Cardiovascular Diabetology

RESEARCH

Open Access

BMI-residualized data uncovers a cluster of people with type 2 diabetes and increased serum ferritin protected from cardiovascular disease

Laura Gallardo-Nuell[1,2,3†], Jordi Blanch[5,7†], Yenny Leal[1,2,3], Daniel E Coral[8], Talita Duarte-Salles[9,10], Giuseppe N. Giordano[8], Paul W. Franks[8], Ewan R. Pearson[11], Geltrude Mingrone[12,13,14], Carel W le Roux[15], Rafael Ramos[5,6,7,4,16*] and José Manuel Fernández-Real[1,2,3,4*]

Abstract

Background Understanding the relationship between serum ferritin levels and cardiovascular outcomes in type 2 diabetes is crucial for improving risk stratification and guiding therapeutic interventions aimed at preventing major adverse cardiovascular events (MACE). This study aimed to identify distinct clusters of individuals with type 2 diabetes who have varying risks of MACE using a data-driven clustering approach.

Methods This retrospective cohort study analyzed data from 49,506 individuals within a multicenter, population-based primary care registry in Catalonia, Spain. Individuals diagnosed with type 2 diabetes at age 35 or older were recruited between January 2010 and December 2021 and followed for at least 10 years. Biomarkers associated with cardiovascular risk—including serum glucose, HbA1c, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, blood pressure, serum ferritin, leukocyte, and monocyte counts—were examined. Clustering analysis was applied to identify patient subgroups, and Cox proportional hazards models were used to assess associations with cerebrovascular events, coronary events, and composite MACE.

Results Five distinct clusters were identified, characterized by differences in serum glucose, HbA1c, lipid profiles, blood pressure, and serum ferritin levels. Individuals with discordantly high serum ferritin levels relative to their body mass index (BMI) exhibited a lower risk of adverse cardiovascular outcomes. In men, hazard ratios (HR) were 0.68 (95% confidence interval [CI]: 0.53–0.87) for cerebrovascular events, 0.65 (95% CI 0.49–0.88) for coronary events, and 0.68 (95% CI 0.56–0.83) for MACE. In women, HRs were 0.81 (95% CI 0.67–0.92) for cerebrovascular events, 0.73 (95% CI 0.57–0.95) for coronary events, and 0.79 (95% CI 0.67–0.92) for MACE.

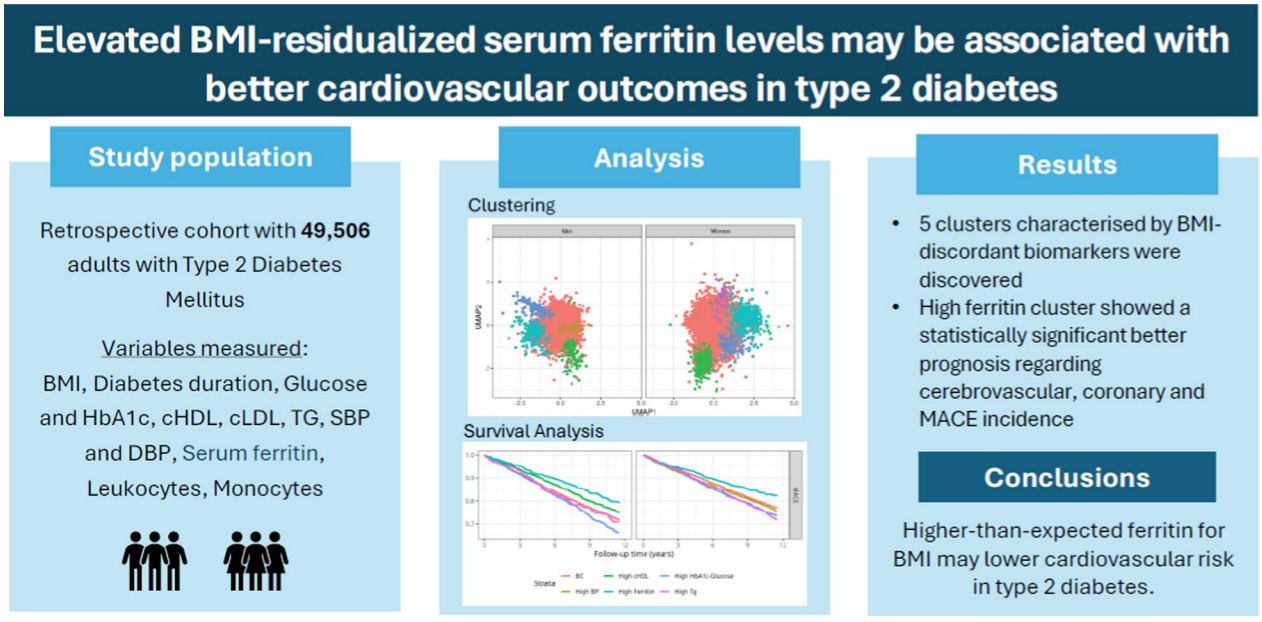
†Laura Gallardo-Nuell and Jordi Blanch contributed equally and share first authorship.

*Correspondence:
Rafael Ramos
rramos.girona.ics@gencat.cat
José Manuel Fernández-Real
jmfreal@idibgi.org

Full list of author information is available at the end of the article

Conclusions Individuals with type 2 diabetes who exhibit higher-than-expected serum ferritin levels relative to their BMI may have a lower risk of cardiovascular events. These findings suggest that ferritin may play a more complex role in cardiovascular risk than previously assumed and highlight the potential for refined risk stratification strategies in type 2 diabetes management.

Graphical abstract



Keywords Type 2 diabetes mellitus, Obesity, Ferritin, BMI, Cardiovascular diseases(s) (CVD)

Background

Type 2 diabetes is a well-recognized major global health threat, affecting millions of individuals worldwide and placing a significant burden on healthcare systems. The prevalence of this chronic disease continues to rise, driven by factors such as ageing populations, urbanisation, and lifestyle changes, including poor diet and physical inactivity. Given its progressive nature, type 2 diabetes is not only associated with high morbidity but also with serious long-term complications, such as cardiovascular disease, neuropathy, nephropathy, and retinopathy, which further exacerbate patient outcomes and healthcare costs [1, 2].

The risk of developing type 2 diabetes (T2D) is determined by a complex interplay of lifestyle, environmental, and genetic factors [3]. Key contributors to the development and progression of T2D include age, body mass index (BMI), smoking, lipid profile, or serum ferritin levels [4]. Increasing age and higher BMI are strongly associated with an elevated risk of T2D, while smoking exacerbates the condition by contributing to insulin resistance and impaired glucose metabolism. Additionally, an unfavourable lipid profile, characterised by elevated levels of LDL cholesterol and triglycerides, further exacerbates the risk of diabetes and its complications. Elevated serum ferritin levels, an indicator of body iron stores, have also been implicated in the pathophysiology of T2D [4]. Collectively, these factors influence the development and progression of T2D and its associated complications, underscoring the need for comprehensive strategies to address these diverse risk elements.

Among these factors, iron plays a notable role. Sullivan's hypothesis in 1981 [5] and the first clinical observation in 1992 by Salonen et al. [6], proposed that increased body iron might favour the development of atherosclerosis through the increased production of reactive oxygen species (ROS). There has been substantial heterogeneity across studies regarding this hypothesis as evidenced in recent meta-analyses [7]. Puzzlingly, more recently, increased plasma transferrin saturation has been shown to putatively protect from coronary artery disease (CAD) or myocardial infarction [8], and J-shaped association of systemic iron status and cardiovascular mortality has been described in CAD patients [9]. For instance, a study showed that among 1,480 patients with stable CAD and 682 individuals without CAD, the risk of having CAD was decreased in patients in the highest quartiles of transferrin saturation, serum ferritin and soluble

transferrin receptor levels [10]. In two studies performed in elderly subjects ≥ 65 and ≥ 71 years of age, low circulating iron was independently associated with the higher occurrence of cardiovascular disease (CVD) and CAD during the long-term follow-up [11, 12].

Moreover, a study showed that in patients with type 2 diabetes and clinically overt CAD, both low and high levels of serum ferritin were associated with a poor prognosis [9].

Ferritin, being an acute phase reactant, often increases in response to inflammation or infection. This could partly explain the paradoxical associations observed with ferritin levels, as elevated ferritin may reflect underlying inflammatory states rather than solely iron status.

In this context, there is increasing awareness regarding the beneficial effects of small increases in ROS levels, as it could lead to increase in ROS defences in response to ROS-dependent damage [13]. As signalling molecules, iron-induced ROS might activate appropriate protective mechanisms.

BMI is commonly used to assess obesity-related health risks and as mentioned, is associated with both type 2 diabetes and cardiovascular diseases. However, BMI alone is often insufficient for accurately classifying obesity-related health risks at an individual level as individuals with similar BMIs can experience significantly different health outcomes, revealing the limitations of relying solely on BMI as a measure of obesity-related health risks [14]. Research has consistently shown a link between elevated serum ferritin levels and obesity. Numerous studies across different populations have highlighted this association, with metabolic syndrome and type 2 diabetes mellitus identified as contributing factors to high serum ferritin levels in individuals with obesity [15–17]. Recent research has highlighted the complex relationship between ferritin levels and obesity, noting that ferritin increases with obesity and that this relationship can confound studies unless properly accounted for. A common method is to adjust for BMI, though this approach may overlook nuanced interactions and subgroup behaviours.

We hypothesised that the variability in iron-cardiovascular outcomes might be explained by the existence of clusters of individuals where some clinical measures are discordant from a linear relationship with BMI.

In the current study, we explored the potential clustering of patients with type 2 diabetes within a population-based cohort, using analytical variables related to CVD. We aimed to identify subpopulations based on expected BMI and compare these clusters to determine if they exhibit different risks of experiencing major adverse cardiovascular events (MACE).

Methods

Study design

In this retrospective study, data was obtained from the Information System for Research in Primary Care (SIDIAP) database, which gathers anonymized longitudinal information of people from the primary care services in a structured way so it can be reliably used for research [18]. This database contains pseudonymised records of over 8 million people since 2006. The information was mapped to Observational Medical Outcomes Partnership Common Data Model OMOP-CDM [19]. SIDIAP covers approximately 75% of the Catalan population, comprising 328 primary care centres managed by the Catalan Institute of Health. SIDIAP contains records related to demographics, diagnoses (coded according to the International Classification of Diseases, 10th revision (ICD-10-CM)), therapeutic and requested procedures, physical examination results, routine measurements and laboratory tests. SIDIAP is linked with other external databases (and anonymization is preserved); we were able to use information from the register of mortality and the pharmacy-invoicing database provided by the Catalan Institute of Health, in which medications are recorded using the Anatomical Therapeutic Chemical (ATC) codes. The quality and representativeness of these data regarding geographical, age, and sex distributions have been documented [18], particularly for cardiovascular risk factors and cardiovascular diseases [20], and have been applied to previous epidemiological reports [21, 22].

Participants

We carried out a retrospective cohort study including persons with type 2 diabetes mellitus who were 35 years old or older at diagnosis. Type 2 diabetes mellitus was defined following a validation algorithm [22]: (a) diagnosis of type 2 diabetes without prior diagnosis of type 1 or undetermined diabetes, (b) diagnosis of undetermined diabetes with another diagnosis of type 2 diabetes without any previous diagnosis of type 1 diabetes, (c) diagnosis of undetermined diabetes and taking non-insulin blood-glucose-lowering drugs (A10B) other than metformin and without diagnosis of type 1 or type 2 diabetes, (d) diagnosis of type 1 diabetes and without prescription of insulin, (e) diagnosis of type 1 diabetes and taking A10B other than metformin, (f) taking A10B other than metformin without any diabetes diagnosis (type 1, type 2 or undetermined). Exclusion criteria were (i) any end-stage kidney disease (eGFR < 15, or dialysis or transplantation), (ii) any diagnosis of depression, seizure disorder or schizophrenia, (iii) previous use of systemic steroids, (iv) any previous diagnosis of malignant tumour, (v) prescription of insulin within 6 months of type 2 diabetes mellitus diagnosis. The recruitment period ran from January 2010 through December 2021. Study entry was

defined by the first diagnosis of type 2 diabetes during this period or 1st January 2010 for people with a previous diagnosis, and the follow-up extended until the first date of the occurrence of an outcome, death, transfer out of the SIDIAP, or end of study period, in December 2021.

Procedures

The collected data included potential confounding variables for our study, measured before the entry date: sex, age, BMI, diabetes duration, glucose and HbA1c levels, lipid fractions (high-density lipoprotein cholesterol (cHDL), low-density lipoprotein cholesterol (cLDL), triglycerides), systolic and diastolic blood pressure (SBP, DBP), serum ferritin, and leukocyte and monocyte counts. The main outcomes, observed after entry date, were fatal and nonfatal coronary heart diseases (acute myocardial infarction (ICD-10CM: I21, I22 or I23), or angina pectoris (ICD-10CM: I20)), fatal and nonfatal cerebrovascular diseases (ischemic stroke (ICD-10CM: I63, I64 or I65 excluding I63.1 and I63.4) or transient ischemic attack (ICD-10CM: G45, G46, I65 or I66)). The primary outcome of the study was major adverse cardiovascular events (MACE), a composite endpoint of coronary heart and cerebrovascular diseases. All the events recorded in SIDIAP have been validated against those observed in the prospective REGICOR cohort [20].

Statistical analysis

We present categorical variables as percentages and continuous variables as means (standard deviations) or medians (interquartile ranges), as appropriate.

To identify our clusters, we followed the procedure described and implemented elsewhere [23]. Briefly, this cluster analysis consists in a multi-stage analytical pipeline that integrates linear modeling, residual extraction, dimensionality reduction, and unsupervised clustering to analyze biomarker data. The primary objectives are to assess the association between a set of biomarkers and BMI (with additional covariates), and subsequently to identify latent subgroups based on the residual structure of these models. The analysis is performed separately in each sex. More details in supplementary methods.

To assess the validity of the identified clusters within the cohort, we randomly generated 100 bootstrap subsets, each comprising at least 90% of the original data. Following, we estimated the clusters in each subset. A cluster was deemed validated if it appeared consistently across all subsets. Subsequently, we assigned each individual to the cluster with the highest probability of belonging, using the cluster with the most frequent occurrence in the probability distribution as their assignment.

To assess the risk of each validated cluster for each outcome, we used Kaplan–Meier survival curves and Cox regression models to calculate the Hazard ratios (HR). The concordant cluster serves as the reference category, and we estimated the HR to change from the concordant cluster to another cluster.

All statistical analyses were carried out using R software vr 4.3.1 [24], including *uwot* package [25] for UMAP and *survival* package [26] for survival analysis. Statistical significance was considered at p values < 0.05.

Role of the funding source The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The SIDIAP database comprises information on 8,036,948 individuals, of whom 983,398 had been diagnosed with type 2 diabetes mellitus in the study period. After excluding individuals based on specific criteria, we obtained a dataset of 463,357 participants (Additional File 1, Supplementary Fig. S1). For our analysis, we conducted a complete-case analysis, resulting in complete information on all covariables for 49,506 individuals. The distribution of the potential confounding variables is described in Table 1. We compared the distribution of these variables with and without performing a complete-case analysis and found no significant differences. The distribution of the variables without performing a complete-case analysis is available in Supplementary Table S1 (Additional File 1).

Figure1A shows the UMAP dimensional reduction of the standardised database according to standardisation variables and the assigned cluster for each individual, stratified by sex. The plot exhibits a circular shape with several protrusions in both sexes. We obtained five clusters for each sex, the concordant cluster (assigned to 78.64% of men and 78.92% of women, Table 2) and four distinct clusters. The concordant cluster is placed in the centre of the map. Figure 1B shows the values of covariable for the centre of each cluster in the standardised scale separating for sexes. For women, the distinct cluster's centres are characterised for having higher values of BP, both diastolic and systolic (6.95% of women, Table 2); higher values of glucose levels, both glucose and HbA1c (5.25%); higher values of Tg (4.60%); or higher values of ferritin (4.32%). For men, the centres are characterised by higher values of cHDL (9.04% of men); higher values of glucose, both glucose and HbA1c (5.43%); higher values of ferritin (4.01%); or higher values of Tg (3.17%). When we refer to "higher values"', we imply that these values are higher in respect to the expected for the given BMI and the other additional covariates. Individuals were assigned to discordant clusters if their probability of belonging to a cluster was higher than 80%. The distribution of individuals with over 80% probability of belonging to a category is shown in Supplementary Table S4 (Additional

Table 1 Distribution of the clinical variables of the SIDIAP cohort used

	Global	Men	Women
N	49,506	19,751	29,755
BMI (kg/m^2)	30.6 (5.5)	29.6 (4.7)	31.4 (5.9)
Age (years)	68.2 (12.5)	66.9 (12)	69 (12.8)
Time T2DM (years)	5.2 (5.3)	4.7 (4.9)	5.5 (5.5)
HbA1c (%)	6.8 (1.4)	6.8 (1.4)	6.8 (1.3)
Glucose (mg/dL)	140.5 (42.6)	142.8 (44.6)	138.9 (41.2)
Leukocytes (10^9/L)	7.3 (2.0)	7.4 (2.1)	7.3 (1.9)
Monocytes (%)	7.7 (2.1)	8.1 (2.1)	7.4 (2)
cLDL (mg/dL)	115.2 (33.0)	112.1 (33.1)	117.2 (32.8)
cHDL (mg/dL)	50.2 (13.0)	46.4 (12.0)	52.7 (13.1)
Tg (mg/dL)	133 [97–184]	130 [94–186]	134 [100–183]
DBP (mmHg)	76.1 (10.0)	76.4 (10.3)	75.9 (9.8)
SBP (mmHg)	135.6 (15.8)	135.7 (15.4)	135.5 (16)
Serum ferritin (ng/mL)	81 [33.7–176]	141 [62–276.9]	58 [25.8–118]
Current smoking	5775 (11.7%)	4001 (20.3%)	1774 (6%)
A10A (Insulin and analogues)	6568 (13.3%)	2313 (11.7%)	4255 (14.3%)
A10B (Blood glucose lowering drugs, excluding insulins)	28,820 (58.2%)	11,433 (57.9%)	17,387 (58.4%)
C10 (Lipid modifying agents)	23,977 (48.4%)	9462 (47.9%)	14,515 (48.8%)
Hypertension medication	35,996 (72.7%)	13,786 (69.8%)	22,210 (74.6%)

Results are expressed as mean and standard deviation (SD) for normal distributed continuous variables, median and interquartile range for non-normal distributed continuous variables and as number and percentage (of the presence) of those qualitative variables. Tg: Triglyceride, Time T2DM: Time since type 2 diabetes mellitus diagnosis, cLDL: LDL cholesterol, cHDL: HDL cholesterol, DBP: Diastolic blood pressure, SBP: Systolic blood pressure

File 1). All clusters exhibit a similar distribution for the standardisation variables, except for the A10A and A10B

Table 2 Distribution of individuals across validated clusters

Cluster	Men		Women	
	N	%	N	%
BC	15,474	78.35%	23,472	78,88%
High HbA1c and glucose	1073	5.43%	1561	5.25%
High Ferritin	793	4.01%	1285	4.32%
High Tg	626	3.17%	1370	4.60%
High BP	–	–	2067	6.95%
High cHDL	1,785	9.04%	–	–

Results are presented as the number of individuals assigned to each cluster with maximum probability and the corresponding percentage of the total population, stratified by sex. BC: Baseline concordant profile Tg: Triglyceride, BP: Blood Pressure, cHDL: HDL cholesterol

medications, which show a higher intake in the cluster with elevated glucose levels (Additional File 1, Supplementary Table S2 and Table S3).

Survival analysis

A survival study was conducted to assess the risk of developing cerebrovascular diseases, coronary disease or MACE. The survival analysis was performed with each individual allocated to the cluster with a probability higher than 80% of belonging.

The high ferritin cluster shows a higher disease-free survival for the three outcomes and for both sexes (Fig. 2) especially for MACE. Also, the high cHDL cluster has a higher disease-free survival for men for coronary disease and MACE. The high HbA1c and glucose level cluster has a lower disease-free survival for all outcomes and sexes. The high ferritin cluster has a significant protective effect for all outcomes and sexes (Fig. 3. HR in men: 0.68, 95% CI 0.53–0.87, $p = 0.0021$ for cerebrovascular, 0.65, 95% CI 0.49–0.88, $p = 0.0051$ for coronary and 0.68 95% CI 0.56–0.83, $p = 0.00014$ for MACE. HR in women: 0.81, 95% CI 0.67–0.92, $p = 0.025$ for cerebrovascular, 0.73, 95%

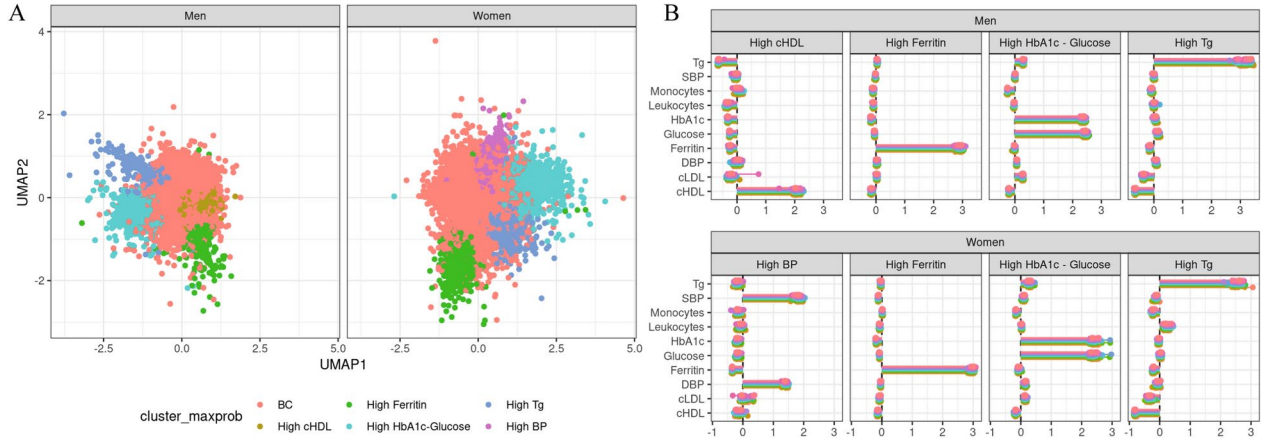


Fig. 1 Concordant and discordant profiles discovered in the SIDIAP. **A** Two-dimensional UMAP projection, representing the distribution of profiles in the dataset. The UMAP is colour-coded based on identified clusters. The left panel corresponds to men, while the right panel corresponds to women. **B** Profile centres for each identified cluster. The distinctive patterns in the profile centres provide a detailed insight into the characteristics defining each cluster. BC: Baseline concordant profile Tg, cLDL: LDL cholesterol, cHDL: HDL cholesterol, DBP: Diastolic blood pressure, SBP: Systolic blood pressure

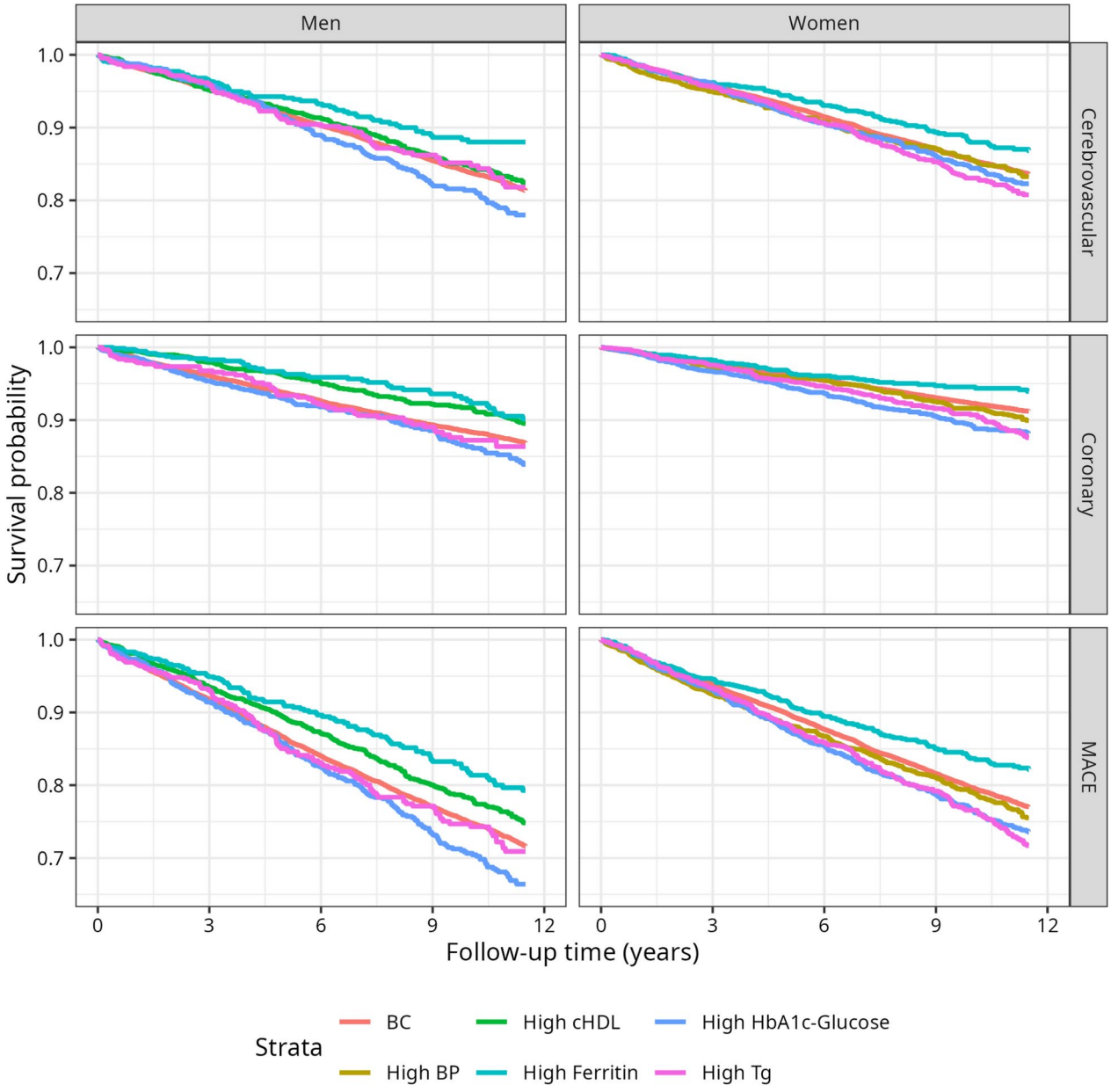


Fig. 2 Disease-free survival comparison between the identified clusters. Kaplan–Meier curve for overall disease-free survival in the identified clusters for the three cardiovascular studied outcomes—cerebrovascular, coronary, and MACE events—stratified by sex. BC: Baseline concordant profile Tg, cLDL: LDL cholesterol, cHDL: HDL cholesterol, DBP: Diastolic blood pressure, SBP: Systolic blood pressure.

CI 0.57–0.95, $p = 0.021$ for coronary, 0.7995% CI 0.67–0.92, $p = 0.003$ for MACE). The high cHDL cluster has a protective effect only for coronary disease and MACE for men (HR of 0.73, 95% CI 0.61–0.87, $p = 0.00065$, and 0.83, 95% CI 0.74–0.93, $p = 0.0015$, respectively). The high Tg cluster has a higher risk of having a coronary disease and MACE only for women (HR of 1.33, 95% CI 1.10–1.61, $p = 0.0032$ and 1.21, 95% CI 1.07–1.36, $p = 0.0026$ respectively). The high HbA1c and glucose level cluster has a higher risk for coronary only for men and MACE in both sexes (HR of 1.16, 95% CI 0.96–1.41, $p = 0.128$ for coronary in men and 1.17, 95% CI 1.02–1.33, $p = 0.0202$ for MACE in men and 1.18, 95% CI 1.05–1.32, $p = 0.0055$ for MACE in women). The analyses were repeated separately for incident and prevalent cases, and no differences were found compared to the overall analysis. The only exception was the high ferritin cluster in incident cases, where the protective effect remained but lost statistical significance due to the smaller sample size.

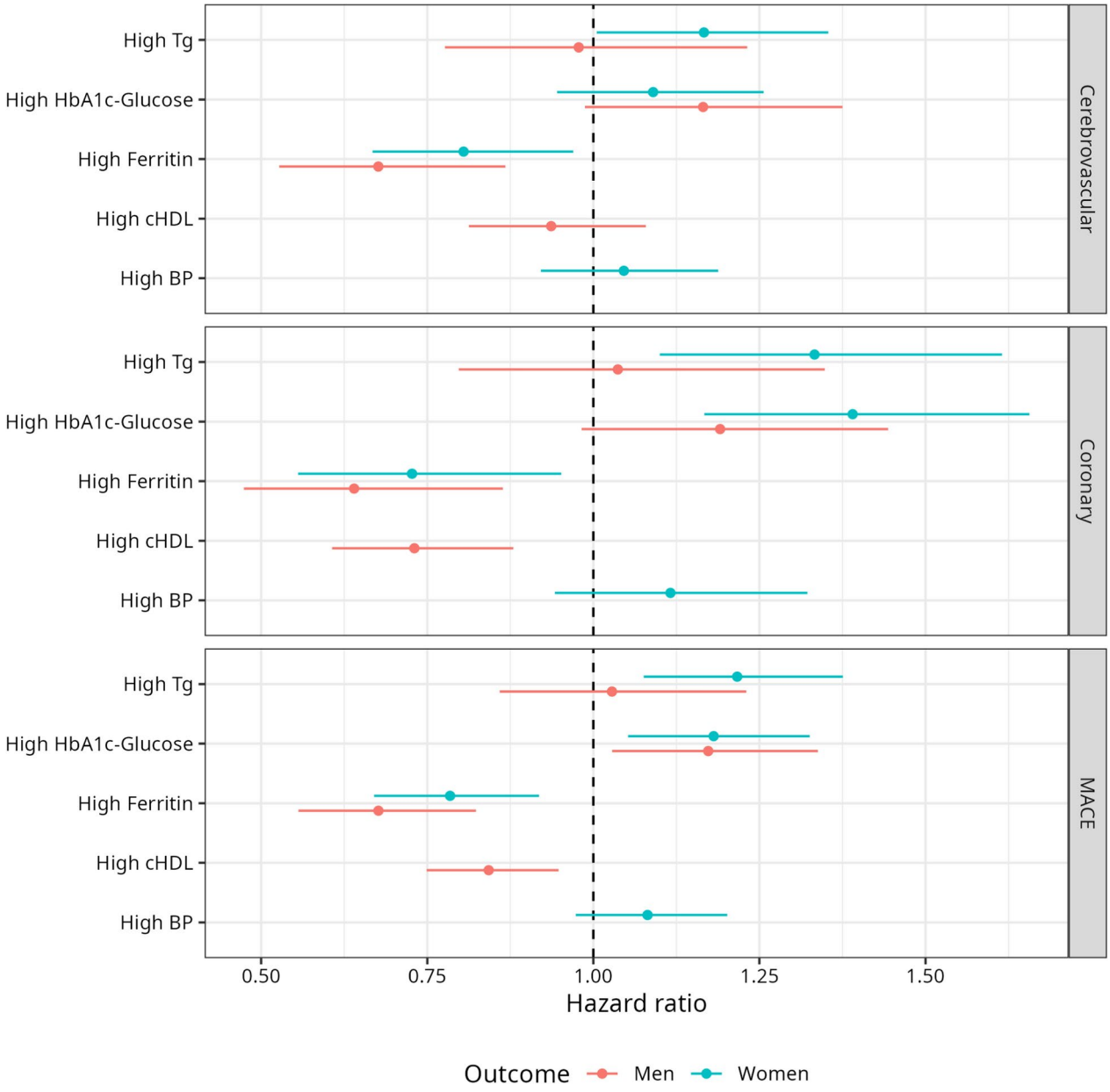


Fig. 3 Sex-stratified Hazard ratios for cardiovascular outcomes across the identified clusters. The figure shows hazard ratios (HR) for the three cardiovascular studied outcomes—cerebrovascular, coronary, and MACE events—stratified by sex

Discussion

The primary objective of this study was to explore the existence of clusters within a type 2 diabetes patient population based on clinical variables and to investigate their impact on the prevalence and incidence of coronary and cerebrovascular diseases.

Previous research has highlighted the heterogeneous nature of type 2 diabetes mellitus [27]. In our study, we identified distinct subpopulations within a type 2 diabetes mellitus cohort, characterised by varying clinical profiles and biomarker levels. These subpopulations exhibited significant differences in their risk for coronary and cerebrovascular diseases. Specifically, we found that a cluster characterised by serum ferritin levels higher than predicted for their BMI was associated with a lower incidence of CVD. This cluster showed both a significantly higher disease-free survival rate and a higher protective effect across various cardiovascular disorders compared to other clusters, mirroring observations in the cHDL cluster among men under certain conditions.

A recent bidirectional Mendelian Randomization (MR) study provides valuable insights into this relationship by investigating the causal links between iron status and obesity-related traits [15]. The study found that

genetically predicted BMI is associated with changes in iron status biomarkers. However, it did not find evidence supporting the causality of iron status on the risk of obesity. This MR study enhances our understanding of the directionality in the relationship between BMI and ferritin levels. It suggests that while obesity (as indicated by BMI) can lead to changes in iron status, changes in iron status do not causally contribute to obesity. These findings align with the observation in our research that higher ferritin levels are more likely a consequence of increased BMI rather than a cause of obesity.

Our findings suggest that individuals in the cluster with higher BMI-residualized serum ferritin levels would be able to sequester more iron and potentially mitigate oxidative stress. This contributes to a complex interplay between ferritin levels and metabolic health in the context of obesity and type 2 diabetes. On the other extreme, clusters characterised by high HbA1c and glucose levels displayed the poorest outcomes in our analysis, aligning with expectations [28]. Similarly, the cluster with elevated triglyceride levels demonstrated consistently lower survival disease-free rates, consistent with the well-established association between high plasma triglyceride levels and cardiovascular disease risk [29].

Conclusions

In conclusion, our investigation revealed distinct clusters within the type 2 diabetes mellitus population, each exhibiting unique clinical profiles and biomarker patterns. Particularly noteworthy was the cluster characterised by elevated serum ferritin levels surpassing BMI predictions, which displayed a significant protective effect across various cardiovascular disorders, underscoring its clinical importance. Our findings suggest that elevated BMI-residualized serum ferritin levels may be linked to improved cardiovascular outcomes in type 2 diabetes patients, potentially through mechanisms involving iron sequestration and the mitigation of oxidative stress.

In summary, this study emphasises the importance of considering ferritin levels in conjunction with BMI when assessing cardiovascular risk in type 2 diabetes mellitus patients. The distinct protective trend observed in the high ferritin cluster warrants further exploration to elucidate underlying mechanisms and potential therapeutic implications.

The higher intake of A10A and A10B in the elevated glucose levels cluster may indicate undertreatment, longer diabetes duration, or a lack of treatment intensification, such as the initiation of insulin therapy.

Several limitations of this study should be considered. First, our analysis was conducted on a small percentage of the original database as ferritin is not routinely measured in individuals with type 2 diabetes, and its determination is often motivated by specific conditions such as anemia, B12 deficiency, or suspected iron overload. This may introduce selection bias, as individuals included in our study may not be representative of the broader type 2 diabetes population. This is an inherent limitation of observational studies using healthcare databases. And although no gold standard assay exists for measuring ferritin levels, different methods measuring serum ferritin have been proven to be comparable [30]. Second, although we adjusted for multiple covariates, residual confounding cannot be ruled out. Unmeasured factors, such as inflammation markers or dietary iron intake, could influence the observed associations. While ferritin is widely used as a marker of iron status, its levels can be affected by inflammation, transferrin saturation, which directly reflects iron availability, could serve as a valuable complementary marker. Additionally, our study population had a relatively high average age (68 years), which may limit the generalizability of our findings to younger individuals with type 2 diabetes. Furthermore, although most women in our cohort were likely postmenopausal, we could not definitively exclude premenopausal women due to the lack of specific data on menopausal status.

Finally, it is important to acknowledge that this study is observational and should be considered hypothesis-generating, causality cannot be established. While our clustering approach identified meaningful subgroups, further validation in independent cohorts is necessary, along with mechanistic studies to clarify the underlying biological processes. Additionally, future research should explore the reproducibility of these clusters using alternative methodologies and assess their clinical utility in risk stratification and management.

Abbreviations

MACE	Major adverse cardiovascular event
HbA1c	Glycated haemoglobin
BMI	Body mass index
cHDL	High-density lipoprotein cholesterol
cLDL	Low-density lipoprotein cholesterol
Tg	Triglycerides
HR	Hazard ratio
T2D	Type 2 diabetes
CAD	Coronary artery disease
CVD	Cardiovascular disease
ROS	Reactive oxygen species
ICD-10-CM	International classification of diseases, 10th revision
ATC	Anatomical therapeutic chemical
UMAP	Uniform manifold approximation and projection
GMM	Gaussian mixture model

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12933-025-02685-w.

Additional file 1.

Author contributions
Conceptualization, J.M.F.-R., R.R., G.N.G., E.P., and D.C.; formal analysis, J.B., L.G.-N., and D.C.; resources, J.M.F.-R., R.R., and Y.L.; data curation, J.B., L.G.-N., and T.D.; writing—original draft preparation, L.G.-N., J.B., and J.M.F.-R.; writing—review and editing, L.G.-N., J.B., Y.L., R.R., D.C., T.D., G.N.G., E.P.,G.M., and C.L.R.; supervision, J.M.F.-R., R.R. and Y.L. All authors have read and agreed to the published version of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding
This work was undertaken as part of the Stratification of Obesity Phenotypes to Optimize Future Therapy (SOPHIA) project. SOPHIA has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 875534. This Joint Undertaking receives support from the European Union's Horizon 2020 Research and Innovation Program, EFPIA and T1D Exchange, the Breakthrough T1D and the Obesity Action Coalition. This work was also partially funded by Generalitat de Catalunya 2021 SGR 01263, 2014 SGR 240, 2014 SGR 902. IDIBGI is a CERCA centre from the "CERCA Programme/Generalitat de Catalunya". The CIBEROBN is an initiative from the Instituto de Salud Carlos III (ISCIII). It was developed in the context of the "Red de Investigación en Cronicidad, Atención Primaria y Prevención y Promoción de la Salud" (RICAPS, the Health Outcomes-Oriented Cooperative Research Networks (RICORS [Grant Number RD21/0016/0001]).

Availability of data and materials
The data collected for this study, including individual participant data and associated data dictionaries, are not publicly available. Access to deidentified participant data or related datasets can be requested for use in other studies, subject to approval by the principal investigator and the ethics committee. Additionally, the study protocol and statistical code are available upon request. Researchers seeking access will be required to submit a formal proposal and, if approved, enter into a data access agreement to ensure participant confidentiality and ethical use.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Diabetes, Endocrinology and Nutrition, Dr. Josep Trueta Hospital, Girona, Spain
[2]Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IDIBGI-CERCA), Girona, Spain
[3]CIBER Centro de Investigación Biomédica en Red de Fisiopatología de la Obesidad y Nutrición (CIBERobn), Madrid, Spain
[5]Grup Investigació en Salut Vascular de Girona (ISV- Girona), Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Catalunya, Spain, Catalunya, Spain
[6]Grup Investigació en Salut Vascular de Girona (ISV- Girona) , Girona Biomedical Research Institute (IDIBGI-CERCA), Dr. Josep Trueta University Hospital, Catalonia, Spain
[7]Network for Research on Chronicity, Primary Care, and Prevention and Health Promotion (RICAPPS), Girona, Spain
[8]Genetic and Molecular Epidemiology Unit, Department of Clinical Science, Lund University Diabetes Centre, Lund University, Helsinborg, Sweden
[9]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), IDIAP Jordi Gol, Barcelona, Catalunya, Spain
[10]Department of Medical Informatics, Erasmus University Medical Center, 3000 Rotterdam, The Netherlands
[11]Population Health and Genomics, University of Dundee, Dundee, UK
[12]Department of Internal Medicine, Catholic University, 00168 Rome, Italy
[13]Fondazione Policlinico Universitario A. Gemelli IRCCS, 00168 Rome, Italy
[14]Division of Diabetes and Nutritional Sciences, School of Cardiovascular and Metabolic Medicine and Sciences, King's College London, London, UK
[15]Diabetes Complications Research Centre, UCD Conway Institute of Biomedical and Biomolecular Research, School of Medicine, University College Dublin, Dublin, Ireland
[4]Department of Medical Sciences, School of Medicine, University of Girona, C/ França s/n, 17007, Girona, Spain
[16]Serveis d'Atenció Primària, Girona, Institut Català de Salut, Catalunya, Spain

Received: 23 December 2024 / Accepted: 12 March 2025
Published online: 26 March 2025

References

1. Zucatti KP, Teixeira PP, Wayerbacher LF, Piccoli GF, Correia PE, Fonseca NKO, et al. Long-term effect of lifestyle interventions on the cardiovascular and all-cause mortality of subjects with prediabetes and type 2 diabetes: a systematic review and meta-analysis. Diabetes Care. 2022;45(11):2787–95.
2. Alberti KGMM, Zimmet PZ. Diabetes: a look to the future. Lancet Diabetes Endocrinol. 2014;2(1):e1-2.
3. Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci. 2014;11(11):1185–200.
4. Liu J, Li Q, Yang Y, Ma L. Iron metabolism and type 2 diabetes mellitus: a meta-analysis and systematic review. J Diabetes Investig. 2020;11(4):946–55.
5. JeromeL S. Iron and the sex difference in heart disease risk. Lancet. 1981;317(8233):1293–4.
6. Salonen JT, Nyyssönen K, Korpela H, Tuomilehto J, Seppänen R, Salonen R. High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men. Circulation. 1992;86(3):803–11.
7. De Das S, Krishna S, Jethwa A. Iron status and its association with coronary heart disease: systematic review and meta-analysis of prospective studies. Atherosclerosis. 2015;238(2):296–303.
8. Kang P, Liu T, Tian C, Zhou Y, Jia C. Association of total iron binding capacity with coronary artery disease. Clin Chim Acta. 2012;413(19–20):1424–9.
9. Ponikowska B, Suchocki T, Paleczny B, Olesinska M, Powierza S, Borodulin-Nadzieja L, et al. Iron status and survival in diabetic patients with coronary artery disease. Diabetes Care. 2013;36(12):4147–56.
10. Grammer TB, Kleber ME, Silbernagel G, Pilz S, Scharnagl H, Tomaschitz A, et al. Hemoglobin, iron metabolism and angiographic coronary artery disease (The Ludwigshafen Risk and Cardiovascular Health Study). Atherosclerosis. 2014;236(2):292–300.
11. Hsu HS, Li CI, Liu CS, Lin CC, Huang KC, Li TC, et al. Iron deficiency is associated with increased risk for cardiovascular disease and all-cause mortality in the elderly living in long-term care facilities. Nutrition. 2013;29(5):737–43.
12. Corti MC, Guralnik JM, Salive ME, Ferrucci L, Pahor M, Wallace RB, et al. Serum iron level, coronary artery disease, and all-cause mortality in older men and women. Am J Cardiol. 1997;79(2):120–7.
13. Ristow M. Unraveling the truth about antioxidants: mitohormesis explains ROS-induced health benefits. Nat Med. 2014;20:709–11.
14. Prillaman M. Why BMI is flawed—and how to redefine obesity. Nature. 2023;622(7982):232–3.
15. Zhou Z, Zhang H, Chen K, Liu C. Iron status and obesity-related traits: a two-sample bidirectional Mendelian randomization study. Front Endocrinol. 2023;14(14):985338.
16. Moreira AC, Mesquita G, Gomes MS. Ferritin: an inflammatory player keeping iron at the core of pathogen-host interactions. Microorganisms. 2020;8(4):589.
17. Mahroum N, Alghory A, Kiyak Z, Alwani A, Seida R, Alrais M, et al. Ferritin—from iron, through inflammation and autoimmunity, to COVID-19. J Autoimmun. 2022;126:102778.
18. Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data resource profile: the information system for research in primary care (SIDIAP). Int J Epidemiol. 2022;51(6):e324–36.
19. Raventós B, Fernández-Bertolín S, Aragón M, Voss EA, Blacketer C, Méndez-Boo L, et al. Transforming the information system for research in primary care (SIDIAP) in Catalonia to the OMOP common data model and its use for COVID-19 research. Clin Epidemiol. 2023;15:969–86.
20. Ramos R, Balló E, Marrugat J, Elosua R, Sala J, Grau M, et al. Validez del Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP) en el estudio de enfermedades vasculares: estudio EMMA. Rev Esp Cardiol. 2012;65(1):29–37.
21. Ramos R, Comas-Cufí M, Martí-Lluch R, Balló E, Ponjoan A, Alves-Cabratosa L, et al. Statins for primary prevention of cardiovascular events and mortality

in old and very old adults with and without type 2 diabetes: retrospective cohort study. BMJ. 2018;5:k3359.
22. Mata-Cases M, Mauricio D, Real J, Bolíbar B, Franch-Nadal J. Is diabetes mellitus correctly registered and classified in primary care? A population-based study in Catalonia. Spain Endocrinol Nutr. 2016;63(9):440–8.
23. Coral DE, Smit F, Farzaneh A, Gieswinkel A, Tajes JF, Sparsø T, et al. Subclassification of obesity for precision prediction of cardiometabolic diseases. Nat Med. 2024. https://www.nature.com/articles/s41591-024-03299-7
24. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2024.
25. Melville J. uwot: the uniform manifold approximation and projection (UMAP) method for dimensionality reduction. 2019. p. 0.2.2. https://CRAN.R-project.org/package=uwot
26. Therneau TM. Survival: survival analysis. 2001. p. 3.8–3. https://CRAN.R-project.org/package=survival
27. Birkeland KI, Kilhovd B, Thorsby P, Torjesen PA, Ganss R, Vaaler S, et al. Heterogeneity of non-insulin-dependent diabetes expressed as variability in insulin sensitivity, β-cell function and cardiovascular risk profile. Diabet Med. 2003;20(1):37–45.
28. Nishimura R, Nakagami T, Sone H, Ohashi Y, Tajima N. Relationship between hemoglobin A1c and cardiovascular disease in mild-to-moderate hypercholesterolemic Japanese individuals: subanalysis of a large-scale randomized controlled trial. Cardiovasc Diabetol. 2011;10(1):58.
29. Padro T, Muñoz-Garcia N, Badimon L. The role of triglycerides in the origin and progression of atherosclerosis. Clín Investig Arterioscler. 2021;33:20–8.
30. Garcia-Casal MN, Peña-Rosas JP, Urrechaga E, Escanero JF, Huo J, Martinez RX, et al. Performance and comparability of laboratory methods for measuring ferritin concentrations in human serum or plasma: a systematic review and meta-analysis. PLoS ONE. 2018;13(5):e0196576.

Publisher's Note