BMJ Open

Characterisation of type 2 diabetes subgroups at diagnosis: the COPERNICAN prospective observational cohort study protocol

Berta Fernandez-Camins,[1,2,3] Bogdan Vlacho ,[3,4] Albert Canudas,[5] Marta Ortega,[6] Minerva Granado-Casas,[3,4,7,8] Alexandre Perera-LLuna,[9,10] Alejandro Boluda-Sanson,[1] Yesmina El-Khattabi-Ofkir,[8] Josep Franch-Nadal,[2,3,4] Didac Mauricio ,[1,3,4,11] COPERNICAN Research Group

To cite: Fernandez-Camins B, Vlacho B, Canudas A, *et al*. Characterisation of type 2 diabetes subgroups at diagnosis: the COPERNICAN prospective observational cohort study protocol. *BMJ Open* 2024;**14**:e083825. doi:10.1136/bmjopen-2023-083825

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-083825).

Received 30 December 2023
Accepted 10 October 2024

For numbered affiliations see end of article.

Correspondence to
Dr Didac Mauricio;
didacmauricio@gmail.com

ABSTRACT

Introduction Type 2 diabetes mellitus (T2DM) is a highly heterogeneous and complex metabolic disease harbouring different metabolic characteristics. Adequate characterisation of subjects is essential to allow the implementation of precision medicine for the prevention, diagnosis, prognosis and treatment of this condition.

Methods and analysis This prospective observational cohort study aims to identify and characterise relevant clinical clusters that are reproducibly associated with various clinical outcomes in T2DM in our Mediterranean region. The COPERNICAN study will include 1200 subjects with newly diagnosed T2DM from 28 primary care centres from the city of Barcelona and the healthcare district of Lleida in Catalonia (Spain). Participants will undergo a comprehensive phenotypic evaluation including, among others, six relevant variables: age, antibodies against glutamic acid decarboxylase, body mass index, glycated haemoglobin (HbA1c), indexes of insulin sensibility (HOMA2-IR) and secretion (HOMA2-beta). We will collect additional comprehensive data on glucose-lowering and other drug treatments, clinical evaluation (including complications), laboratory parameters, advanced lipoprotein profile, dietary habits and physical activity. The linkage with the population database will be done to perform a pragmatic follow-up of participants as part of their usual clinical care. A state-of-the-art cluster analysis (k-means and hierarchical clustering) will be performed.

Ethics and dissemination The present study complies with all the ethical aspects and protection of participant subjects complying with all current local and European Union legislation. All Ethics Committees from the institutions involved in the study (IR Sant Pau Ethics Committee, Ethics Committee for Drug Research at IDIAP Jordi Gol and University Hospital of Bellvitge Ethics Committee for Research) approved this protocol. Confidentiality and anonymity of the data are ensured according to the current Spanish Organic Law 3/2018 of 05 December.

Trial registration number ClinicalTrials.gov. registration number NCT05333718, 27 January 2023.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Multicentre prospective cohort study of people with newly diagnosed type 2 diabetes mellitus (T2DM).
⇒ Comprehensive phenotyping to identify and better characterise T2DM subgroups; real-world data-enriched assessment.
⇒ Patients with T2DM with longer disease duration will not be included.
⇒ Non-supervised cluster analysis could vary, depending on the number and type of clinical variables included.

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is an increasingly prevalent condition that reduces the quality of life and life expectancy of people affected, mainly due to the occurrence of chronic complications.[1] It is a highly heterogeneous metabolic disease with a complex aetiology involving a combination of different pathophysiological, environmental and genetic factors. In addition, the clinical characteristics of the disease and its progression may vary greatly between subjects, leading to insufficient metabolic control. However, the current classical T2DM diagnostic criteria seem to be poorly designed for capturing the heterogeneity of each subject clinical characteristics, the disease course and the response to medications and the risk of complications.[2] Therefore, T2DM heterogeneity remains an issue when it comes to making decisions regarding the management and prognosis of people with this condition. Being able to subclassify patients with T2DM into groups might allow a suitable precision medicine approach to diabetes management, thereby identifying subsets of patients who are at the highest risk for disease progression

or particular complications and/or patients most likely to benefit from particular therapeutic management strategies.[2 3]

The majority of current subclassification approaches use either a genetic or phenotypic approach. Genetic subtype approaches have included polygenic scores that aim to better define diabetes subtypes. In contrast to monogenic and T1DM, full and cluster-specific polygenic scores for T2DM have yet to show a sufficient predictive ability to warrant their use in routine clinical practice.[4 5] Identifying more precise genetic subtypes of T2DM may become more predictive in future iterations, enabling targeted therapies.[6] The main phenotypic approach was proposed by Ahlqvist *et al* in 2018, which focused on five clinical subtypes of T2DM that differed according to disease progression and risk of diabetic complications using six clinical variables at the time of diagnosis (anti-glutamic acid decarboxylase (GAD) autoantibodies, age at diagnosis, body mass index (BMI), glycated haemoglobin (HbA1c) at diagnosis and homeostatic model assessment estimates of insulin secretion capacity (HOMA2) and beta and insulin resistance (HOMA2-IR)).[7] The characteristics of these clusters have been successfully replicated in other independent cohorts using different clustering methods and variables, arguably becoming a well-replicated approach.[8–14] More recently, studies have further looked at the clinical utility of phenotypic stratification approaches using clinical trial data to identify subtypes and clinical features associated with response to treatment.[13 15]

Overall, identifying clinically relevant subgroups of T2DM in relation to outcomes may lead to a better understanding of the patients in our care in our region and allow us to implement different precision medicine strategies in our country. It is important to establish strategic initiatives to help us test and implement all the new potential precision medicine tools for possible prevention, diagnosis, prognosis and diabetes treatment. Thus, this project is relevant as a pioneer study in our country.

The study's aims and hypothesis

We aim to identify clinically relevant clusters in patients with T2DM in our Mediterranean region (Catalonia, Northeast Spain) which differ according to outcomes such as clinical features, disease progression, onset of comorbidities and complications, treatment response and lipoprotein profile. We will initially use the following variables at the time of diagnosis to identify clusters: age, anti-GAD antibodies, BMI, HbA1c, homeostatic model and others to estimate beta cell function and insulin resistance through C peptide quantification. The outcomes that we will analyse will include the following: clinical characteristics at the time of diagnosis, such as sociodemographic characteristics (sex, age, education level and deprivation level), associated comorbidity frequency, including micro- and macrovascular complications, presence of non-alcoholic fatty liver disease, subclinical femoral and carotid atherosclerosis, advanced lipoprotein profile, lifestyle variables (dietary habits and physical activity), use of antidiabetic drugs and other treatments, as well as diabetes-related metabolomics and genetic profile. During the follow-up, we aim to study the progression of the disease and the appearance of different diabetes-related micro- and macrovascular complications.

Hypothesis

The main hypothesis of this study is that there is heterogeneity in T2DM subjects in our Mediterranean region, similar to those previously described in other European populations. We also hypothesise that the different identified subgroups will show a different profile regarding other clinical characteristics at the time of diagnosis.

METHODS AND ANALYSIS

Settings

We will create a prospective cohort of people with recently diagnosed T2DM from different primary healthcare centres in Catalonia, Northeast Spain. The study recruitment started on 4 March 2022 and is planned to be finalised on 31 December 2024. Currently, during the recruitment period, 28 healthcare centres are involved in the recruitment of the participants: 2 reference hospitals and 26 primary healthcare centres from different public healthcare providers.

Eligibility criteria

The inclusion criteria are as follows: people from both sexes and at least 18 years of age who attended healthcare centres from two healthcare districts, that is, the city of Barcelona and the healthcare region of Lleida, recently diagnosed T2DM (with a maximum disease duration of 3 months at inclusion), with a diagnosis established according to the American Diabetes Association criteria[1] and providing informed consent by signing the informed consent form. People with a diagnosis of other types of diabetes (type 1, MODY, gestational or by other causes) will be excluded.

Sample size and power calculations

This is a descriptive study without a confirmatory analysis of a specific hypothesis. Therefore, the sample size has not been calculated according to one expected hypothesis or one specific desirable association measure. Additionally, we have assessed the number of subjects in previous studies similar to the current one in other countries;[8 10] they all used a similar approach with a similar or lower sample size than the one proposed here. We plan to recruit 1200 participants across 28 healthcare centres over a 24–36-month period. This estimate has been based on an expected incidence of four T2DM per 1000 person-years according to our own study of our population, as part of a multinational collaboration.[16 17]

Patient and public involvement

None are involved.

Patient recruitment

Preselection

All potential candidates will be identified by their healthcare professionals (physicians and nurses) in the participating sites. Once the study criteria are confirmed, potential candidates will be invited to participate and informed about the study details. Once the participant agrees to participate, a specific preselection study visit will be scheduled, where study criteria will be again confirmed, and the participant will provide written informed consent.

Inclusion

Each patient providing written informed consent will be included. Once informed consent is granted, study procedures and data collection will be done by completing an electronic case report form). At the inclusion visit, the investigators will perform a standard medical exam. Lifestyle-related questionnaires will also be self-administered during the visit. An appointment will be made for routine clinical blood and urine parameters as recommended by current guidelines within 2 weeks of the initial visit. Biological blood and urine samples will be shipped and stored at the biobanking facilities for future diabetes-related studies. Routine screening for complications will be carried out including foot and eye assessments. Additionally, the participant's consent will be obtained for future follow-up through extraction of clinical data from their electronic medical records through the population databases available in our system (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP)/Programa d'analítica de dades per a la recerca i la innovació en salut (PADRIS)).

Follow-up visits and discontinuation

The follow-up of the participants will be done at 12 and 24 months through linkage to the above-mentioned databases. We have planned a long-term continuous follow-up beyond the study duration. In case of premature discontinuation from the study, the reason for discontinuation and data will be collected.

Baseline and follow-up measures

At inclusion, we will collect data on sociodemographic data (age, sex, ethnicity, country of origin and educational and deprivation levels), toxic habits (alcohol, as explained below; and tobacco and drug use), clinical data (weight; height; BMI; waist circumference; blood pressure; ankle–brachial index; monofilament, as explained below; vibration perception, as explained below; pedal pulses and fundus photography), specific data on T2DM (date of diagnosis, initial symptoms of hyperglycaemia and family history of T2DM), comorbidities of interest (both relevant clinical comorbidities such as hypertension and dyslipidaemia as well as the presence of complications of diabetes, including microvascular and macrovascular complications), standard laboratory data, pharmacological treatment (glucose-lowering treatment and all other concomitant medications), diet and physical activity. Additionally, in the healthcare district of Lleida, a specific substudy for subclinical carotid and femoral atherosclerosis will be done.

Alcohol consumption will be differentiated between risk consumption and absence of risk consumption in which risk consumption has been defined as the following: adult population (≥18 years of age) with consumption of ≥4 standard drink units per day (SDU/day) in men and ≥2–2.5 SDU/day in women. Any consumption in a population with comorbidities, activities or treatment that would advise against consumption will also be considered risk consumption. Monofilament evaluation will be performed in four testing sites: plantar surface of the first, third and fifth metatarsals and plantar surface of the first toe. Vibration perception will be assessed by using a standard tuning fork on the surface of the peroneal malleolus to evaluate the presence or absence of sensibility. See table 1 for more details on the variables to be collected.

During the follow-up, we will collect data on toxic habits, clinical data, new comorbidities or complications of diabetes, concomitant medication and, in case of premature discontinuation, the reason and date for the discontinuation including any causes of death. The study visits and procedure plan can be found in table 2.

Main study variables

The main study variables are age, anti-GAD antibodies, HbA1c, BMI, homeostatic model 2 to assess beta cell function (HOMA2-beta) and the homeostatic model 2 to assess insulin resistance (HOMA2-IR).[7]

Diabetes-related autoantibodies (anti-GAD and IA-2) will be measured using two commercially available ELISA kits: GAD65 autoantibodies ELISA and IA-2 autoantibodies ELISA.[18 19] Optimal cut-off values for positivity are set at 5 and 15 U/mL for GAD65 and IA-2Ab, respectively.

HOMA2-IR and HOMA2-beta will be calculated using the following calculator: Diabetes Trials Unit, University of Oxford: HOMA-2 Calculator.[20] C-Peptide will be determined through the high-sensitivity direct chemiluminescent immunoassay CLIA analyser (LIASON XL 'Flash' chemiluminescence technology, CLIA) and compatible assay kit.[21]

Physical activity questionnaires

Physical activity will be evaluated using the Spanish short version of the International Physical Activity Questionnaire (IPAQ). The short IPAQ form contains seven questions that evaluate four domains: leisure physical activity, domestic and gardening activities, work-related physical activity and transportation-related physical activity.[22] The activity is then classified into low, medium or high, depending on the energetic consumption estimated for each activity.

Dietary questionnaires

Adherence to the Mediterranean diet will be evaluated with the 14-item Mediterranean Diet Adherence Screener

Table 1 Description of study variables

Variable	Definition
Age*	Years
Sex*	Male, female
Ethnicity*	Country of origin
Educational level*	International Standard Classification of Education
Deprivation level*	MEDEA Index‡[29]
Tobacco use*	Smoker, non-smoker, former smoker
Alcohol intake*	Risk consumption, absence of risk consumption
Drug use*	Consumer, non-consumer
Weight*	Kilograms (kg)
Height*	Metres (m)
Body mass index*	Kilograms/metre^2
Waist circumference*	Centimetres
Blood pressure*	Systolic blood pressure (SBP) and diastolic blood pressure (DBP) in mm Hg
Ankle–brachial index*	Ankle SBP divided by arm SBP
Monofilament test*	Evaluation of the four points for the Semmes Weinstein monofilament test[30]
Vibration perception[31] *	Present/absent
Pedal pulses*	Present/absent
Fundus photography*†	Normal/signs of diabetic retinopathy
Date of diagnosis*	DD/MM/YYYY
Initial symptoms of hyperglycaemia*	Weight loss, polyuria, polydipsia, polyphagia, presence of ketosis, hospital admission due to decompensation, hyperglycaemia
Family history of diabetes*	Presence of and number of family members (from parents, siblings and offspring) with a history and type of diabetes
Dyslipidaemia*†	CIE10 codes: E78.xx
Hypertension*†	CIE10 codes: I10.xx, I12.xx, I13.xx, I15.xx
Peripheral artery disease*†	CIE10 codes: I70.xx, I73.xx
Heart failure*†	CIE10 codes: I50.xx
Ischaemic heart disease*†	CIE10 codes: I20.xx, I25.xx
Stroke*†	CIE10 codes: G45.xx, G46.xx, I63.xx, I69.xx
Liver disease (excluding steatosis)*†	CIE10 codes: K70.xx, K71.xx, K72.xx, K73.xx, K74.xx, K75.xx, K77.xx
Liver steatosis*†	CIE10 codes: K75.81, K76.xx
Renal insufficiency*†	CIE10 code: N18
Acute pancreatitis*†	CIE10 codes: K85.xx
Diabetic nephropathy*†	CIE10 codes: E11.2, N08.3, R80
Diabetic neuropathy*†	CIE10 codes: E11.4, G59, G56, G62, G63, G90, G99
Diabetic retinopathy*†	CIE10 codes: E11.4, G59, G56, G62, G63, G90, G99
Glucose-lowering drugs*†	ATC/DDD: A10, posology, frequency, dispensing.
Lipid-lowering drugs*†	ATC/DDD: C10, posology, frequency, dispensing.
Antiplatelet agents*†	ATC/DDD: B01AC, posology, frequency, dispensing.
Anticoagulant drugs*†	ATC/DDD: B01AA, posology, frequency, dispensing.
Anti-hypertensive agents*†	ATC/DDD: C07, posology, frequency, dispensing.
Standard laboratory parameters*†	Glucose, HbA1c, creatinine, urate, urine albumin/creatinine ratio, blood count, AST, ALT, GGT, TSH, free T4, total cholesterol, cLDL, cHDL, triglycerides
Antibodies to islet cell antigens*	anti-GAD, anti-IA2
Beta-cell function and insulin sensitivity indexes*	HOMA2-IR and HOMA2-beta, C-peptide

*Source: assessment at inclusion visit.
†Source: follow-up data from eCAP/SIDIAP/PADRIS.
‡The MEDEA (Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales) Index is a deprivation index constructed by census tract.
ALT, alanine aminotransferase; AST, aspartate aminotransferase; ATC, anatomic, therapeutic and chemical drug classification; cHDL, high-density lipoprotein cholesterol; CIE10, International Classification of Diseases 10th version; cLDL, low-density lipoprotein cholesterol; DDD, defined daily dose; eCAP, Primary Care Clinical Station (Estació Clínica d'Atenció Primària); GAD, glutamic acid decarboxylase antibodies; GGT, gamma-glutamyl transferase; HbA1c, glycated haemoglobin; HOMA2-beta, Homeostatic Model Assessment of beta-cell function; these two variables will be calculated using Diabetes trials Unit University of Oxford HOMA Calculator; HOMA2-IR, Homeostatic Model Assessment for Insulin Resistance 2; IA2, tyrosine phosphatase-related islet antigen 2 antibodies; PADRIS, data analytics programme for health research and innovation (programa d'analítica de dades per a la recerca i la innovación en salut); SIDIAP, Information System for the Development of Research in Primary Care (Sistema d'Informació pel Desenvolupament de la Investigació en Atenció Primària); T4, thyroxine; TSH, thyroid-stimulating hormone.

questionnaire.[23] This questionnaire is an adaptation of a previously validated 9-item questionnaire used in the study's control group PREDIMED-Plus.[24] It contains 12 questions about the frequency of consumption of different foods, and two of those questions relate to the

Table 2 Variables and procedure plan throughout the study performance

	Preselection	Inclusion visit	SIDIAP/PADRIS electronic medical record follow-up
Selection criteria review	x	x	
Informed consent		x	
Sociodemographic data*		x	
Clinical data†		x	x
History of T2DM		x	
Comorbidities (see table 1)		x	x
Specific mortality			x
Laboratory data‡		x	x
Smoking, alcohol or drug use		x	
Pharmacological treatment§		x	x
Diet and physical activity		x	
Carotid and femoral ultrasound¶		x	
Biological samples stored in biobanking facility		x	
Advanced lipoprotein profile (Liposcale)[32]		x	

*Age, gender and primary care area.
†Weight, height, waist circumference, SBP/DBP, heart rate and the Mediterranean Diet Adherence Screener questionnaire.
‡The main variables will be collected during the inclusion visit, and the rest of the laboratory data will be collected through the SIDIAP/PADRIS database.
§Pharmacological treatment will be collected during the inclusion visit and at follow-up through SIDIAP/PADRIS.
¶This will be registered during the inclusion visit only in centres from the Lleida healthcare area.
DBP, diastolic blood pressure; PADRIS, data analytics programme for health research and innovation (*programa d'analítica de dades per a la recerca i la innovación en salut*); SBP, systolic blood pressure; SIDIAP, Information System for the Development of Research in Primary Care (*Sistema d'Informació pel Desenvolupament de la Investigació en Atenció Primària*); T2DM, type 2 diabetes mellitus.

regular consumption of typical Spanish Mediterranean diet foods.

Subclinical femoral and carotid atherosclerosis

Previous studies from our group have shown the early presence of subclinical atheromatous disease in individuals with T2DM and pre-diabetes.[25 26] A substudy on the prevalence of carotid and femoral subclinical atherosclerosis will be performed with an expected recruitment of 509 participants in the Lleida healthcare area. Carotid and femoral ultrasound examination will be done by B-mode ultrasound and colour Doppler to identify early asymptomatic subjects with signs of femoral or carotid atherosclerosis, following standard procedures conducted in other studies previously performed by our group.[26] The presence of atheromatous plaques will be assessed in 10 vascular territories: internal, external, bulb and common carotid and common and superficial femoral arteries. Atheromatous plaque is defined as a focal wall thickening encroaching into the arterial lumen by at least 50% of the surrounding intima-media thickness value or with a thickness of 1.5 mm at minimum measuring from the media adventitia interference to the intima–lumen surface, according to the American Society of Echocardiography Consensus Statement[27] and the Mannheim Consensus.[28]

Analysis plan

A cluster analysis (k-means and hierarchical clustering) following the method described by Ahlqvist *et al* will be performed to analyse the collected and cleaned data.[7] Male and female subjects will be grouped separately to avoid stratification regarding gender-related differences in the group variables. The mean and standardised deviation will be calculated for continuous variables; continuous variables greater than 5 SD from the mean will be excluded. Categorical variables will be included as binary variables. The two-step cluster analysis will be performed, in which first, the optimal number of clusters is estimated, and subsequently, a hierarchical clustering is performed using record probability as a measure of distance and the Schwarz Bayesian inference criterion for clustering. The stability of the groups will be evaluated by resampling the data set 2000 times and calculating the Jaccard similarity coefficient to the original group. Generally, stable clusters should produce a Jaccard similarity coefficient over 0.75. Once the clusters are defined, we will assign the same names used in the original study, depending on the distribution of the cluster characteristics. The statistical analysis will be performed using R data analysis software (version>4.20.1), and cluster analysis will be performed using the flexible procedures for clustering pack, version>2.2–10. Each subject will be assigned to a group according to the Euclidean distance to each group centre. A descriptive analysis will be used to evaluate differences between subtypes, tested using Fisher's exact test, ANOVA test and post hoc comparisons. A p value lower than 0.05 will be considered statistically significant in the phenotype analysis and will be corrected by multiple comparisons. Furthermore, Benjamini-Hochberg's correction will be used to determine the significance of multiple tests (by variables). After the recruiting period, a comparative analysis will be performed to ensure that there are no significant biases found between patients who participated and those who did not agree to join the study. This will be done using the Hanlon method.

DISCUSSION

The COPERNICAN study is a multicentre prospective observational study with the aim of characterising people with T2DM at diagnosis in a Mediterranean region to identify relevant comprehensive clinical subgroups associated with differences in clinical outcomes. We expect to complete the recruitment by the end of 2024 and initial follow-up at the end of 2026; however, we plan to follow the whole cohort in the long term. The unique combination of clinical data collected at inclusion and during the follow-up and the state-of-the-art cluster analysis will allow us to better understand the subtypes of this complex disease in our country.

We chose a prospective observational study design and follow-up based on real-world clinical care at clinical sites to reflect as much as possible the usual care and management of patients with T2DM in our healthcare settings. Moreover, including reference hospitals and primary care centres in recruiting the study participants might increase the validity of the study. Participants are being/will be selected using a consecutive sampling approach for each patient being seen at the clinic as part of their routine clinical care to control T2DM, based on availability and willingness to take part in the study. Moreover, we intend to reduce the possibility of any clinical or therapeutic intervention since the site clinical researchers adhere to routine practice according to the local T2DM guidelines and recommendations. Only a few laboratory parameters have been added to the usual routine lab assessment (peptide C, antigens anti-GAD and IA-2) to ensure that we are not including subjects with other types of diabetes. Samples for future genetic studies will also be collected. Additionally, information about the prescribed treatment will be prospectively collected. The questionnaires regarding diet and physical activity will be self-administered at the inclusion visit.

Limitations

One of the limitations may be the relative representativeness of the study participants. We opted for a realistic approach that allows both viability and sufficient representativeness. This proposal addresses this issue and may allow us to cooperate in this field with other national and international researchers. There is a possibility of a low rate of recruitment due to different reasons such as research site investigator motivation and usual healthcare work burden. The absence of rural primary care teams among the participating centres might be a limitation regarding its representativeness of the Catalonian population. Missing data could also occur since this is a prospective observational study; however, we have adopted a pragmatic risk-based continuous monitoring of the collected data to help us prevent or diminish the risk of this limitation. Another limitation lies in the cluster analysis, an exploratory unsupervised analysis. Some forms of cluster analysis may produce slightly different results each time the statistical analysis is run.

Ethics and dissemination

The present study complies with all the ethical requirements and with the protection of the participant subjects as it complies with the legal ethical Spanish regulation (Real Decreto 957/2020 and the Biomedical Investigation Law 14/2007, 3 July) and is conducted according to the Declaration of Helsinki in its last revision. Since this is an observational study, it is considered that the patients' participation carries minimal risk. This risk is similar to the one the patient would have throughout usual clinical practice without participating in the study. This protocol was evaluated by the Ethics Committee (CEI) before the inclusion of the patients (Hospital de la Santa Creu i Sant Pau Ethics Committee, ref. HSCSP 21/369 (OBS) dated 17 July 2021.; Ethics Committee for drug research (CEIm) at IDIAP Jordi Gol, CEIm ref. 21/266-P dated 15 December 2021; University Hospital of Bellvitge Ethics Committee for Research, ref. PR438/21 (CSI 21/86) dated 09 February 2022). The study protocol was registered on ClinicalTrials.gov with Trial Registration Number NCT05333718. Any data required by the protocol could be subject to audits by the sponsor. The results of this study will be published irrespective of the outcomes.

Data deposition

For this study, a database has been created as per the local regulatory legal framework (Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales). The investigators and/or authorised personnel will record all the study data on an electronic case report form (eCRF).

Author affiliations
[1]Department of Endocrinology and Nutrition, IR Sant Pau, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[2]Department of Medicine, University of Barcelona, Barcelona, Spain
[3]DAP_CAT group, Unitat de Suport a la Recerca Barcelona Ciutat, Institut Universitari d'Investigació en Atenció Primària Jordi Gol, IDIAP Jordi Gol, Barcelona, Spain
[4]CIBER of Diabetes and Associated Metabolic Diseases, Instituto de Salud Carlos III, Barcelona, Spain
[5]Escola de Doctorat. Departament de Medicina, Infermeria i ciències de la Salut, University of Lleida, Lleida, Spain
[6]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, IDIAP Jordi Gol, Barcelona, Spain
[7]Department of Nursing and Physiotherapy, University of Lleida, Lleida, Catalunya, Spain
[8]Health Care Research Group (GRECS), Lleida Institute of Biomedical Research Foundation Dr Pifarré, Lleida, Spain
[9]Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine, CIBER-BBN, Madrid, Spain
[10]B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Polytechnic University of Catalonia, Barcelona, Spain
[11]Faculty of Medicine, University of Vic - Central University of Catalonia, Vic, Spain

Collaborators COPERNICAN Research Group: Berta Fernandez-Camins, Bogdan Vlacho, Albert Canudas, Marta Ortega, Minerva Granado-Casas, Alexandre Perera-LLuna, Alejandro Boluda-Sanson, Yesmina El Khattabi Ofkir, Josep Franch-Nadal, Dídac Mauricio, Idoia Genua Trullos, Inka Miñambres, Natalia Riera, Antonia Valsera Robles, M Eulalia Pradas Sala, Norma Ballestar Palacios, Olga Vazquez Soler, Cristina Ledesma Serrano, Ana Cámara Caño, Teresa Villuendas, Begona Ribas Lopez, Ana Maria Pedro Pijoan, Eva Gonzalez Platas, Eva Garcia, Marta Hernández, Raquel Martí, Eva Serra Llavall, Montserrat Torra Solé, Maite Puig Solé, Ana Isabel Areny Tallaví, Marc Olivart, Neus Miro Vallve, Àngels Molló, Marta Puig Fito, Jesús Pujol Salud, Cristina Garcia Serrano,Eva Miquel Fernàndez, Lidia Aran Sole, Antonieta Lafarga, Cristina Dominguez Amador, Sandra Guerrero, Cristina Farras Salles, Maria Boldú Franque, Lorena Arnay, Mireia Marin, Laura Turo Pere, Anna Maria Torné Cortés, Laia Llubes Arrià, Cecilia Bañeres Argiles, Assumpció Florensa Flix, Paula Garcia, Manuel Mata Cases, Carla Muñoz, Isabel Bobe, Elena Hernández Boluda, María Begoña Cañas Diaz, Xavier Peligros Palma, Francisco Cegri Lombando, Francesc Xavier Cos Claramunt, Anna Martinez Sanchez, Gabriel Cuatrecasas, Anna Amorós, Blanca Simon, Anna Massana Raurich, Maria Isabel Prieto Fernández, Maria Olga Alvarez Fernández, Robert Cabanes Gómez, Ascensión Niubo Lopez, Marta Crespo Boixasa, Maria Miñana Castellanos, Assumpció Altaba Barcelo, Estel Felez Carroba, Carme Olius Peyrats, Eva Marsol Prieto, Eva Ferris Gallart, Mari Carmen Ruiz Salido, Alicia Mostazo Muntané, Marta Tàrrega Rubio, Maria Lluïsa Calonge Carbonell, Roser Ros Barnadas, Elena Artal Traveria, Zulema Martí Oltra, Carolina Lapena Estella, Olga Poveda Jovellar, Ana Carrera Rodriguez,

Sonia Aldea Gomez, Xavier Herraiz Fabregat, Elisabet Llop Lozano, Ignacio Alborch Simó, Laura Güell Espigol, Pilar Pedro Benages, Javier Parramon Nuñez, Paloma Prats, Berta Tió, Carlos Arias, Pelayo Martinez, Cristian Llacer Pinos, Roser Ferrer Costa, Luz Maria Cruz Carlo.

Contributors DM and JF-N contributed to the conception and design of the study. DM is the scientific coordinator of the project. JF-N and MO are local coordinators for the project for primary care centres from Barcelona and Lleida. DM designed the hypothesis and objectives. BV and MG contributed to the study supervision and execution. AC, AB-S, YE-K-O and COPERNICAN Research Group are responsible for data collection. AP-L will be conducting the formal analysis. BF-C and BV developed the draft manuscript. All authors critically revised the manuscript and approved the submitted version. DM and JF-N share the senior authorship. DM is the guarantor.

Funding This research received financial support from the Spanish Ministry of Economy and Competitiveness (www.mineco.gob.es) PID2021-122952OB-I00, DPI2017-89827-R and Instituto de Salud Carlos III, Ministry of Health (PI21/01163). This work has been financed by a PERIS grant from the Primary Care Oriented Research Projects program with the reference number SLT021/21/000030 granted by the Generalitat de Catalunya (2022-2024). CIBER for Diabetes and Associated Metabolic Diseases (CIBERDEM; leading group CB15/00071) and the Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), are initiatives of Instituto de Salud Carlos III (ISCIII)

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Not applicable.

ORCID iDs
Bogdan Vlacho http://orcid.org/0000-0003-3768-8112
Didac Mauricio http://orcid.org/0000-0002-2868-0250

REFERENCES

1 American Diabetes Association Professional Practice Committee. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes - 2022. *Diabetes Care* 2022;45:17–38.
2 Deutsch AJ, Ahlqvist E, Udler MS. Phenotypic and genetic classification of diabetes. *Diabetologia* 2022;65:1758–69.
3 Chung WK, Erion K, Florez JC, *et al*. Precision medicine in diabetes: a Consensus Report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* 2020;63:1671–93.
4 Goodrich JK, Singer-Berk M, Son R, *et al*. Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. *Nat Commun* 2021;12:3505.
5 DiCorpo D, LeClair J, Cole JB, *et al*. Type 2 Diabetes Partitioned Polygenic Scores Associate With Disease Outcomes in 454,193 Individuals Across 13 Cohorts. *Diabetes Care* 2022;45:674–83.
6 Dornbos P, Koesterer R, Ruttenburg A, *et al*. A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. *Nat Genet* 2022;54:1609–14.
7 Ahlqvist E, Storm P, Käräjämäki A, *et al*. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–9.
8 Zaharia OP, Strassburger K, Strom A, *et al*. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 2019;7:684–94.
9 Zou X, Zhou X, Zhu Z, *et al*. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *The Lancet Diabetes & Endocrinology* 2019;7:9–11.
10 Bello-Chavolla OY, Bahena-López JP, Vargas-Vázquez A, *et al*. Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Res Care* 2020;8:e001550.
11 Anjana RM, Baskar V, Nair ATN, *et al*. Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. *BMJ Open Diabetes Res Care* 2020;8:e001506.
12 Slieker RC, Donnelly LA, Fitipaldi H, *et al*. Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* 2021;64:1982–9.
13 Dennis JM, Shields BM, Henley WE, *et al*. 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* 2019;7:442–51.
14 Tanabe H, Saito H, Kudo A, *et al*. Factors Associated with Risk of Diabetic Complications in Novel Cluster-Based Diabetes Subgroups: A Japanese Retrospective Cohort Study. *J Clin Med* 2020;9:2083:1–12:.
15 Pigeyre M, Hess S, Gomez MF, *et al*. Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia* 2022;65:206–15.
16 Rojo-Martínez G, Valdés S, Soriguer F, *et al*. Incidence of diabetes mellitus in Spain as results of the nation-wide cohort di@bet.es study. *Sci Rep* 2020;10:2765.
17 Magliano DJ, Chen L, Islam RM, *et al*. Trends in the incidence of diagnosed diabetes: a multicountry analysis of aggregate data from 22 million diagnoses in high-income and middle-income settings. *Lancet Diabetes Endocrinol* 2021;9:203–11.
18 Chen S, Willis J, Maclean C, *et al*. Sensitive non-isotopic assays for autoantibodies to IA-2 and to a combination of both IA-2 and GAD65. *Clin Chim Acta* 2005;357:74–83.
19 Palomer X, Mauricio D, Rodríguez-Espinosa J, *et al*. Evaluation of two nonisotopic immunoassays for determination of glutamic acid decarboxylase and tyrosine phosphatase autoantibodies in serum. *Clin Chem* 2004;50:1378–82.
20 Hermans MP, Levy JC, Morris RJ, *et al*. Comparison of tests of beta-cell function across a range of glucose tolerance from normal to diabetes. *Diabetes* 1999;48:1779–86.
21 Hörber S, Orth M, Fritsche A, *et al*. Comparability of C-Peptide Measurements - Current Status and Clinical Relevance. *Exp Clin Endocrinol Diabetes* 2023;131:173–8.
22 Román Viñas B, Ribas Barba L, Ngo J, *et al*. Validación en población catalana del cuestionario internacional de actividad física. *Gac Sanit* 2013;27:254–7.
23 García-Conesa M-T, Philippou E, Pafilas C, *et al*. Exploring the Validity of the 14-Item Mediterranean Diet Adherence Screener (MEDAS): A Cross-National Study in Seven European Countries around the Mediterranean Region. *Nutrients* 2020;12:2960:1–18:.
24 Martínez-González MA, Fernández-Jarne E, Serrano-Martínez M, *et al*. Development of a short dietary intake questionnaire for the quantitative estimation of adherence to a cardioprotective Mediterranean diet. *Eur J Clin Nutr* 2004;58:1550–2.
25 Sánchez E, Betriu À, López-Cano C, *et al*. Characteristics of atheromatosis in the prediabetes stage: a cross-sectional investigation of the ILERVAS project. *Cardiovasc Diabetol* 2019;18:154.
26 Rubinat E, Marsal JR, Vidal T, *et al*. Subclinical Carotid Atherosclerosis in Asymptomatic Subjects With Type 2 Diabetes Mellitus. *J Cardiovasc Nurs* 2016;31:E1–7.
27 Stein JH, Korcarz CE, Hurst RT, *et al*. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine. *J Am Soc Echocardiogr* 2008;21:93–111; .
28 Palanca A, Castelblanco E, Betriu À, *et al*. Subclinical atherosclerosis burden predicts cardiovascular events in individuals with diabetes and chronic kidney disease. *Cardiovasc Diabetol* 2019;18:93.
29 Domínguez-Berjón MF, Borrell C, Cano-Serral G, *et al*. Constructing a deprivation index based on census data in large Spanish cities(the MEDEA project). *Gac Sanit* 2008;22:179–87.
30 Bedogni G, Bellentani S, Miglioli L, *et al*. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006;6:33.
31 Reggars JW. Vibratory Sensation Testing. *COMSIG Rev* 1995;4:14–5.
32 Mallol R, Amigó N, Rodríguez MA, *et al*. Liposcale: a novel advanced lipoprotein test based on 2D diffusion-ordered 1H NMR spectroscopy. *J Lipid Res* 2015;56:737–46.