large-scale population cohorts, could provide new insight on coronary heart disease (CHD) risk assessment. We demonstrate the resulting metabolic characteristics of the SOM-based subgroups and compare their risk assessment abilities to those of univariate subgroups based on a well-known causal CHD biomarker, apolipoprotein B (apoB) [13–16].

## 2. Materials and methods

### 2.1. Population cohorts

The Northern Finland Birth Cohort 1966 (NFBC66) was set up in the two northernmost provinces of Finland to study factors associated with preterm birth and morbidity during follow-up (www.oulu.fi/nfbc). Originally, a total of 12,058 children (96% of all births in 1966 in the region) were born into the cohort. For this study, we utilised data from 46-year sample collection in which a well representative 52% of the original cohort attended [17]. NMR-based lipoprotein data (96% fasting samples) were available from 5789 participants.

FINRISK 1997 (FINRISK97) is a nationally representative cohort, established by the Finnish National Institute for Health and Welfare to monitor middle-aged population health outcomes and risk factors [18]. Originally 8444 participants aged 25–74 years were recruited and 15-year follow-up data was available. NMR-based lipoprotein profiling was from semi-fasted (minimum 4 h of fasting before blood was drawn) serum samples from 7607 participants (mean age 48 ± 13 years).

### 2.2. Apolipoprotein, lipid and lipoprotein subclass analyses

An NMR spectroscopy-based methodology that is currently widely applied in large-scale epidemiology and genetics was applied [8–12]. This platform is powerful in lipoprotein subclass analysis and its large-scale epidemiological applications have recently been reviewed [9]. Briefly, the method provides apolipoprotein A-I (apoA-I) and B concentrations and standard clinical lipids; low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol as well as total cholesterol and triglycerides. In addition, quantitative data on lipoprotein particle concentrations and their main lipid constituents (phospholipids, triglycerides, cholesteryl esters and free cholesterol molecules) for 14 lipoprotein subclasses are obtained. The lipoprotein subclasses are characterised by particle size as follows: very-low-density lipoprotein (VLDL) fraction consists of extremely large (average diameter > 75 nm), very large (64 nm), large (53.6 nm), medium (44.5 nm), small (36.8 nm) and very small (31.3 nm) particles. Intermediate-density lipoprotein (IDL) particles are on average 28.6 nm in diameter. LDL particles are divided into three subclasses; large (25.5 nm), medium (23.0 nm) and small (18.7 nm). HDL fraction consists of four subclasses; very large (14.3 nm), large (12.1 nm), medium (10.9 nm) and small (8.7 nm).

### 2.3. Univariate subgrouping – apolipoprotein B quartiles

Apolipoprotein B quartiles were calculated and used in the survival analysis.

### 2.4. Multivariate subgrouping – self-organizing map analysis

The SOM analyses were undertaken with the Numero software package [5] in the R environment. Details and practicalities of SOM analysis are described elsewhere [5]. All analyses here were based on a total of 44 lipoprotein subclass measures (collectively referred to as input variables): particle, triglyceride and total cholesterol (cholesteryl esters and free cholesterol summed together in each subclass particle) concentration for six VLDL subclasses, IDL, three LDL subclasses, four HDL subclasses, apoB and apoA-I. Input variables were pre-processed with previously published tools within the Numero R package [5] (rank-based transform for men and women separately and normalized to the range between −1 and 1). The SOM analysis was first performed in NFBC66 to identify the apparent subgroups based on the extensive lipoprotein data. The resulted (trained) SOM and the pre-defined subgroups were then applied directly to classify the participants in the FINRISK97 cohort using an identical set of input variables. To visualise the overall lipid profile in each subgroup, z-scores of log-transformed measures were calculated as (subgroup mean – all data mean)/all data SD.

As a sensitivity analysis, fully independent SOM training, subgrouping and survival analysis were performed solely on the basis of the FINRISK97 data; the characteristics of the resulting SOM subgroups as well as the Kaplan-Meier curves were very similar. Excluding participants with a prevalent CHD at baseline (n = 199) or not including apoB as an input variable in the SOM analysis yielded essentially the same results. The interpretation and conclusions regarding the comparison between univariate and multivariate subgrouping were identical.

### 2.5. Survival analyses

Participants in FINRISK97 with prevalent CHD (n = 199) and those with missing data or outliers (n = 7) were removed. Final analyses for the 15-year follow-up had 7306 participants with 575 incident CHD

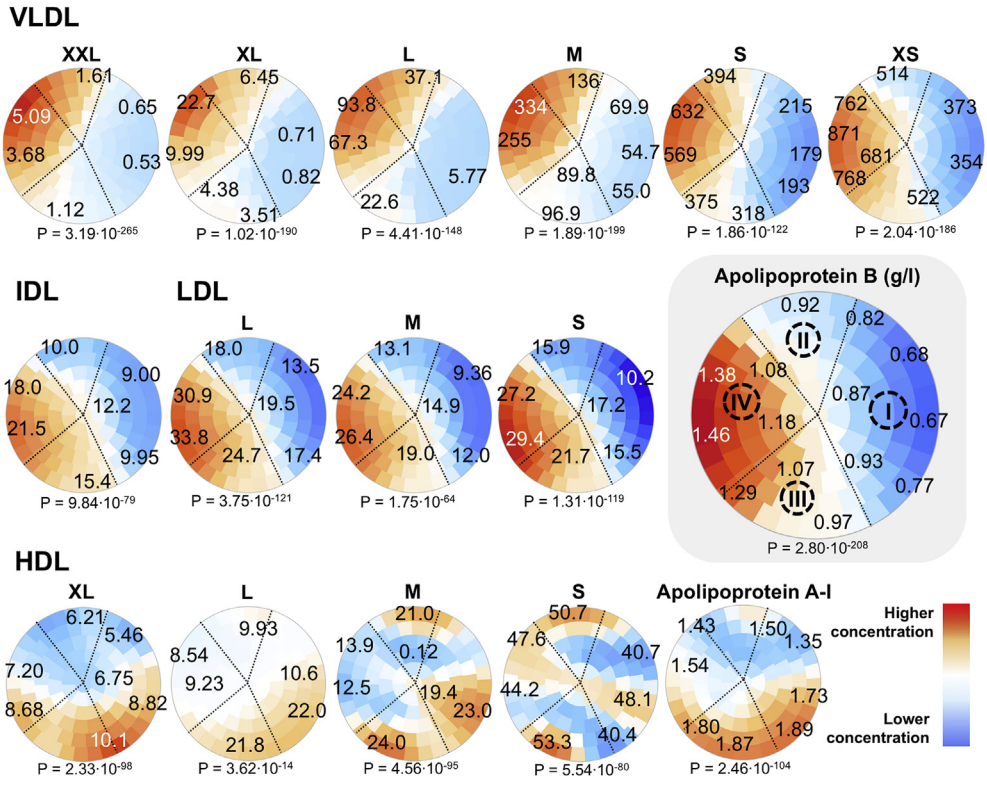


**Fig. 1.** Statistical colourings (component planes) of circulating lipoprotein particle, apolipoprotein A-I and apolipoprotein B concentrations on the self-organizing map. Each component plane shows colouring on the same SOM. The SOM illustrated is for the FINRISK97 cohort data for 7607 participants using 44 input variables (particle, triglyceride and total cholesterol concentration for 14 lipoprotein subclasses as well as apoA-I and apoB concentrations). The SOM organisation and the areas depicting the four population subgroups are from an independent training of the SOM for 5789 participants in the NFBC66. Subgroup I is characterised by the lowest, and subgroup IV the highest, mean apoB and related triglyceride and cholesterol concentrations. Subgroups II and III have intermediate concentrations of apoB but have elevated cholesterol and elevated triglycerides, respectively (see Fig. 2 for further details). The colour scale indicates deviation from the population mean with respect to random fluctuations that could be expected by chance; red refers to higher and blue for lower concentrations. The numbers on selected units tell the local mean value for that particular region in the original measurement unit (the values for VLDL are $10^{-10}$ mol/l, for IDL and LDL $10^{-8}$ mol/l, for HDL $10^{-7}$ mol/l, and for apoA-I and apoB g/l). The SOM is a two-dimensional organisation of the participants based on multi-dimensional input data, in this case 44 variables describing lipoprotein metabolism. The position on the map is unique and dependent on the input variable profile; thus each individual is always in the same place on each component plane. The P value below each component plane indicates the probability of observing equivalent regional variability for random data. Abbreviations: SOM, self-organizing map; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein; XXL, extremely large; XL, very large; L, large; M, medium; S, small; XS, very small; apoA-I, apolipoprotein A-I; apoB, apolipoprotein B.

events (defined as fatal or nonfatal myocardial infraction, cardiac revascularization, or unstable angina). Kaplan-Meier curves for each SOM-based subgroup were calculated for incident first CHD events. Identical analyses were performed for the apoB quartiles. Analyses were performed in R statistical language.

## 3. Results

### 3.1. Data-driven population subgrouping – SOM-based analysis

Results from the SOM-based classification of the participants in the FINRISK97 cohort are illustrated in Fig. 1. All 44 lipoprotein subclass measures (input variables) were used in the SOM but component planes (colourings of the SOM for individual variables) are shown only for the particle concentrations of the 14 lipoprotein subclasses as well as for apoA-I and apoB. The component planes for total cholesterol and triglyceride concentrations were very similar to the corresponding particle concentration ones shown. The component planes shown in Fig. 1 demonstrate strong regional patterns, particularly for the apoB-containing lipoprotein fractions (VLDL, IDL and LDL) and are labelled from I to IV in the ascending order of the subgroup mean apoB concentration. However, it is important to note that the circulating apoB concentration distributions overlap between all four subgroups, and heavily between the adjacent subgroups, as illustrated in Fig. 2.

Subgroup I is characterised by the lowest, and subgroup IV the highest, mean apoB concentration and the related triglyceride and cholesterol concentrations. The key separators for the adjacent subgroups II and III are VLDL and LDL particle concentrations, respectively. Thus, subgroup II is characterised by elevated triglycerides and rather low cholesterol; the situation is vice versa for subgroup III. The lipoprotein subclass particle concentration histograms together with apoB and apoA-I concentrations for the four subgroups are illustrated in Fig. 2. The characteristics and behaviour for VLDL, IDL and LDL subclasses is systematic however not identical within and between the subgroups. The HDL subclasses behave in a more heterogeneous manner, however subgroup I being characterised by the highest and the subgroup IV with the lowest HDL particle concentrations.

### 3.2. Risk of coronary heart disease by multivariate and univariate subgroups

Results from the survival analyses in the FINRISK97 for the SOM-based population subgroup are presented in Fig. 3A and for the apoB quartiles in Fig. 3B. The Kaplan-Meier curves based on the quartiles of circulating apoB concentrations show striking dose-response behaviour in contradiction to the subgroups from the multivariate SOM analysis. Despite substantially different lipoprotein subclass concentration profiles for SOM-based subgroups II and III (Fig. 2) they have highly overlapping curves to incident CHD (Fig. 3A).

## 4. Discussion

In this study, we present a novel application of an artificial intelligence algorithm, so-called self-organizing maps, to define subgroups based on detailed lipoprotein profiles in large population-based cohorts. Four distinct subgroups in relation to apolipoprotein B and A-I as well as to lipoprotein subclasses were characterised and the subgroup-specific risk for incident coronary heart disease risk evaluated. An instinctive expectation might be that utilisation of comprehensive multivariate data on lipoprotein metabolism would lead to better understanding and also to better estimation of coronary heart disease risk. However, the results did not fully support this expectation but a univariate analyses, using only apolipoprotein B quartiles, led to a better and more logical estimation of incident disease risk. In spite of that, the

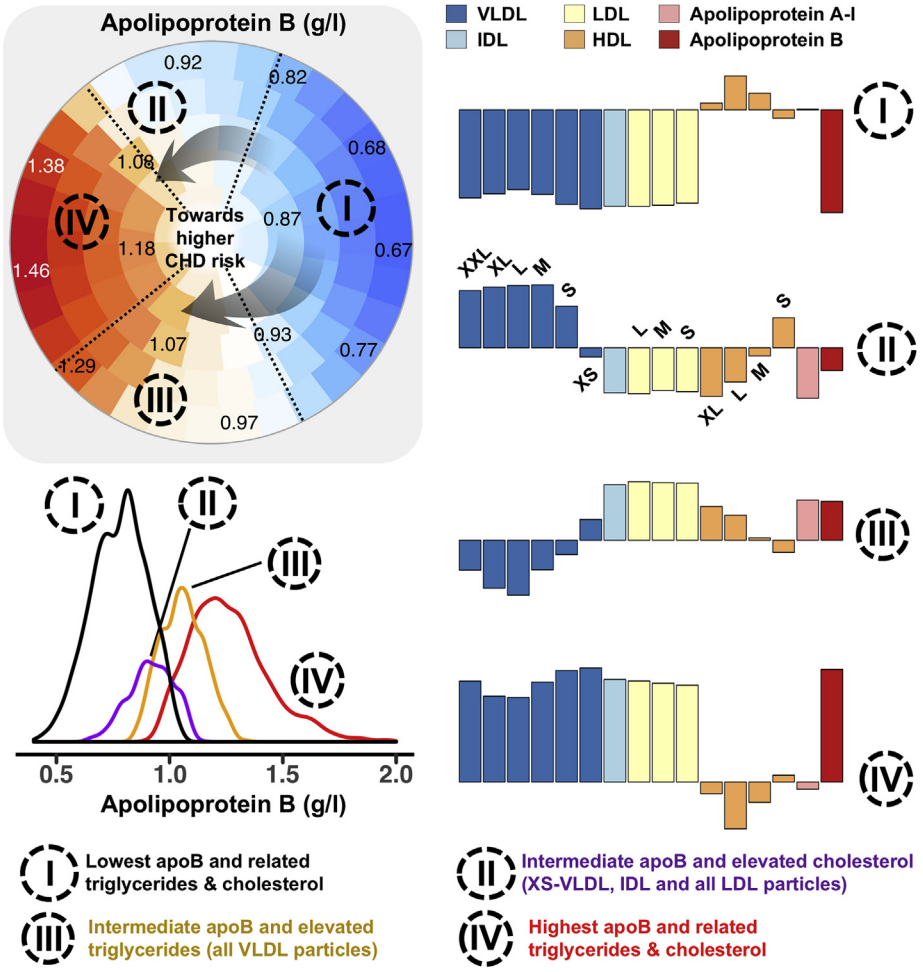


**Fig. 2.** Characteristics of the lipoprotein subclass profiles for the SOM-based population subgroups.
Analysis details and abbreviations are as explained in the caption for Fig. 1. Subgroup I is characterised by the lowest concentrations of apoB and coherently low concentrations for all apoB-containing lipoprotein subclasses. The opposite is the case for subgroup IV. Subgroups II and II have intermediate and rather similar concentrations for apoB. However, subgroup II is characterised by rather high concentrations of VLDL subclasses and rather low concentrations of IDL and LDL subclasses. The situation is opposite for subgroup III. XS-VLDL subclass shows intermediary behaviour between VLDL and LDL subclasses in subgroups II and III. The values shown in the histograms are z-scores of log-transformed measures ((subgroup mean – all data mean)/all data standard deviation). Note that while the mean apoB concentrations differ for the different subgroups, their distributions are heavily overlapping, particularly between the adjacent subgroups. The subgroup interpretations therefore are characteristic for the entire group and not necessarily for a single individual within the group, exactly as anticipated in population epidemiology [5–8].

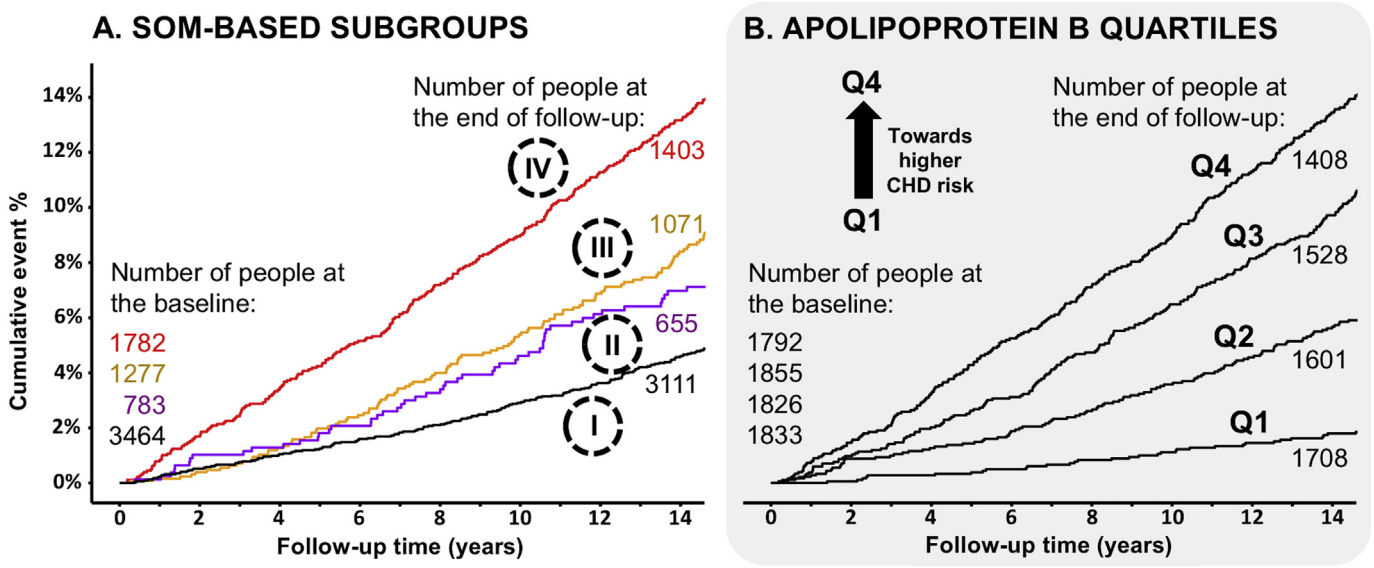


**Fig. 3.** Survival analysis of multivariate and univariate population subgroups.
In part A the subgroups are based on the SOM analyses illustrated and detailed in Figs. 1 and 2 and in B the subgroups represent apolipoprotein B quartiles. The Kaplan-Meier curves demonstrate the 15-year cumulative event risk for incident first coronary heart disease events (7306 individuals with 575 events including fatal or nonfatal myocardial infraction, cardiac revascularization or unstable angina). Individuals with baseline CHD were excluded from the analysis. The apolipoprotein B quartiles show excellent dose-response behaviour and better description of the overall risk than the participant subgroups from the multivariate SOM analysis.

data-driven SOM analysis gave invaluable detailed information on how lipoprotein metabolism, at the subclass level, relates to the risk of CHD. This kind of information cannot be obtained via univariate analysis and has biological and potentially translational value, though the results show that data-driven analysis is not optimal for outcome risk assessment [19].

The subgroup with the highest concentrations for non-HDL particles (VLDL, IDL and LDL) and for circulating apolipoprotein B represented the highest risk for CHD. Conversely, the subgroup characterised by the lowest values for these measures represented the lowest risk. The intermediate subgroups with elevated triglycerides and with elevated cholesterol had substantially different VLDL and LDL subclass profiles, yet a comparable concentration of apoB and overlapping event curves. Thus, even though multidimensional and comprehensive data on lipoprotein subclass profiles were used, the apoB concentrations in the population subgroups appeared to be directly related to the CHD risk, even in the presence of major variation in cholesterol and triglyceride concentrations.

In this context is would be good to note that the circulating apoB concentrations overlap between all four SOM-based subgroups. This is obviously in contradiction with the apoB quartiles that by definition are separate. Our interpretation for the situation is that in the multivariate SOM analysis, with the equally weighted set of lipoprotein measures, inclusion of some variables that might not directly (or not as well as apoB) relate to the incident CHD events, though they markedly vary between individuals, is likely to diminish the predictive values of the subgroups. While the multivariate metabolic lipoprotein data on lipoprotein subclasses is noteworthy in understanding details related to lipoprotein metabolism, a single good (in this case causal) biomarker is likely to be more useful from the predictive perspective.

Even though the results regarding the role of apoB as a single predictive biomarker may not be instinctive from the data analysis point of view, they are not surprising from the biological perspective and add to the burgeoning evidence for the fundamental role of apoB-containing lipoprotein particles in the development of atherosclerosis and in defining the risk for CHD [13–16,20–23].

Particularly Mendelian randomization analyses, using genetic instrument in large-scale studies, have played a crucial role in increasing our knowledge on the key causal molecular players in CHD [24]. A recent extensive study by Ference at al. [13], comparing the effects of genetic modification of lowering triglycerides with the lowering of LDL cholesterol, convincingly showed that the clinical benefit of lowering triglycerides as well as LDL cholesterol is proportional to the absolute change in the circulating apoB concentration. Thus, the apoB-containing lipoprotein particles appear to be the key factor, not the lipids per se transported in these particles. However, the apoB protein molecule does not circulate without lipids, so if there is an apoB molecule, there are also lipid molecules, but the apoB seems to the biological component that defines the way [13–16,20–23].

These results have general implications on data-driven subgrouping in epidemiology and potential translational applications. SOMs have been successfully used to identify metabolically different subgroups in patients with type 1 diabetes and thus to gain deeper understanding of population diversity and multi-morbidity [1,2,5,6]. However, the risk prediction for specific clinical endpoints is a separate issue calling for careful and detailed analysis [19] and should not be conflated with exploratory studies. Even though individuals can be clustered with several different methods and on the basis of various metabolic data, these measures might not be optimal from the risk assessment perspective and the subgrouping may thus not provide general clinical utility. Theoretically, an unsupervised clustering is likely to suffer from a large number of variables that carry information not related to the outcome. However, all lipoprotein data used in this work in the SOM analyses were associated with CHD, and thus we consider this a negligible phenomenon in this study.

Our results are conceptually consistent with a recent study in which data-driven cluster analysis was applied in patients with newly diagnosed type 2 diabetes [25]. In two type 2 diabetes related trials the authors applied a previously optimistically presented data-driven population subgrouping [3] and compared the clinical utility of this subgroup-based approach to predicting patient outcomes with an alternative strategy of developing models for each outcome using simple patient characteristics. Their conclusion was that for the best clinical utility, approaches using specific phenotypic measures to predict specific outcomes would most likely perform better than assigning patients to subgroups [25]. This independent finding supports our results and interpretation in relation to the data-driven population subgrouping *versus* apolipoprotein B in assessing the risk of CHD.

### 4.1. Conclusion

The results presented here provide evidence to temper some of the enthusiasm towards multivariable stratification and risk profiling in epidemiology and potential translational applications. Population subgroups with distinct metabolic and risk profiles can be identified but any incremental clinical utility should be specifically determined by rigorous testing against the most validated existing markers.

## Financial support

PO is supported by the Emil Aaltonen Foundation. JK and MAK are supported by a research grant from the Sigrid Juselius Foundation, Finland. The cohorts and this work have also been supported by funding from the Academy of Finland, Novo Nordisk Foundation and EU.

## Author contributions

All listed authors meet the requirements for authorship. Concept and design: PO, MAK. Clinical data: MP, MRJ, MAK. Lipoprotein analyses: MAK. Analysis plan and interpretations: PO, SK, JK, VPM and MAK. Statistical analyses: PO and SK. Draft manuscript: PO, MAK. All authors commented the manuscript and agreed to its content. Overall responsibility: PO, MAK.

## Declaration of competing interest

The authors declared they do not have anything to disclose regarding conflict of interest with respect to this manuscript.

## References

[1] M. Ala-Korpela, Data-driven subgrouping in epidemiology and medicine, Int. J. Epidemiol. 48 (2019) 374–376, https://doi.org/10.1093/ije/dyz040.
[2] R. Lithovius, I. Toppila, V. Harjutsalo, C. Forsblom, P.H. Groop, et al., Data-driven metabolic subtypes predict future adverse events in individuals with type 1 diabetes, Diabetologia 60 (2017) 1234–1243, https://doi.org/10.1007/s00125-017-4273-8.
[3] E. Ahlqvist, P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, 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 (2018) 361–369, https://doi.org/10.1016/S2213-8587(18)30051-2.
[4] C.W. Seymour, J.N. Kennedy, S. Wang, C.H. Chang, C.F. Elliott, et al., Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis, J. Am. Med. Assoc. 321 (2019) 2003–2017, https://doi.org/10.1001/jama.2019.5791.
[5] S. Gao, S. Mutter, A. Casey, V.-P. Mäkinen, Numero: a statistical framework to define multivariable subgroups in complex population-based datasets, Int. J. Epidemiol. 48 (2019) 369–374, https://doi.org/10.1093/ije/dyy113.
[6] V.P. Mäkinen, C. Forsblom, L.M. Thorn, J. Wadén, D. Gordin, et al., Metabolic phenotypes, vascular complications, and premature deaths in a population of 4,197 patients with type 1 diabetes, Diabetes 57 (2008 Sep) 2480–2487, https://doi.org/10.2337/db08-0332.
[7] L.S. Kumpula, S.M. Mäkelä, V.P. Mäkinen, A. Karjalainen, J.M. Liinamaa, et al., Characterization of metabolic interrelationships and in silico phenotyping of lipoprotein particles using self-organizing maps, J. Lipid Res. 51 (2010) 431–439, https://doi.org/10.1194/jlr.D000760.
[8] M. Ala-Korpela, G. Davey Smith, Metabolic profiling-multitude of technologies with great research potential, but (when) will translation emerge? Int. J. Epidemiol. 45

(2016) 1311–1318, https://doi.org/10.1093/ije/dyw305.
[9] P. Würtz, A.J. Kangas, P. Soininen, D.A. Lawlor, G. Davey Smith, et al., Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies, Am. J. Epidemiol. 186 (2017) 1084–1096, https://doi.org/10.1093/aje/kwx016.
[10] J. Kettunen, A. Demirkan, P. Würtz, H.H.M. Draisma, T. Haller, et al., Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA, Nat. Commun. 7 (2016) 11122, https://doi.org/10.1038/ncomms11122.
[11] A.E. Locke, K.M. Steinberg, C.W.K. Chiang, S.K. Service, A.S. Havulinna, et al., Exome sequencing of Finnish isolates enhances rare-variant association power, Nature 572 (2019) 323–328, https://doi.org/10.1038/s41586-019-1457-z.
[12] T. Tukiainen, J. Kettunen, A.J. Kangas, L.P. Lyytikäinen, P. Soininen, et al., Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci, Hum. Mol. Genet. 21 (2012) 1444–1455, https://doi.org/10.1093/hmg/ddr581.
[13] B.A. Ference, J.J.P. Kastelein, K.K. Ray, H.N. Ginsberg, M.J. Chapman, et al., Association of triglyceride-lowering LPL variants and LDL-C–lowering LDLR variants with risk of coronary heart disease, J. Am. Med. Assoc. 321 (2019) 364–373, https://doi.org/10.1001/jama.2018.20045.
[14] A.D. Sniderman, M. Pencina, G. Thanassoulis, ApoB: the power of physiology to transform the prevention of cardiovascular disease, Circ. Res. 124 (2019) 1425–1427, https://doi.org/10.1161/CIRCRESAHA.119.315019.
[15] M. Ala-Korpela, The culprit is the carrier, not the loads: cholesterol, triglycerides and apolipoprotein B in atherosclerosis and coronary heart disease, Int. J. Epidemiol. (2019), https://doi.org/10.1093/ije/dyz068.
[16] J. Borén, K.J. Williams, The central role of arterial retention of cholesterol-rich apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a triumph of simplicity, Curr. Opin. Lipidol. 27 (2016) 473–483, https://doi.org/10.1097/MOL.0000000000000330.
[17] M.R. Järvelin, U. Sovio, V. King, L. Lauren, B. Xu, et al., Early life factors and blood pressure at age 31 years in the 1966 Northern Finland birth cohort, Hypertension 44 (2004) 838–846, https://doi.org/10.1161/01.HYP.0000148304.33869 (ee).
[18] K. Borodulin, E. Vartiainen, M. Peltonen, P. Jousilahti, A. Juolevi, et al., Forty-year trends in cardiovascular risk factors in Finland, Eur. J. Public Health 25 (2015) 539–546, https://doi.org/10.1093/eurpub/cku174.
[19] M. van Smeden, F.E. Harrell Jr., D.L. Dahly, Novel diabetes subgroups, Lancet Diabetes Endocrinol. 6 (2018) 439–440, https://doi.org/10.1016/S2213-8587(18)30124-4.
[20] K. Skålén, M. Gustafsson, E.K. Rydberg, L.M. Hultén, O. Wiklund, et al., Subendothelial retention of atherogenic lipoproteins in early atherosclerosis, Nature 417 (2002) 750–754, https://doi.org/10.1038/nature00804.
[21] I. Tabas, K.J. Williams, J. Borén, Subendothelial lipoprotein retention as the initiating process in atherosclerosis: update and therapeutic implications, Circulation 116 (2007) 1832–1844, https://doi.org/10.1161/CIRCULATIONAHA.106.676890.
[22] S.D. Proctor, D.F. Vine, J.C. Mamo, Arterial retention of apolipoprotein B(48)- and B(100)-containing lipoproteins in atherogenesis, Curr. Opin. Lipidol. 13 (2002) 461–470, https://doi.org/10.1097/00041433-200210000-00001.
[23] J.L. Goldstein, M.S. Brown, A century of cholesterol and coronaries: from plaques to genes to statins, Cell 161 (2015) 161–172, https://doi.org/10.1016/j.cell.2015.01.036.
[24] M.V. Holmes, M. Ala-Korpela, G.D. Smith, Mendelian randomization in cardiometabolic disease: challenges in evaluating causality, Nat. Rev. Cardiol. 14 (2017) 577–590, https://doi.org/10.1038/nrcardio.2017.78.
[25] J.M. Dennis, B.M. Shields, W.E. Henley, A.G. Jones, A.T. Hattersley, Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data, Lancet Diabetes Endocrinol. 7 (2019) 442–451, https://doi.org/10.1016/S2213-8587(19)30087-7.