

BMJ Open SALMANTICOR study. Rationale and design of a population-based study to identify structural heart disease abnormalities: a spatial and machine learning analysis

Jose Ignacio Melero-Alegria,¹ Manuel Cascon,¹ Alfonso Romero,² Pedro Pablo Vara,¹ Manuel Barreiro-Perez,¹ Victor Vicente-Palacios,¹ Fernando Perez-Escanilla,³ Jesus Hernandez-Hernandez,¹ Beatriz Garde,¹ Sara Cascon,⁴ Ana Martin-Garcia,¹ Elena Diaz-Pelaez,¹ Jose Maria de Dios,⁵ Aitor Uribarri,¹ Javier Jimenez-Candil,¹ Ignacio Cruz-Gonzalez,¹ Baltasara Blazquez,⁶ Jose Manuel Hernandez,⁶ Clara Sanchez-Pablo,¹ Inmaculada Santolino,⁷ Maria Concepcion Ledesma,⁸ Paz Muriel,² P Ignacio Dorado-Diaz,¹ Pedro L Sanchez¹

To cite: Melero-Alegria JI, Cascon M, Romero A, *et al.* SALMANTICOR study. Rationale and design of a population-based study to identify structural heart disease abnormalities: a spatial and machine learning analysis. *BMJ Open* 2019;**9**:e024605. doi:10.1136/bmjopen-2018-024605

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-024605>).

Received 11 June 2018
Revised 6 October 2018
Accepted 18 December 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Pedro L Sanchez;
pedrolsanchez@secardiologia.es

ABSTRACT

Introduction This study aims to obtain data on the prevalence and incidence of structural heart disease in a population setting and, to analyse and present those data on the application of spatial and machine learning methods that, although known to geography and statistics, need to become used for healthcare research and for political commitment to obtain resources and support effective public health programme implementation.

Methods and analysis We will perform a cross-sectional survey of randomly selected residents of Salamanca (Spain). 2400 individuals stratified by age and sex and by place of residence (rural and urban) will be studied. The variables to analyse will be obtained from the clinical history, different surveys including social status, Mediterranean diet, functional capacity, ECG, echocardiogram, VASERA and biochemical as well as genetic analysis.

Ethics and dissemination The study has been approved by the ethical committee of the healthcare community. All study participants will sign an informed consent for participation in the study. The results of this study will allow the understanding of the relationship between the different influencing factors and their relative importance weights in the development of structural heart disease. For the first time, a detailed cardiovascular map showing the spatial distribution and a predictive machine learning system of different structural heart diseases and associated risk factors will be created and will be used as a regional policy to establish effective public health programmes to fight heart disease. At least 10 publications in the first-quartile scientific journals are planned.

Trial registration number NCT03429452.

Strengths and limitations of this study

- To obtain data on the prevalence and incidence of structural heart disease in the setting of a population-based study enrolling a total of 2400 individuals, stratified by age, sex and by place of residence (rural and urban), in a Spanish community.
- To create a population-based established control group providing availability of normative reference values quantification for echocardiographic, ECG, VASERA, biochemical and genetic parameters.
- To show the spatial distribution of the different patterns of structural heart disease through the spectrum of age and sex and between urban and rural residences.
- To develop a predictive model of structural heart disease using cardiovascular heterogeneous data (including images and machine learning techniques).
- To establish the study as the global observatory on cardiovascular health research and development of the regional healthcare government to support effective public health programme implementation.

INTRODUCTION

Each year heart diseases cause almost 4 million deaths in Europe and the USA, that is, one out of four deaths.^{1 2} Although the number of deaths from heart disease has decreased, the burden of heart disease is increasing. In 2015, more than 85 million people in Europe were living with cardiovascular disease.² The increases in the prevalence of classical cardiovascular risk factors, dietary factors, physical activity and probably other social factors make the largest contribution to the risk of heart

disease. Overall, cardiovascular disease healthcare costs in the European Union and the USA have increased rapidly over the last 10 years; currently surpassing €200 billion a year.^{2,3}

In this sense, public health delivery planning requires reliable information about contemporary population-level disease prevalence and incidence. Furthermore, community healthcare systems should obtain and provide their own data before implementing any effective health programme as these regional systems are highly influenced by geographical diversity, the availability of resources and infrastructure, and the characteristics of healthcare systems and patterns of reimbursement.⁴ This is well illustrated by the attention of myocardial infarction where the exchange of accurate and timely information between the healthcare community, decision-makers and the public programme effects has been essential.⁵⁻⁸

Policies need to consider both standardised rates, which describe disease prevalence and incidence independently of changes in population, and absolute numbers of patients affected, which describe the impact of the disease on the population, political commitment, resources and services of interest.^{4,9} Limited data exist on estimation of heart disease prevalence in a population setting. Previous studies have frequently been based on selected cohorts, which may not represent the general population.¹⁰⁻¹³ Other studies have restricted case identification to those made in general practice consultations or hospital admissions.¹⁴⁻¹⁶ However, it is only by considering presentations across the whole spectrum of structural heart disease that the full burden of the disease can be captured and an accurate distinction can be made between the incident and prevalent cases. Thus, contemporary population-based studies of heart disease prevalence and incidence are needed to inform resource planning and research prioritisation but current evidence is scarce.

Spatial analysis is a great tool to investigate population behaviour, relations and consequently determine future action plans or policies. Spatial methods are varied, ranging from descriptive spatial analysis to complex interpolation algorithms. Gaussian process (GP) procedures, such as cokriging, have distinct advantages over conventional spatial prediction techniques.¹⁷ They allow researchers to include measured spatial variability in the geostatistical estimation process and they smooth predicted values based on the proportion of total sample variability accounted by random noise. Furthermore, GP helps mitigate the effect of variable sample density caused by hot spots (some zones are usually oversampled). Hence, geostatistical techniques are suitable methods to apply on population studies.

Furthermore, the volume of quantitative and imaging data, generated by population studies, will also be a key driver in the future for research and how we provide care. In this sense, machine learning (ML) to train algorithms to recognise cardiac damage on a better level, avoiding diagnostic errors and improving the early identification of the disease offers new approaches to leveraging

the increasing volume of data available for analyses.¹⁸⁻²¹ Thus, we are convinced that ML can play a key role in population-based epidemiological studies when trying to recognise patients disease vulnerability earlier.

The objectives of this study are: to obtain data on the prevalence and incidence of structural heart disease in a population setting; to show the spatial distribution of the different patterns of structural heart disease through the spectrum of age and sex and between urban and rural; to develop a predictive model of structural heart disease using cardiovascular heterogeneous data (including images and ML techniques); to generate new hypotheses which might contribute to healthcare research and to political commitment to obtain resources and support effective public health programme implementation.

In this article, we describe the design, data and imaging acquisition, analysis methods and quality assurance metrics for the SALMANTICOR study.

METHODS

Study design and participants

The SALMANTICOR study is a cross-sectional descriptive population-based study of the prevalence of structural heart disease and their risk factors that will enrol a total of 2400 individuals, stratified by age, sex and by place of residence (rural and urban), in a Spanish community: Salamanca. Structural heart disease refers to any of the following heart abnormalities including congenital heart disease, cardiomyopathies, valvular heart disease, ischaemic heart disease, pericardial diseases and rhythm or conduction disorders.

The province of Salamanca is located on the Western Spain, bordered in the West by Portugal. It has an area of 12.349 km² and had a population of 342 857 people in 2014; 167 459 (49%) male and 175.398 (51%) female citizens. It is divided into 362 municipalities; more than half are villages with fewer than 300 people. In fact, 227 878 (67%) people live in 10 municipalities of more than 5000 individuals that will be considered for future analysis as urban areas and 114 581 (33%) people live in the rest of municipalities and consequently will be considered as rural areas.

Spain's and consequently Salamanca's healthcare system is public, guaranteeing universal coverage. In total, 98.7% of the population are insured for this public Spanish healthcare system. In Salamanca, a total of 35 primary health centres throughout the province provide healthcare services to the overall population: 18 to the urban-considered municipalities and 17 to the rural-considered municipalities (figure 1).

Individuals aged ≥18 years included in the lists of all primary healthcare facilities of the province of Salamanca represented the reference population of 295 975 subjects: mean age 52.9±19.8 years; 52.4% females; 61.3% residing in urban areas. A sample size of 2400 subjects is calculated based on an expected prevalence of structural heart disease of 6% with a CI of 95% and a 1%

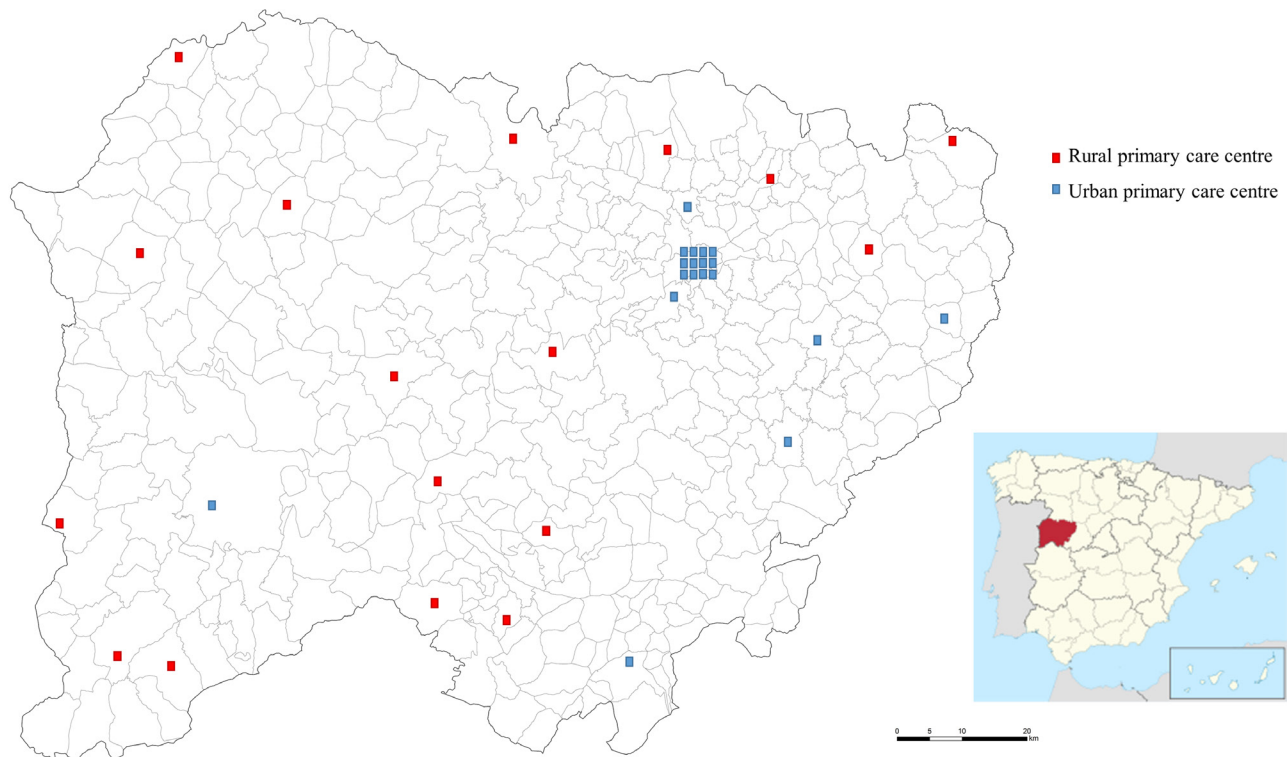


Figure 1 Province of Salamanca map and distribution of the total of 35 primary health centres: 18 in urban-considered municipalities (blue) and 17 in rural-considered municipalities (red). Municipalities of more than 5000 individuals are considered as urban areas in the SALMANTICOR study.

precision. In order to obtain the necessary sample size, 35% more requests for participation will be made, estimating errors of location from the healthcare database or refuses to participate in the study. Thus, 3564 people will be randomly selected from the primary care lists.

Cohort participants will undergo a basal examination visit, in these primary healthcare centres, between 2015 and 2018. Surviving participants are expected to return for a 5 and 10-year follow-up visit. Institutional review committee approval was obtained and all participants will provide informed consent. The SALMANTICOR study is designed to provide echocardiographic parameters characterising cardiac structure and function in all individuals. SALMANTICOR participants will undergo surveillance for cardiovascular events, including heart failure, incident coronary heart disease and all-cause mortality.

Medical investigation process

Medical history, surveys completion and examinations will be obtained at the subject's primary care referral centre and will be analysed and interpreted centrally at the University Hospital of Salamanca. A complete medical history, physical examination and the surveys completion checkout will be performed by a cardiologist in a separate office, where examinations and blood sample extraction will be performed. Echocardiographic measures will be initially performed. Participant's blood pressure and VASERA measures will be taken within 30 min after starting the echocardiographic examination

and after the subject will be resting for 10 min. ECG will be performed after VASERA to finalise with the blood sample extraction.

Questionnaires

After obtaining written informed consent, trained interviewers will use a structured questionnaire to collect baseline data in face-to-face interviews at the time of physical examination. Self-reported diseases will be verified by individuals' primary care doctors according to recognised international standards. The questionnaire will collect information on demographics and cardiovascular risk factors, cardiovascular and non-cardiovascular medical history, physical examination, medication, socioeconomic status, dietary habits as well as lifestyle and physical activity (table 1).

Echocardiographic assessment

A standardised echocardiography ultrasound examination, including M-mode, two dimensional (2D), spectral, colour flow and tissue Doppler, will be performed by a certified technical professional using Philips CX-50 scanner with a standard 2.5–3.5 MHz phased-array probe. Image acquisition will be performed using a preprogrammed acquisition protocol (table 2); following the American and European Society of Echocardiography recommendations.^{22–24} All studies will be acquired and stored digitally on a local picture archiving and communication system and transferred from field primary care

Table 1 Questionnaires

Name of the questionnaire	No of variables	Principal variables	Time of completion
Demographics and Cardiovascular risk factors	12	Sex, age, residence, smoking, alcohol consumption, hypertension, hypercholesterolaemia, diabetes, previous heart disease, family history	5 min
Cardiovascular and non-cardiovascular history	23	Coronary heart disease, arrhythmias, valvulopathies, heart failure, cardiac healthcare visits in the past and where (public or private attention), stroke, vascular peripheral disease, bleeding history, chronic kidney disease, chronic lung disease, asthma, rheumatic disease, depressive disorder, dementia, anxiety, dependency	12 min
Physical examination	8	Body mass index, abdominal perimeter, heart rate, oxygen saturation, blood pressure, heart murmurs and sounds	8 min
Medication	24	Aspirin, clopidogrel, ticagrelor, prasugrel, warfarin, acenocumarol, dabigatran, rivaroxaban, apixaban, edoxaban, betabloquers, ACE inhibitors, RAAS antagonists, calcium channel blocker, diuretics, aldosterone inhibitors, statin, ezetimibe, fibrates, ivabradine, ranolazine, proton-pump inhibitor, NSAIDs, corticoids	10 min
Socioeconomic status	13	Marital status, education, employment, annual income, homeownership, housing quality, medical coverage	8 min
Dietary habits and lifestyle	39	No of meals, diet, beverage, salt, bread, olive oil, coffee, chocolate and potatoes dietary counselling, Mediterranean diet adherence, no of sleeping hours, siesta practice, pet ownership	12 min
Physical activity	7	No of days, no of hours, intensity	5 min
Total	126		60 min

ACE, Angiotensin-converting enzyme; NSAIDs, nonsteroidal anti-inflammatory drugs; RAAS, renin-angiotensin-aldosterone system.

centres to a secure server at the Salamanca University Hospital on the same day via a dedicated virtual private network connection. Development of the imaging and analysis protocol, field centre echocardiography manual of operations, reading centre manual of operations, field centre sonographer, training of sonographer occurred from July 2015 to October 2015, followed by the initiation of the SALMANTICOR visit in November 2015, which was continued until May 2018.

For patients in sinus rhythm, >3 full cardiac cycles will be recorded for each view, with recording beginning once the view is optimised. For subjects in atrial fibrillation, >5 acquisitions per view will be recorded. Sonographers are instructed to continuously optimise both imaging depth and sector width to maintain a frame rate of 50–80 frames per second. Sonographers are also instructed to adjust 2D gain and compression, when necessary, to optimally demonstrate left ventricle endocardial borders. The colour Doppler Nyquist limit will be set at 64 cm/s. Colour Doppler gain will be set just below the level at which random background noise will be seen. Sonographers will optimally align spectral Doppler parallel to the direction of the blood flow of interest. Sonographers

will optimise the baseline shift and velocity range so that the spectral envelope will occupy approximately three-fourths of the display. All spectral Doppler acquisitions will be performed with a sweep speed between 75 and 100 cm/s, and a sample volume length of 3 mm for pulsed-wave Doppler. The tissue Doppler sample volume will be placed at the level of an annulus (mitral and tricuspid) and the baseline shift and velocity range will be optimised. All tissue Doppler acquisitions will be performed with similar acquisitions of spectral Doppler with a filter setting of 100 Hz.

Echocardiograms will be obtained at the subject's primary care referral centre and sonographers will not perform any measurements on the images obtained because all measurements will be analysed and interpreted centrally at the University Hospital of Salamanca. All SALMANTICOR echocardiograms will be read by a certified cardiologist and over-read by a board-certified cardiologist with expertise in echocardiography variables assessment (table 3). Over-reads of echocardiograms will be performed to confirm the accuracy of key quantitative measurements and to identify clinically important findings. Inter and intrareader reproducibility was assessed

Table 2 Echocardiographic imaging protocol required views

Parasternal position	
Parasternal long axis	Two-dimensional (2D) imaging (at deep depth) 2D imaging (at shallow depth) Colour Doppler of the mitral and aortic valves (AVs)
Parasternal short axis, AV level	2D imaging of AV Colour Doppler of AV 2D imaging of right ventricular outflow tract (RVOT) Colour Doppler of RVOT Pulsed-wave (PW) and Continuous-wave (CW) Doppler of RVOT
Parasternal short axis, mitral valve level	2D imaging
Parasternal short axis, left ventricle apex	2D imaging
Apical position	
Apical four-chamber view	2D imaging 2D imaging, focused/zoomed of left ventricle (LV) 2D imaging, focused on left atrium Colour Doppler of mitral valve/left atrium PW Doppler of mitral flow CW Doppler of mitral flow Tissue Doppler imaging (TDI) of septal and lateral mitral annulus
Apical four-chamber view, focused on the right ventricular	2D imaging Colour Doppler of tricuspid valve/right atrium CW Doppler of tricuspid regurgitation TDI of lateral tricuspid annulus
Apical five-chamber view	2D imaging Colour Doppler of left ventricular outflow tract (LVOT) PW of LVOT flow CW of transaortic flow
Apical two-chamber view	2D imaging 2D imaging focused/zoomed on LV 2D imaging focused on left atrium Colour Doppler mitral valve/left atrium
Apical three-chamber view	2D imaging 2D imaging focused/zoomed on LV 2D imaging focused on left atrium Colour Doppler mitral valve/left atrium Colour Doppler of AV PW of LVOT flow CW of transaortic flow
Subcostal view	
Inferior vena cava	2D imaging (5 s acquisition)

before initiating the trial. For inter-reader reproducibility, intraclass correlation values ranged from 0.85 to 0.99 with left atrial volume and left ventricular end-diastolic volumes having the highest intraclass correlation values (0.97–0.99). Intraclass correlation values were slightly better from intrareader assessments for all measures.

Vascular function assessment

Cardio-Ankle Vascular Index (CAVI), brachial-ankle pulse wave velocity (baPWV) and Ankle-Brachial Index

(ABI) will be estimated using the VaSera VS-1500 device (Fukuda Denshi) as described by our group.²⁵ The baPWV will be calculated, as well as CAVI, which provides a more accurate estimation of the atherosclerosis degree. CAVI integrates cardiovascular elasticity derived from the aorta to the ankle pulse velocity through an oscillometric method; it is used as a good measure of vascular stiffness and does not depend on blood pressure.²⁶ CAVI values will be automatically calculated by substituting the stiffness parameters in the following equation to detect the vascular elasticity and the baPWV; where p is the blood density, P_s and P_d are systolic blood pressure and diastolic blood pressure in mm Hg, respectively, and baPWV is measured between the aortic valve and ankle.

$$\text{stiffness parameter } \beta = 2p \times \frac{1}{(P_s - P_d)} \times \left(\frac{P_s}{P_d} \right) \times \text{baPWV}^2$$

The average coefficient of the variation of CAVI is <5%, which is small enough for clinical use and confirms that CAVI has favourable reproducibility.^{27 28} CAVI and ABI will be measured in the resting position. baPWV is estimated using the following equation; where t_{ba} is the time, the same waves were transmitted to the ankle.

$$\text{baPWV} = \frac{(0.5934 \times \text{height [cm]} + 14.4724)}{t_{ba}}$$

For the study, the lowest ABI and the highest CAVI and baPWV obtained will be considered. CAVI is classified as normal (CAVI <8), borderline (8 ≤ CAVI <9) and abnormal (CAVI ≥ 9). Abnormal CAVI represents subclinical atherosclerosis, and baPWV ≥ 17.5 is considered abnormal.^{29 30} ABI ≤ 0.9 is considered abnormal.

ECG examination

ECG examination will be performed using a General Electric MAC 3500 ECG System (Niskayuna, New York, USA), which automatically measures wave voltage and duration. ECG will be performed by the same nurse trained to carefully standardised procedures for ECG acquisition. The standard 12-lead ECGs will be obtained at a paper speed of 25 mm/s, an amplitude of 10 mm/1 mV and a filter range 0.04–40 Hz from all patients. ECG tracing will be interpreted in a similar way to the echocardiographic protocol by an independent cardiologist and over-read by a board-certified cardiologist with expertise in ECG at the University Hospital of Salamanca. ECG measurements and interpretations will be done following standard methods^{31 32} (table 4).

Laboratory test

Venous blood sampling will be performed at the end of the examination after participants have fasted and abstained from smoking, consumption of alcohol and caffeinated beverages for 12 hours, following the protocol used in our hospital for other multidisciplinary projects.²⁵ A total of 20 mL of venous blood will be drawn for research testing. Blood will be drawn as follows: ethylenediaminetetraacetic acid (EDTA) 10 mL and serum 10 mL.

Table 3 Echocardiographic parameters

Structure and function assessment	No of variables	Principal variables
Aorta and atria and ventricles	39	Ascending aorta (mm), left ventricular diastolic dimension (mm), LV systolic dimension (mm), left ventricular mass index (g/m^2), left atrial volume index by biplanar Simpson method (mL/m^2), right ventricular diastolic dimension (mm), right atrial volume index (mL/m^2), biplanar Simpson left ventricular ejection fraction (%), mitral E-wave (cm/s), mitral A-wave (cm/s), mitral E/A, mitral deceleration time (cm/s), pulmonary artery systolic pressure (mm Hg), mitral E/e'septal annulus, mitral E/e'lateral annulus, mitral E/e'average of annulus
Valves	41	Aortic valve jet peak velocity (m/s), aortic mean gradient (mm Hg), aortic cups number, aortic valve calcification, aortic regurgitation presence and grade, mitral valve calcification, mitral mean gradient (mm Hg), mitral pressure half time (ms), mitral prolapse, mitral regurgitation presence and grade, tricuspid regurgitation presence and grade, pulmonary regurgitation presence and grade
Pericardium	3	Pericardial effusion presence and grade

Aliquots of plasma (3×2 mL), serum (4×2 mL) and white cell pellet (3×2 mL) will be stored in freezers (−80°C) until the analysis. All biomaterial (serum, plasma and white blood cells) will be stored in the Instituto de Investigación Biomédica de Salamanca biobank. Referral for biobanking is carried out through a specific electronic database. Biochemical tests include N-terminal pro-brain natriuretic peptide (NT-proBNP), troponin, haemoglobin, blood cell count, thrombocytes, ferritin and iron, transferrin and iron saturation, potassium, sodium and creatinine, glycated haemoglobin, plasma glucose, aspartate aminotransferase, alanine aminotransferase, total cholesterol, triglycerides, high-density lipoprotein (HDL) and low-density lipoprotein (LDL), uric acid, high-sensitive C reactive protein, thyroid-stimulating hormone. Further, biomarkers indicative of different pathophysiological mechanisms relevant to heart disease will be analysed. A white cell pellet will be used for genotyping.

Results and outcomes

After the clinical history is performed and the echocardiogram and ECG are interpreted, a clinical report is sent to the patient and to the primary care medical doctor. Individuals needing a further evaluation will be sent to the cardiology department through a preference standardised protocol.

Individuals will be contacted at 5 years intervals to ascertain the clinical status and to repeat the described basal evaluations. Clinical outcomes will include cardiovascular major adverse cardiac events (MACE), commencing dialysis and first hospitalisation.

Statistical analysis

Casual and multivariate inference

Data input will be stored in a database designed for the project. Normal distribution of variables will be verified using the Kolmogorov-Smirnov test. Quantitative variables will be displayed as mean±SD if normally distributed or as the median (IQR) if asymmetrically distributed

and qualitative variables will be expressed as frequencies. Analysis of the difference of means between variables of two categories will be carried out using a Student's t-test or a Mann-Whitney U test, as appropriate, while qualitative variables will be analysed using a χ^2 test. To analyse the relationship between qualitative variables of more than two categories and quantitative variables, an analysis of variance and the least significant difference test will be used in the post hoc tests. The relationship of quantitative variables to each other will be tested using Pearson's or Spearman's correlation as appropriate. Analysis of covariance will be performed to adjust the variables that can affect the results as confounders. A multivariate analysis of variance will be performed in cases with more than one dependent variable to identify whether changes in the independent variables have significant effects on the dependent variables. The association between the variables studied will be performed by multiple linear regression. Data will be analysed using the SPSS V.23.0 statistical package (SPSS). A $p < 0.05$ will be considered as statistically significant.

Spatial analysis

Additionally, this research aims having a spatial understanding of the structural heart disease abnormalities in the province of Salamanca. Such a demanding task will be carried out by applying different statistic procedures as multiple factor analysis (MFA) and cokriging.

MFA is an extension of principal component analysis (PCA) tailored to handle distinct variables (quantitative, categorical or frequency) and different data tables collected on the same observations.³³ MFA is put into practice depending on the data tables and the variables types: in the case of quantitative variables a PCA is applied; multiple correspondence analysis (MCA) is applied in case of categorical variables³⁴ and correspondence analysis (CA) for frequency variables.³⁵ Cokriging is a multivariate geostatistical procedure used for interpolation

Table 4 12-lead ECG parameters

Rhythm	Sinus rhythm Auricular tachycardia Atrial fibrillation Common atrial flutter Uncommon atrial flutter Nodal rhythm Atrial ectopics Ventricular ectopics Atrial paced rhythm Ventricular paced rhythm with sinus activity Ventricular paced rhythm with atrial fibrillation Atrial and ventricular paced rhythm
Heart rate	
P wave	P duration Sinus P morphology Pulmonary P morphology Interatrial block
PQ time	
Atrioventricular (AV) block	Not present First-degree AV block Second-degree AV block, Mobitz I Second-degree AV block, Mobitz II 2:1 AV block Third-degree or complete AV block
QRS duration	
QRS axis	
RR time	
QT time	
QT corrected time	
Brugada pattern	Not present Type I Type II Type III
Early repolarisation pattern	Not present Inferior Lateral Inferior and lateral
Bundle branch configuration	Not present Complete left bundle branch block Complete right bundle branch block Incomplete left bundle branch block Incomplete right bundle branch block
Intraventricular conduction disturbances	
Fascicular block configuration	Not present Left anterior fascicular block Left posterior fascicular block
Notch QRS presence	
Left ventricular hypertrophy	
Delta waves presence	
Repolarisation changes of digitalis	
Pathological Q-waves presence and position	
Significant ST elevation	
Significant ST depression	
Negative T-waves presence and position	

purposes.³⁶ This method is a generalisation of a multivariate linear-weighted regression model, where weights depend on distance, direction and orientation of the neighbouring data to the unsampled location.

In the SALMANTICOR study, we will further combine MFA and cokriging. In our case, we have two different levels of observations, participants and municipalities. As a mathematical comparison, municipalities contain participants, therefore, if we want to extend our investigation to a spatial analysis we need to use the resulting MFA projections on their corresponding municipality areas and then apply a cokriging analysis on the unsampled municipalities (figure 2) (online supplementary data). This combination will provide a spatial understanding of the Salamanca population and will cover the whole analysis, however, if we want to focus on a specific questionnaire, we could skip the MFA and look at the results obtained from the MCA, PCA or CA and then apply a cokriging analysis. In addition, if we require analysing a particular item from a questionnaire, we could also perform the analysis. To summarise, we have a versatile methodology that permits to study as concrete aspects as a wider analysis of our study.

The R packages FactoMineR and Gstat will be used in order to apply MFA and cokriging, respectively.^{37,38} An additional code will be shared in a public Github repository.

Machine learning

The SALMANTICOR study will also be analysed following the ML pipeline represented in figure 3. Our first step will consist in the development of scalable methods for ML optimisation with the aim to develop a first approach to the predictive structural heart disease model. Our ML model will start from ingesting raw data, leveraging data processing techniques to wrangle, process and engineer meaningful features and attributes from this data (feature engineering). The derived features are attributes or properties shared by all the independent units on which analysis or prediction is to be done. In our case, clinical variables and variables quantified from imaging data will be chosen. Features will be combined with scalable ML algorithms, including deep learning process and automatic extraction of data functionalities, in order to develop the model (fit model). The model's basic behaviour and functionalities will be tested to develop a robust and reliable model (training model). We will validate, train and improve the ML model in a trial an error process until satisfactory model performance (validation). The SALMANTICOR study sample will be randomly divided into a train dataset (70% of the sample) and a validation dataset (30% of the sample), following previous published ML models.³⁹ We will use our train dataset to fit our ML model and the validation dataset to evaluate our results. This process will be repeated multiple times to guarantee a robust fit without overfitting. We will build our predictor models using: random forest, gradient boosting, logistic regression, K-nearest neighbours, support vector machine, linear discriminant analysis and naive Bayesian network

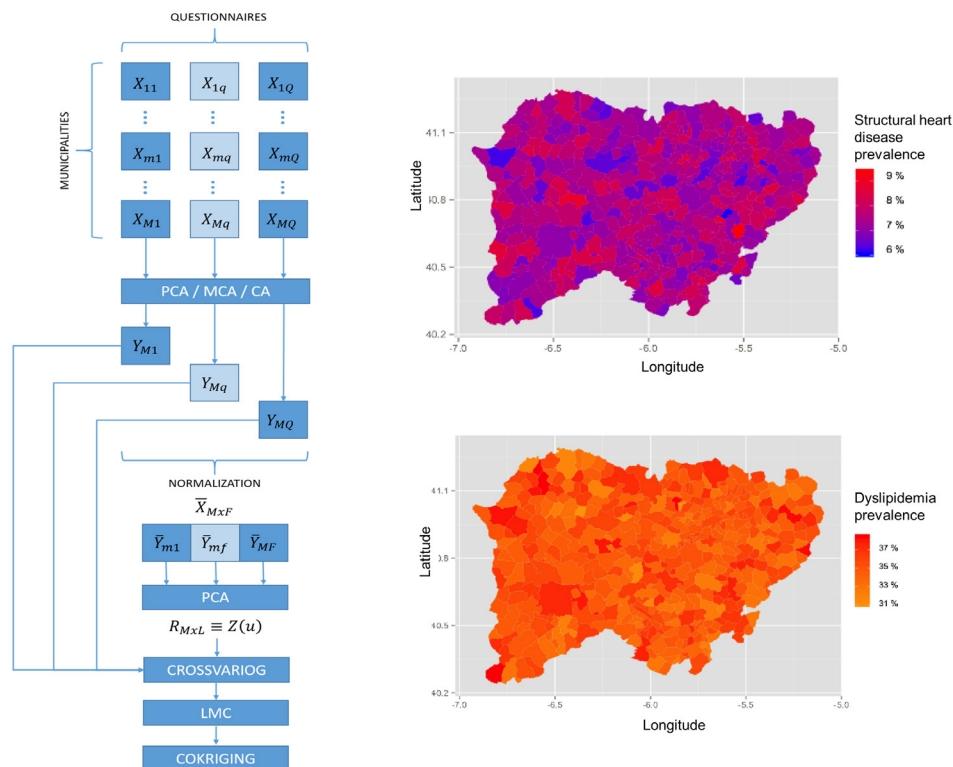


Figure 2 The left panel represents the spatial analysis pipeline that SALMANTICOR will use for map plotting purposes. We will combine multiple factor analysis and cokriging. We will inquire and analyse participants from municipalities and questionnaires. Initially, for quantitative variables principal component analysis (PCA) is applied; for categorical variables, multiple correspondence analysis (MCA); and for frequency variables, correspondence analysis (CA). We will then assemble the normalised data in a single table that is analysed via PCA to describe the spatial behaviours of our samples within crossvariograms (crossvariog). We then will apply a linear model coregionalization (LMC) to finally interpolate the results over the different municipalities of the province of Salamanca using cokriging. Maps in the right panel represent municipal spatial patterns examples of how we will represent municipal (Salamanca is divided into 362 municipalities) distribution of structural heart disease and dyslipidaemia prevalence.

models (online supplementary data). Our ML pipeline set-up will compare the performance of each algorithm on the dataset using a set of carefully selected evaluation criteria (ie, classification accuracy, logarithmic loss, confusion matrix, area under curve, F1 score, mean absolute error, mean squared error) and the categorisation of the specific cardiac problem.

For the realisation of these ML models, we will use free software (Python) and free open-source unified workbench such as Scikit-learn.⁴⁰

Quality control

Different processes will be carried out to guarantee study data quality and thus maximise the validity and reliability of measurements of the results. To this effect, field work operation manuals have been prepared. These documents specify the adequate procedure for performing each test. All of these actions will confirm adequate performance of each procedure. Monthly meetings will be held with the principal investigator of the study to analyse the entire process, and an annual report on study progress will be prepared.

Ethical review board and dissemination plan

Participants will be required to sign an informed consent form prior to participation in the study, in accordance with the Declaration of Helsinki and WHO standards for observational studies. Participants will be informed of the objectives of the project and of the risks and benefits of the examinations made. None of the examinations poses life-threatening risks for the type of participants to be included in the study. The study includes the obtaining of biological samples (including genetics analysis); the study participants will, therefore, be informed in detail. The confidentiality of the recruited participants will be ensured at all times in accordance with the provisions of current legislation on personal data protection (15/1999 of December 13), and the conditions contemplated by Act 14/2007 on biomedical research.

We will use a variety of methods to ensure that our work will achieve maximum visibility. Publication of our study protocol provides an important first step towards this direction. In this paper, we have sought to offer a comprehensive overview of relevant literature, while underlining current research gaps that necessitated the design and implementation of the SALMANTICOR study. Similarly,

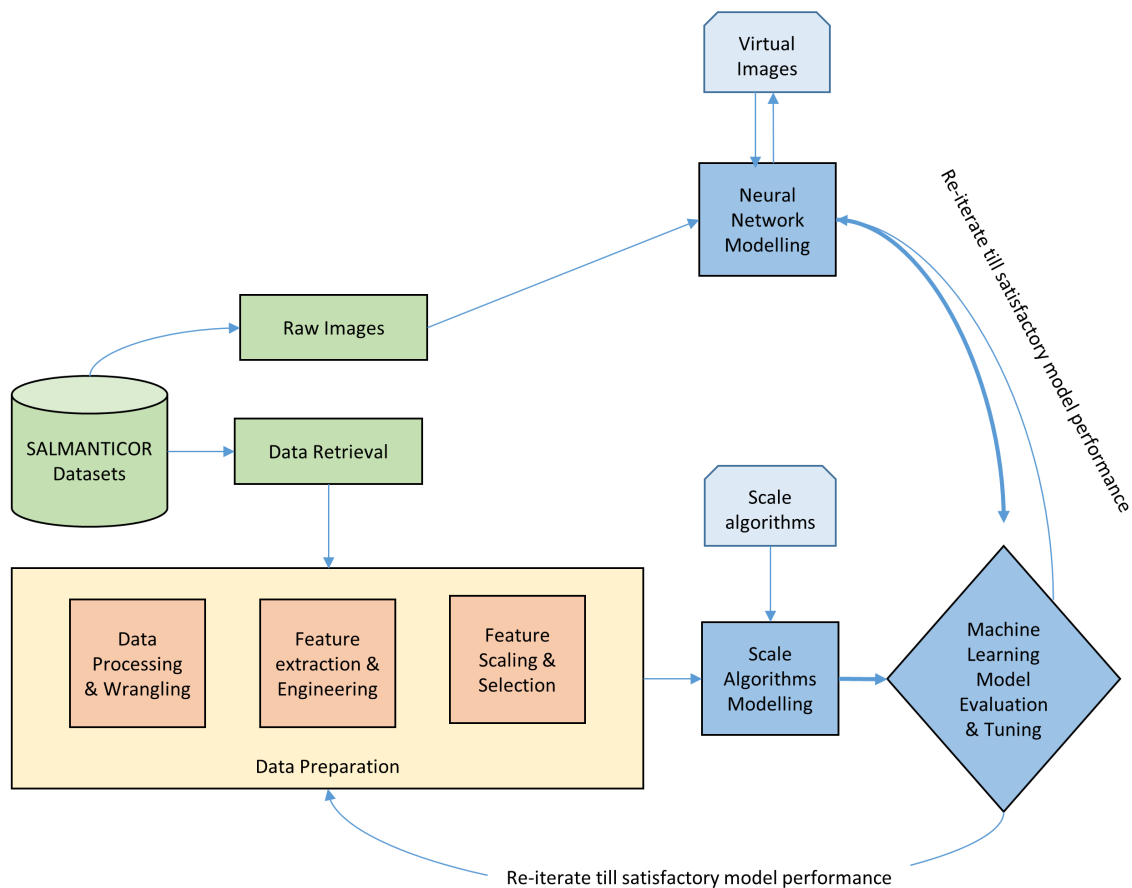


Figure 3 Machine learning (ML) pipeline for the SALMANTICOR study. The learning algorithm will take heterogeneous data that will be preprocessed to create input data for the ML algorithm. Furthermore, raw images will also be used in the ML algorithm using neural network modelling. The output of the ML algorithm will also need to be processed and improved until a satisfactory model is developed.

the study results, given their applicability and implications for the general population, will be disseminated in research meetings and in at least 10 articles published in scientific journals. Finally, population-based control groups are difficult to obtain, specifically in case–control cardiovascular studies where structural heart disease has to be rolled out. The SALMANTICOR study will provide availability of normative reference values quantification for echocardiographic, ECG, biochemical, genetics, VASERA and other parameters. Thus, international cooperation sharing data and participating in Horizon 2020 programmes with the SALMANTICOR population are contemplated.

Patient and public involvement

Patients' representatives will have an increasingly present voice in the SALMANTICOR study. There is currently an only patient organisation for heart disease in the province of Salamanca, 'El paciente experto'. This organisation has provided counselling in the design of the study, will jointly interpret the results of the study with the investigators of SALMANTICOR, will help to disseminate them to society and will be involved when establishing new policies for health improvement and education empowerment with the administration to halt the epidemic of cardiovascular disease.

A clinical report will be sent to all participants and their primary care medical doctors immediately after the clinical history is performed and the echocardiogram and ECG interpreted. Finally, the global and most important observations from the SALMANTICOR study will be sent by letter to all participants and to all doctors, primary care and specialists, of the province of Salamanca through the Medical College of Salamanca and our health Administration.

Data statement

Our data will be accessed at the Institute of Research of the University Hospital of Salamanca. Furthermore, our dataset will be published in a public repository. Additional code for our spatial analysis will be shared in a public Github repository.

DISCUSSION

A major strength of the SALMANTICOR study is the selection of a representative population-based cohort across primary care, with a probable significant number of structural heart disease cases of each age, sex and place of residence category to allow overall and subpopulation analyses. This population-based approach increases the

generalisability of the findings compared with surveys that addressed cardiovascular risk factors but have never included an echocardiographic assessment.^{11 14 41–44} Moreover, in view of the similarity of trends in cardiovascular disease and population ageing from Spain with other developed countries,⁴⁵ our findings are likely to be broadly applicable to them.

Echocardiography in the SALMANTICOR study is designed to address three specific aims. The first one is to characterise the abnormalities of cardiac structure and function in a community-based sample and to assess how these abnormalities vary by place of residence (rural or urban), age and sex. The study uses standard and novel echocardiographic techniques to characterise five specific domains of cardiac structure. These data will be used to define the population distribution of these measurements and to determine their relationship with the cardiovascular risk factors, including hypertension, diabetes mellitus, coronary disease, renal insufficiency and prognostically relevant biomarkers such as NT-proBNP and high-sensitivity troponin.

The second aim is to investigate ventricular–arterial coupling in addition to the association of cardiac structure and function with arterial stiffness assessed by CAVI, baPWV and ABI.

The third aim is to prospectively examine the extent to which these non-invasive measures associate with incidences of adverse cardiovascular outcomes and to determine the degree to which these associations also vary by age, sex and by place of residence (rural or urban). By accomplishing these objectives, this study is developing an echocardiographic imaging database that will facilitate future investigations to compare these echocardiographic measures both with studies previously performed in other countries,^{12 13} and to be used as a very well-established control group. Furthermore, our study will provide availability of normative reference values quantification for ECG, biochemical, genetics, VASERA and other parameters.

Adequate public health and service delivery planning requires reliable information about contemporary population-level disease incidence. Salud Castilla y León (SACYL) is the regional healthcare government authority of Castilla y León providing universal access to health services for 2.5 million people. SACYL is closely integrated with other public services and policies as part of a holistic approach to improving population health. In this sense, our study data will be used to understand the cardiovascular health needs of our community and to improve people's health and well-being, and how they can be developed. SALMANTICOR will be established as the global observatory on cardiovascular health research and development of SACYL, since we will include real-time data about the burden of cardiovascular disease, people's social circumstances and living conditions, lifestyles and diet, economic factors, access to healthcare and other services, as well as our genes, age and sex. In addition to understating the overall picture of our population's health, the data will be disaggregated

to identify inequalities for example by gender, sex and urban or rural place of residence. This will support the prioritisation of interventions depending on the needs of different groups and will require effective actions for the prediction and prevention of cardiovascular disease; from macropolicies down to individuals and families, empowering people to take control of their health. In this sense, two new medical technology research lines have been identified by the SALMANTICOR investigators: exploring the use of spatial methods and exploring modern computational methods developed in the field of ML.

The use of spatial methods in healthcare research enables disease distribution patterns to be identified and has become popular in the field of public health.^{46–48} Cancer and other disease mortality atlases have shown us that many risk factors of a territorial nature, influence geographical patterns, making it possible to select disease indicators and so reveal their geographical structure.^{49 50} However, the number of spatial analyses published in major epidemiology journals is still very low.⁵¹ One of the reasons is that the application of spatial methods requires specific training and has resulted in their substitution with less optimal methods from healthcare research. Therefore, it is important to promote spatial methods, especially those which are simple to interpret in the field of population-based studies and which could be potentially used in combination with other computational methods to facilitate interpretation, prediction and healthcare policies. Cardiology spatial analysis has been developed mainly in optimisation problems and prevalence prediction. As an example of optimisation, travel time isochrones analysis has been deployed in different facilities in order to identify exposed areas and act accordingly.⁵² Nevertheless, prevalence predictions are the most common geostatistical techniques in healthcare and it is not an exception in cardiology.^{53 54}

The incorporation of ML in medicine holds promise for substantially improved healthcare delivery.^{18–21} ML provides methods, techniques and tools that can help solving diagnostic and prognostic problems in a variety of cardiac medical domains.^{55–63} Furthermore, ML offers new approaches to leveraging the growing volume of heterogeneous data, including imaging data, available for analyses. To date, ML has been used in two broad and highly interconnected areas: automatization of tasks that might otherwise be performed by a human and generation of clinically important knowledge. However, it is argued that the successful implementation of ML methods can help the integration of computer-based systems in the healthcare environment providing opportunities to really improve the efficiency of medical care and to be used as a regional policy to establish effective public health programmes. In this sense, the SALMANTICOR study represents an excellent opportunity to explore ML algorithms for estimating and ranking the impact of environmental and classical risk factors in the development of structural heart disease in a population-based setting.

Author affiliations

¹Department of Cardiology, Hospital Universitario de Salamanca, Instituto de Investigación Biomédica de Salamanca (IBSAL), Facultad de Medicina, Universidad de Salamanca, and CIBERCV, Salamanca, Spain

²Miguel Armijo Primary Care Centre, Salamanca, Spain

³San Juan Primary Care Centre, Salamanca, Spain

⁴Robleda Primary Care Center, Salamanca, Spain

⁵Salamanca Primary Care Centre Management, Salamanca, Spain

⁶Miranda del Castañar Primary Care Centre, Salamanca, Spain

⁷Santa Marta Primary Care Centre, Salamanca, Spain

⁸Peñaranda de Bracamonte Primary Care Centre, Salamanca, Spain

Acknowledgements We thank all primary care physicians and personnel helping with the development of the study. We thank Philips Iberica and Obra Social 'La Caixa' for their support. We especially thank participants in the study and apologise for any inconvenience we could have caused. We thank the involvement of the Salamanca patient organisation 'El paciente experto', for providing counselling to Salmanticor and for further promoting the dissemination of the results to the society and to the regional government.

Contributors JIM-A: data acquisition, surveys completion, physical, electrocardiographic and VASERA examinations, design of the work, drafting the work and revising it critically, final approval of the version to be published. MC: data acquisition, surveys completion, conception and design of the work, drafting the work and revising it critically, final approval of the version to be published. AR: conception and design of the work, interpretation of data, drafting the work of revising it critically, primary care coordination, final approval of the version to be published. PPV: echocardiographic data acquisition, interpretation of data, final approval of the version to be published. MB-P: conception and design of the echocardiographic protocol, analysis and interpretation of echocardiographic data, drafting the work and revising it critically for important intellectual content, final approval of the version to be published. VV-P: conception and design of the spatial and machine learning analysis, analysis and interpretation of data, drafting the work and revising it critically for important intellectual content, final approval of the version to be published. FP-E: conception and design of the work, interpretation of data, primary care coordination, final approval of the version to be published. JH-H: conception and design of the electrocardiographic protocol, analysis and interpretation of ECG data, drafting the work and revising it critically for important intellectual content, final approval of the version to be published. BG: conception and design of the lifestyle, Mediterranean and exercise surveys, analysis and interpretation of data, final approval of the version to be published. SC: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. AM-G: analysis and interpretation of echocardiographic data, final approval of the version to be published. ED-P: analysis and interpretation of echocardiographic data, final approval of the version to be published. JMdD: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. AU: conception and design of the work (surveys), analysis and interpretation of data, final approval of the version to be published. JJ-C: conception and design of the work, analysis and interpretation of ECG data, final approval of the version to be published. IC-G: conception and design of the work (surveys), analysis and interpretation of data, final approval of the version to be published. BB: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. JMH: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. CS-P: data acquisition, surveys completion, physical, electrocardiographic and VASERA examinations, final approval of the version to be published. IS: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. MCL: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. PM: conception and design of the work, coordinator of 5 out of 35 primary care centres, acquisition of data, final approval of the version to be published. PID-D: conception and design of the spatial and machine learning analysis, analysis and interpretation of data, drafting the work and revising it critically for important intellectual content, final approval of the version to be published. PLS: conception and design of the study, interpretation of data, drafting the work, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This study was supported by national (PI14/00695, Institute of Health Carlos III, Spanish Ministry of Economy and Competitiveness) and community (GRS1030/A/14, SACYL, Junta Castilla y León) competitive grants and by the Spanish Cardiovascular Network (RIC and CIBERCV) from the Institute of Health Carlos III, Spanish Ministry of Economy and Competitiveness, Obra Social 'la Caixa' and Philips Iberica Healthcare division.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study has been approved by the clinical research ethics committee (CEIC) of the health area of Salamanca ('CEIC of Salamanca Health Area, 29 September 2014).

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. CDC, NCHS. Data are from the Multiple Cause of Death Files, 1999-2015, as compiled from data provided by the 56 vital statistics jurisdictions through the Vital Statistics Cooperative Program. 2017 (Accessed 6 Dec 2017).
2. European Cardiovascular Disease Statistics 2017. Data are from the European Heart Network (EHN), a Brussels-based Alliance of heart foundations and likeminded non-governmental organisations throughout Europe, with member organisations in 25 countries. 2017 www.ehnheart.org/cvd:statistics.html (Accessed 6 Dec 2017).
3. Mozaffarian D, Benjamin EJ, Go AS, *et al*. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* 2015;131:e29-322.
4. Pearson TA, Palaniappan LP, Artinian NT, *et al*. American Heart Association Guide for Improving Cardiovascular Health at the Community Level, 2013 update: a scientific statement for public health practitioners, healthcare providers, and health policy makers. *Circulation* 2013;127:1730-53.
5. Gerber Y, Weston SA, Enriquez-Sarano M, *et al*. Contemporary risk stratification after myocardial infarction in the community: performance of scores and incremental value of soluble suppression of tumorigenicity-2. *J Am Heart Assoc* 2017;6.
6. Dondo TB, Hall M, Timmis AD, *et al*. Geographic variation in the treatment of non-ST-segment myocardial infarction in the English National Health Service: a cohort study. *BMJ Open* 2016;6:e011600.
7. Zhang L, Desai NR, Li J, *et al*. National Quality Assessment of Early Clopidogrel Therapy in Chinese Patients With Acute Myocardial Infarction (AMI) in 2006 and 2011: insights from the China Patient-Centered Evaluative Assessment of Cardiac Events (PEACE)-Retrospective AMI Study. *J Am Heart Assoc* 2015;4.
8. Regueiro A, Bosch J, Martín-Yuste V, *et al*. Cost-effectiveness of a European ST-segment elevation myocardial infarction network: results from the Catalan Codi Infart network. *BMJ Open* 2015;5:e009148.
9. Conrad N, Judge A, Tran J, *et al*. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet* 2018;391.
10. Dawber TR, Meadors GF, MOORE FE. Epidemiological approaches to heart disease: the framingham study. *Am J Public Health Nations Health* 1951;41:279-86.
11. Teo K, Chow CK, Vaz M, *et al*. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J* 2009;158:1-7.
12. Shah AM, Cheng S, Skali H, *et al*. Rationale and design of a multicenter echocardiographic study to assess the relationship between cardiac structure and function and heart failure risk in a biracial cohort of community-dwelling elderly persons: the Atherosclerosis Risk in Communities study. *Circ Cardiovasc Imaging* 2014;7:173-81.
13. Vasan RS, Xanthakis V, Lyass A, *et al*. Epidemiology of left ventricular systolic dysfunction and heart failure in the framingham study: an echocardiographic study over 3 decades. *JACC Cardiovasc Imaging* 2018;11.

14. Yusuf S, Hawken S, Ounpuu S, *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937–52.
15. O'Donnell MJ, Xavier D, Liu L, *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010;376:112–23.
16. Chambers J, Kabir S, Cajeat E. Detection of heart disease by open access echocardiography: a retrospective analysis of general practice referrals. *Br J Gen Pract* 2014;64:e105–11.
17. Englund EJ. A variance of geostatisticians. *Math Geol* 1990;22:417–55.
18. Deo RC. Machine Learning in Medicine. *Circulation* 2015;132:1920–30.
19. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216–9.
20. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med* 2017;376:2507–9.
21. Shameer K, Johnson KW, Glicksberg BS, *et al.* Machine learning in cardiovascular medicine: are we there yet? *Heart* 2018;104:1156–64.
22. Lang RM, Badano LP, Mor-Avi V, *et al.* Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2015;16:233–71.
23. Marwick TH, Gillebert TC, Aurigemma G, *et al.* Recommendations on the use of echocardiography in adult hypertension: a report from the European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE). *Eur Heart J Cardiovasc Imaging* 2015;16:727–54.
24. Douglas PS, Garcia MJ, Haines DE, *et al.* ACCF/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. *J Am Coll Cardiol* 2011;57:1126–66.
25. Gomez-Marcos MA, Martinez-Salgado C, Gonzalez-Sarmiento R, *et al.* Association between different risk factors and vascular accelerated ageing (EVA study): study protocol for a cross-sectional, descriptive observational study. *BMJ Open* 2016;6:e011031.
26. Takaki A, Ogawa H, Wakeyama T, *et al.* Cardio-ankle vascular index is a new noninvasive parameter of arterial stiffness. *Circ J* 2007;71:1710–4.
27. Shirai K, Hiruta N, Song M, *et al.* Cardio-ankle vascular index (CAVI) as a novel indicator of arterial stiffness: theory, evidence and perspectives. *J Atheroscler Thromb* 2011;18:924–38.
28. Shirai K. Analysis of vascular function using the cardio-ankle vascular index (CAVI). *Hypertens Res* 2011;34:684–5.
29. Hu H, Cui H, Han W, *et al.* A cutoff point for arterial stiffness using the cardio-ankle vascular index based on carotid arteriosclerosis. *Hypertens Res* 2013;36:334–41.
30. Kawai T, Ohishi M, Onishi M, *et al.* Cut-off value of brachial-ankle pulse wave velocity to predict cardiovascular disease in hypertensive patients: a cohort study. *J Atheroscler Thromb* 2013;20:391–400.
31. Macfarlane PW, Katibi IA, Hamde ST, *et al.* Racial differences in the ECG--selected aspects. *J Electrocardiol* 2014;47:809–14.
32. Rijnbeek PR, van Herpen G, Bots ML, *et al.* Normal values of the electrocardiogram for ages 16–90 years. *J Electrocardiol* 2014;47:914–21.
33. Escofier B, Pagès J. Multiple factor analysis (AFMULT package). *Comput Stat Data Anal* 1994;18:121–40.
34. Guisado-Clavero M, Roso-Llorach A, López-Jimenez T, *et al.* Multimorbidity patterns in the elderly: a prospective cohort study with cluster analysis. *BMC Geriatr* 2018;18:16.
35. Benzecri JP, Dunod P. L'Analyse des Données. Volume II. L'Analyse des correspondances. Paris Dunod. 1973.
36. Wackermagel H. *Multivariate geostatistics: an introduction with applications*. New York, NY: Springer-Verlag, 2003.
37. Le S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* 1990;25:1–18.
38. Pebesma EJ. Multivariate geostatistics in S: the gstat package. *Comput Geosci* 2004;30:683–91.
39. Frizzell JD, Liang L, Schulte PJ, *et al.* Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol* 2017;2:204–9.
40. Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
41. Grau M, Elosua R, Cabrera de León A, *et al.* [Cardiovascular risk factors in Spain in the first decade of the 21st Century, a pooled analysis with individual data from 11 population-based studies: the DARIOS study]. *Rev Esp Cardiol* 2011;64:295–304.
42. Masiá R, Pena A, Marrugat J, *et al.* High prevalence of cardiovascular risk factors in Gerona, Spain, a province with low myocardial infarction incidence. REGICOR Investigators. *J Epidemiol Community Health* 1998;52:707–15.
43. Rigo Carratalá F, Frontera Juan G, Llobera Cànaves J, *et al.* [Prevalence of cardiovascular risk factors in the Balearic Islands (CORSAIB Study)]. *Rev Esp Cardiol* 2005;58:1411–9.
44. Félix-Redondo FJ, Fernández-Bergés D, Fernando Pérez J, *et al.* [Prevalence, awareness, treatment and control of cardiovascular risk factors in the Extremadura population (Spain). HERMEX study]. *Aten Primaria* 2011;43:426–34.
45. Roth GA, Johnson C, Abajobir A, *et al.* Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol* 2017;70:1–25.
46. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect* 2004;112:998–1006.
47. Abellan JJ, Richardson S, Best N. Use of space-time models to investigate the stability of patterns of disease. *Environ Health Perspect* 2008;116:1111–9.
48. Kontopantelis E, Stevens RJ, Helms PJ, *et al.* Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open* 2018;8:e020738.
49. Ho NT, Thompson C, Nhan LNT, *et al.* Retrospective analysis assessing the spatial and temporal distribution of paediatric acute respiratory tract infections in Ho Chi Minh City, Vietnam. *BMJ Open* 2018;8:e016349.
50. López-Abente G, Aragonés N, Pérez-Gómez B, *et al.* Time trends in municipal distribution patterns of cancer mortality in Spain. *BMC Cancer* 2014;14:535.
51. Auchincloss AH, Gebreab SY, Mair C, *et al.* A review of spatial methods in epidemiology, 2000–2010. *Annu Rev Public Health* 2012;33:107–22.
52. Forouzanfar MH, Alexander L, Anderson HR, *et al.* Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:2287–323.
53. Przybysz R, Bunch M. Exploring spatial patterns of sudden cardiac arrests in the city of Toronto using poisson kriging and hot spot analyses. *PLoS One* 2017;12:e0180721.
54. Ogunniyi MO, Holt JB, Croft JB, *et al.* Geographic variations in heart failure hospitalizations among medicare beneficiaries in the Tennessee catchment area. *Am J Med Sci* 2012;343:71–7.
55. Garcia EV, Cooke CD, Folks RD, *et al.* Diagnostic performance of an expert system for the interpretation of myocardial perfusion SPECT studies. *J Nucl Med* 2001;42:1185–91.
56. Paul AK, Shill PC, Rabin MRI, *et al.* Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. *Applied Intelligence* 2018;48:1739–56.
57. Raghavendra U, Fujita H, Gudigar A, *et al.* Automated technique for coronary artery disease characterization and classification using DD-DTDWT in ultrasound images. *Biomed Signal Process Control* 2018;40:324–34.
58. Alizadehsani R, Zangooei MH, Hosseini MJ, *et al.* Coronary artery disease detection using computational intelligence methods. *Knowl Based Syst* 2016;109:187–97.
59. Tan JH, Hagiwara Y, Pang W, *et al.* Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med* 2018;94:19–26.
60. Alizadehsani R, Hosseini MJ, Khosravi A, *et al.* Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. *Comput Methods Programs Biomed* 2018;162:119–27.
61. Arabasadi Z, Alizadehsani R, Roshanzamir M, *et al.* Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput Methods Programs Biomed* 2017;141:19–26.
62. Acharya UR, Fujita H, Lih OS, *et al.* Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowl Based Syst* 2017;132:62–71.
63. Acharya UR, Fujita H, Adam M, *et al.* Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: a comparative study. *Inf Sci* 2017;377:17–29.