The Cantabria Cohort, a protocol for a population-based cohort in northern Spain

Marta Alonso-Peña[1*], Trinidad Dierssen[1,2], Maria José Marin[1], Jessica Alonso-Molero[1,2], Inés Gómez-Acebo[1,2], Inés Santiuste[1], Jeffrey V. Lazarus[3,4], Pascual Sanchez-Juan[5,6], Galo Peralta[1], Javier Crespo[1,2,7], Marcos Lopez-Hoyos[1,2,7] and Cantabria Cohort Collaborators

Abstract

Cantabria Cohort stems from a research and action initiative lead by researchers from Valdecilla Research Institute (IDIVAL), Marqués de Valdecilla University Hospital and University of Cantabria, supported by the regional Goverment. Its aim is to identify and follow up a cohort that would provide information to improve the understanding of the etiology and prognosis of different acute and chronic diseases. The Cantabria Cohort will recruit between 40,000–50,000 residents aged 40–69 years at baseline, representing 10–20% of the target population. Currently, more than 30,000 volunteers have been enrolled. All participants will be invited for a re-assessment every three years, while the overall duration is planned for twenty years. The repeated collection of biomaterials combined with broad information from participant questionnaires, medical examinations, actual health system records and other secondary public data sources is a major strength of its design, which will make it possible to address biological pathways of disease development, identify new factors involved in health and disease, design new strategies for disease prevention, and advance precision medicine. It is conceived to allow access to a large number of researchers worldwide to boost collaboration and medical research.

Keywords Population-based cohort, Lifestyle, Socio-economic factors, Biobank, Big data, Precision medicine, Longitudinal study, Spain

*Correspondence:
Marta Alonso-Peña
malonso@idival.org
[1] Valdecilla Research Institute (IDIVAL), Santander 39011, Spain
[2] Faculty of Medicine, University of Cantabria, Santander 39011, Spain
[3] Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain
[4] CUNY Graduate School of Public Health and Health Policy (CUNY SPH), New York, NY, USA
[5] CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, 28220 Madrid, Spain
[6] Alzheimer's Centre Reina Sofia-CIEN Foundation-ISCIII, 28031 Madrid, Spain
[7] Marques de Valdecilla University Hospital, Santander 39008, Spain

Background

Recent reports have shown that non-communicable diseases, especially cardiovascular and metabolic diseases, were the main cause of morbi-mortality in Spain in 2019 [1]. Therefore, risks factors related to lifestyle, behavior and environment are placing a heavy burden on the Spanish population's health [1]. At the same time, incidence of these diseases and an aging population threaten the sustainability of the Spanish universal health system. That stated, preventive, predictive and personalized medicine are urgently needed; to achieve such an aim, extensive data and knowledge will be fundamental.

Long-term follow-up cohort studies have played a crucial role in understanding modifiable factors associated with the development of chronic disease. In the early 1950s, initial results from the Framingham Heart Study

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 2 of 13

were published [2, 3], beginning a new era in epidemiology [4]. Large prospective cohort studies have established themselves as the most appropriate epidemiological design for research in the field of multimorbidity in real-life conditions [5]. Over the last decades, the number of large prospective cohorts have increased [6–10].

The Framingham Heart study selected its study population from a well-defined geographic area, the city of Framingham (Massachusetts, USA). This same strategy was carried out in Europe. One of the most remarkable examples was that of the European Prospective Investigation into Cancer and Nutrition (EPIC), which was an initiative of the International Agency for Research on Cancer (IARC) [6]. More recently, national cohort studies have been developed [7–10].

In Spain, several efforts have been undertaken to establish prospective, population-based cohorts [11–19]. However, to date, long-term, population-based and multipurpose studies covering a specific territory have not been attemped. Conversely, a national cohort study has just been launched [20]. Health system management in Spain takes place at the regional (autonomous community) level, which in practice results in 17 autonomous health systems within Spain [1].

In this context, the Cantabria Cohort was launched in late 2020 and it includes residents in Cantabria, an autonomous community located in northern Spain. Cantabria covers an area of 5,330 km^2, hosting a total population of 584,507 inhabitants, 45% living in the region's capital [21] (Fig. 1). Moreover, the Cantabrian Public Health System is organized into 42 health areas for primary care. It has four hospitals; the Hospital Universitario Marqués de Valdecilla (HUMV) is the tertiary and referenced hospital in the region. This center helps implement large-scale studies, reducing the necessity of coordination among different institutions and procedures. Further, due to the widespread COVID-19 vaccination campaign in 2021, a huge effort to integrate multiple public databases and update contact information of the population has facilitated the start of the project. Cantabria Cohort stems from a research and action initiative to improve the regional health system and advance the health-related Sustainable Development Goals [22]. The present paper describes the design and implementation of the Cantabria Cohort and the processes related to the biological sample collection and data acquisition. It also provides an overview of the governing board, quality assurance, and legal and ethical aspects, as well as future research opportunities and cooperation.

Study objectives

This project's main objective is to identify and follow up a cohort that would provide baseline information on lifestyles, socio-economic aspects, and morbidity of
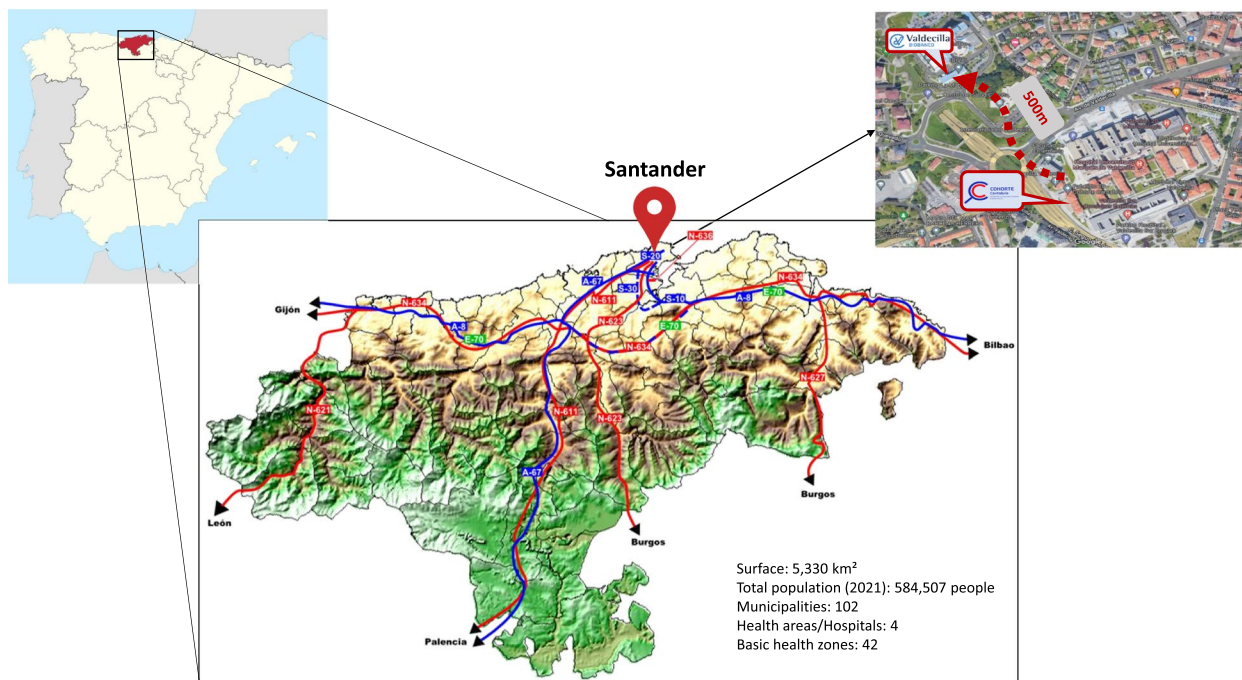

Fig. 1 Location of study centers and main transport connections in the region [21]. Santander is the region's capital. *Map source: Wikimedia Commons*

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 3 of 13

the Cantabrian population. This information would be related to health events registered during the follow-up to improve the understanding of the etiology and prognosis of different acute and chronic diseases. The repeated collection of biomaterials combined with broad information from participant questionnaires, medical examinations, actual health system records and other public data sources is a major strength of its design. The study is planned and built from a multidisciplinary research and action perspective. It is conceived to allow access to a large number of researchers worldwide to boost collaboration and medical research. Overall, five general objectives are pursued by the Cantabria Cohort. Within their general scope, more specific research questions and projects will be included.

Integration and use of real-world data to improve health research

One of the main objectives is the development of systems to use official and public secondary data sources for scientific research. This implies the integration, validation and analysis of dispersed data, as well as the legal and administrative governance necessary to assure the safety and ethical use of the sources. In this regard, real-world evidence is an emerging domain in health research and several European projects have been launched to create collaborative networks in which clinical data can be shared and integrated to define common data models, standardize vocabulary and delineate government structures, such as the European Health Data & Evidence Network (EHDEN) [23]. Cantabria Health Service and Valdecilla Research Institute (IDIVAL) joined EHDEN. Our project has taken advantage of the first steps taken to create a dataset using the Observational Medical Outcomes Partnership (OMOP) Common Data Model [24] and boost its transformation into a longitudinal database for implementation in research. Official data sources other than health records will also be included in the project.

Identifying lifestyle risk factors and their involvement in chronic diseases

The first set of aims is to increase the understanding of the role of lifestyle factors (e.g., smoking, alcohol consumption, physical activity, dietary patterns, body composition, occupation and environmental conditions) in developing major forms of chronic disease, with a special emphasis on obesity, metabolic syndrome and its related conditions. Of note, 60.4% of the Cantabria population is considered to have a sedentary lifestyle [25] and incidence of such conditions is increasing worldwide [26]. By integrating different sources of data, as described below, we aim to gain insights into the natural history and causal

physiopathological pathways. In the Cantabria region, up to 27% of the population older than 15 years old has hypercholesterolemia, 22.6% suffers from hypertension, 19% low-back pain, 15.2% mental illness, 7.9% has diabetes, 5.6% arthrosis and 1.7% suffers from chronic obstructive pulmonary disease [27]. We also aim to document the impact of these risk factors in terms of epidemiology and public health strategies.

Evaluation of geographic and socio-economic disparities in health and healthcare

Another objective of the Cantabria Cohort is to increase awareness about the causes of social and regional disparities in health. According to the National Health System Annual Report 2020–2021, the positive perception of health in the population aged 15 and over is clearly lower among people with a basic level of education and below, especially among women [25]. Besides, the AROPE (At risk of poverty or social exclusion) rate in Cantabria is 19.4% [25]. Socio-economic position and psychosocial factors will be evaluated as health determinants via calculation of HOUSES index [28] from primary questionnaires and Spanish Cadastre and ecological secondary (Atlas of Urban Vulnerability) socio-economic sources of data [29, 30]. Finally, information will be collected on the use of health services, medical interventions and medications.

Assessing biomarkers for the early detection of diseases and disease risks

Determining biomarkers has become one of the main objectives of biomedical research. Technological advances, especially in "omic" sciences, have revolutionized research on disease biomarkers and even the identification of health markers. However, applying such "omics" requires large collections of biological samples to be consistent and avoid statistic bias [31]. Therefore, one of the main objectives of Cantabria Cohort is to collaborate with other international partners in the investigation and development of omics studies to identify new health and disease biomarkers, increase biological replicates and introduce diversity into the genetic and environmental backgrounds of the populations being analyzed.

Surveillance of viral hepatitis and HIV in Cantabria

The World Health Organization (WHO) set the ambitious goals of eliminating viral hepatitis B and C as a public health threat by 2030 and reducing the number of people newly infected with HIV [26]. To achieve these goals, the scale-up of direct-acting antiviral therapies is a necessary cornerstone. It cannot, however, be implemented without micro-elimination programmes that facilitate identification of undiagnosed cases. Once

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 4 of 13

diagnosed, viral hepatitis and HIV can be treated, efficiently reducing the risk of transmission [32, 33]. Furthermore, HCV, HBV and HIV are spread by the same mechanisms, so it is common that individuals infected with one of these viruses are also infected by the others [34]. According to the last reports of the Spanish Directorate General of Public Health, in collaboration with the Cantabrian Autonomous System of Epidemiological Surveillance, the notification of new VIH cases has been reducing from 53 total cases in 2009 to 14 in 2020 [35], while the incidence of hepatitis B is the highest in the country (2.23 cases per 100,000 inhabitants in 2020) [36] and that of hepatitis C is also among the highest in Spain (2.75 cases per 100,000 inhabitants in 2020) [37]. Thus, the Cantabria Cohort also aims to reduce HCV, HBV and HIV incidence in Cantabria and enhance its treatment by serological testing in all participants.

Study design and methods

For the development and execution of these study protocol, we have follow the recommendations of SPIRIT 2013 guideline [38].

Study population and recruitment

The Cantabria Cohort will recruit 40,000–50,000 residents aged 40–69 years at baseline. An information campaign in various general media was launched in 2021.

Thereafter, study participants were recruited through any of the following: 1) voluntary registration on the study website (www.cohortecantabria.com) or direct telephone contact and 2) random selection (stratified by sex and age) using the Cantabrian Public Health System population database. Afterwards, participants are contacted by telephone; they are informed about all the phases of the study, their rights of withdrawal, the possible disadvantages of participating in the study, and other fundamental characteristics. People who agree to participate are sent the Informed Consent together with the full Participant Information Sheet and an appointment at the study center, HUMV (Fig. 2). Recruitment and calls are carried out by trained personnel and will continue until the sample size is reached or follow-up is initiated, which is planned for April 2024.

No financial compensation for participation is planned, but participants receive, if requested, a report of their test results (i.e., blood test results and anthropometric data). Further, the medical team reviews all discordant analytical results and health records to decide whether any clinical finding should be communicated to the study participants or need further medical attention. Recruitment started on 20 April 2021 and it is still ongoing. The expected response rate in the global randomly selected population was around 50%, which was confirmed during recruitment (response rate in the whole randomly
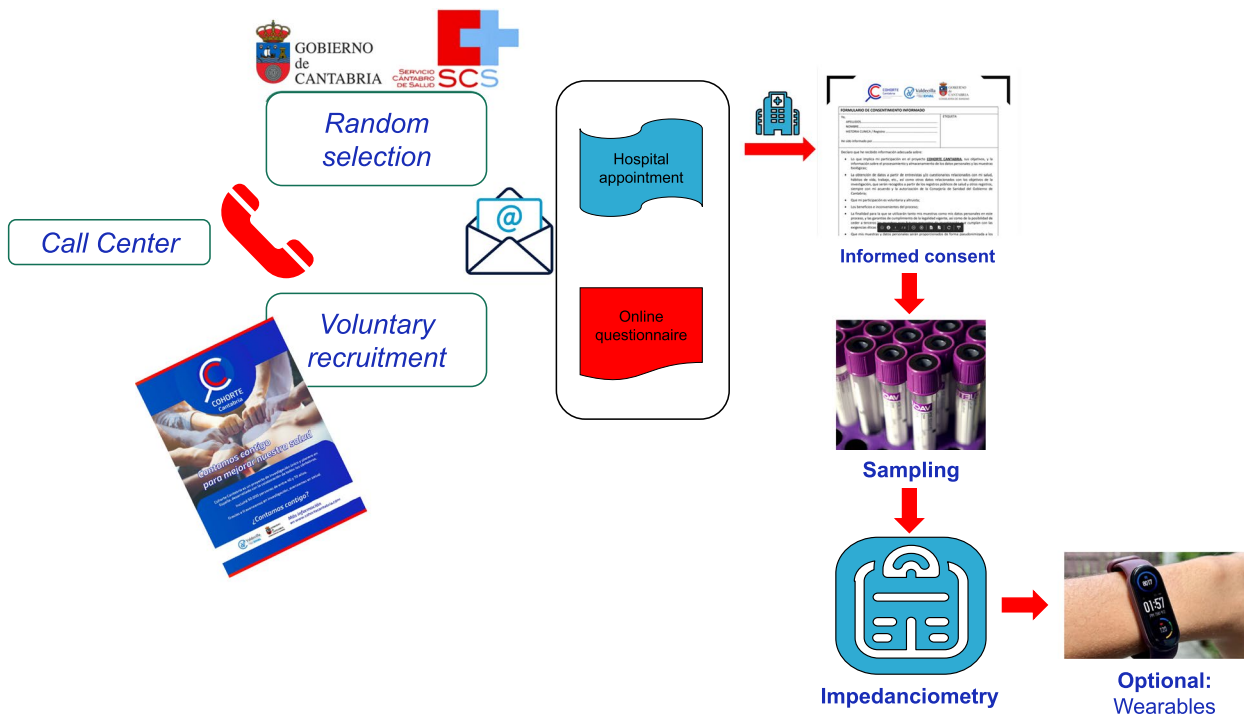

Fig. 2 Selection process, patient recruitment and collection of information for the study

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 5 of 13

selected population is currently 43.6%). At the moment of publication, more than 30,000 individuals have already been included in the study, of whom 57.8% were randomly selected. Due to potential bias in recruitment, type of recruitment (voluntary/random) has been recorded for each individual in the study. Once recruitment is closed, deep analysis will be undertaken and published to compare the demographics of the voluntary group and the random group against each other (See preliminary data on Supplementary Table 2), and against the Cantabrian population. Moreover, in order to evaluate the reproducibility of our findings, we will perform sensitivity analysis that excludes voluntarily registered participants. Data to be released for future research will always include the variable "type of participation" so that researchers can include it in multivariate models to control for its potential confounding effect.

Baseline examination and data collection

Data collection starts with the telephone call wherein current contact information is updated, if necessary. Once identification, contact information and appointment have been confirmed, a self-administrated questionnaire is delivered as a link to REDCap platform via email [39]. It can also be given on paper in a format ready for digital reading data extraction in the case of non-Internet users. Digital questionnaires must be completed before the appointment while paper questionnaires must be returned by post or in person.

Medical examinations last 10–15 min and include blood extraction for biobanking and basic analytics and measurements of anthropometric characteristics [40]. Finally, a random subset of participants is invited to carry a wearable activity wristband.

All participants will be re-invited for follow-up examinations (re-assessment) every three years after their baseline recruitment, including approximately the same examinations.

Laboratory analysis and biological samples

Biological material from all patients will be treated in accordance with Law 14/2007 on Biomedical Research [41]. Blood samples will be used to measure conventional analytical parameters as well as serological markers of viral infection (Table 1, Supplementary Table 1). If the participant expressly authorizes the donation of samples to the Valdecilla Biobank, two more tubes of blood are withdrawn. In addition, authorization is requested for integration into the Valdecilla Biobank of the surplus tissue samples from therapeutic or diagnostic surgical procedures available at Cantabrian hospitals' pathology departments. The samples kept in the custody

of the Biobank will be governed by the provisions of Royal Decree 1716/2011, of November 18 [42].

Study questionnaire

This questionnaire includes the level of education attained, gross income, quality of life, diet, physical activity, family history, work activity and other habits, anthropometric data, tobacco and alcohol consumption, housing characteristics, etc., measured by externally validated surveys (Table 1, Supplementary Table 1). The questionnaire is structured in six independent modules to ease its completion and requires approximately one hour.

Body composition

Bioelectric impedance analysis (BIA) has emerged as a validated method for evaluating body composition [40] and detecting malnutrition [46, 47] (Table 1, Supplementary Table 1). Seca mBCA 515/514 and 274 digital stadiometers (Hans E. Rüth, Barcelona, Spain) are used in this context to obtain an immediate overview of the distribution of muscles, fat and water in the body by 8-point bioelectrical impedance analysis of 19 different frequencies and seven body segments.

Wearable activity wristband

Participants may be offered an electronic bracelet (Xiaomi MI Smart Band 5) that will allow us to obtain information on: 1) physical activity; 2) sleep; and 3) heart rate. Specifically, a measurement of steps and the type of physical activity is performed every minute while heart rate, every 10 min for 21 days. Afterward, the device is connected to the researchers' smartphones as previously described [48]. Until now, around 10% of study participants have accepted to wear the electronic bracelet and provide information on daily physical activity.

Data from public health records

Cantabria Health Service and IDIVAL joined EHDEN in 2020 [49]. Currently, the IDIVAL database in EHDEN is a cross-sectional dataset representing citizens who receive public health assistance from the Cantabrian Health Service between 2016 and 2020. The information provided by the region's primary care is related to annotations, diagnoses (converted from ICPC2 to SNOMED), clinical variables and vaccines. From the hospital setting, appointments, tests (SNOMED), diagnoses (converted to SNOMED from IDC10), variables and specific information on Hospital Pharmacy have been included. Laboratory results (LOINC) and electronic prescription information from both settings are included. Data transformation and validation were performed as previoculy described [50]. The Cantabria Cohort will take advantage

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 6 of 13

Table 1 Data obtained at baseline examination from laboratory, impedanciometry and questionnaires

Source	Category	Data/Instrument
Laboratory	Biochemistry	Serum glucose, urea, creatinine, sodium, potassium, uric acid, ALT, AST, GGT, alkaline phosphatase, triglycerides, HDL, LDL and total cholesterol, albumin, calcium, C-reactive protein
	Hormones	TSH, T4
	Hepatology	APRI, FIB4
	Haematology	Leukocytes, neutrophils, lymphocytes, monocytes, eosinophils, basophils, erythroblasts, red blood cells, hemoglobin, hematocrit, mean corpuscular hemoglobin, erythrocyte distribution width, mean corpuscular volume, mean corpuscular hemoglobin concentration, platelets, mean platelet volume
	Microbiology	Hepatitis B, Hepatitis C, HIV
Impedanciometry	Anthropometric measurements	Height, weight, body mass index, waist circumference
	Body composition	Fat mass, visceral adipose tissue, lean mass, skeletal muscle mass, total body water, extracellular water
	Energy	Energy expenditure, energy expenditure at rest, stored energy
	Raw impedanciometry data	Bioimpedance vector, impedance, reactance, resistance, phase angles
Questionnaire	Socioeconomic	Marital status, level of education, employment status, annual income, housing characteristics, place of residence
	Demographic	Age, sex, ethnicity, phototype
	Quality of life	Spanish version of the SF-36 Health Survey [43]
	Diet	13-item Mediterranean diet adherence screener and 9-item scale of adherence to a low-fat diet [44]
	Physical activity	Spanish version of the physical activity questionnaire [45]
	Sleep habits	The items employed to get this information come from GCAT Genomes for life project [15]
	Tobacco use/exposure	GCAT Genomes for life project [15]
	Alcohol consumption	GCAT Genomes for life project [15]
	Family history of disease	Progenitors, siblings and descendants. These items were got from GCAT Genomes for life [15]
	Memory	Cognitive Complaints Questionnaire [40]

ALT Alanine transaminase, *AST* Aspartate transaminase, *GGT* Gamma-glutamyltransferase, *HDL* High-density lipoprotein, *LDL* Low-density lipoprotein, *TSH* Thyroid-stimulating hormone, *T4* Thyroxine, *APRI* AST-to-Platelet Ratio Index, *FIB4* Fibrosis index based on four factors, *HIV* Human immunodeficiency virus

of the inclusion of its target population in IDIVAL dataset. Thus, all information regarding public health assistance is directly linked to Cantabria Cohort Database. Furthermore, procedures have been developed to use national healthcare databases to allow for the identification and validation of diseases throughout follow-up.

Non-users of public health assistance from the Cantabrian Health Service will be identified and asked to provide updates on their health records every three years. Until now, only 2.4% of recruited participants are non-users.

Data on main diagnoses and comorbidities obtained from health databases will be validated annually from the 10% of the randomly chosen sample. The validation will be performed on a selection of the most prevalent diagnoses in the cohort. The reference criterion for validation will be a review of clinical history by two trained physicians to evaluate the sample independently and in masked form. Internal validity parameters will be calculated: sensitivity and specificity with their 95% CI and external validity (predictive values). In addition,

inter-rater concordance will be performed using the kappa index.

Other data sources
In addition to data from public health services, there is a multitude of documentary sources from different public and governmental administrations with information that can help to understand the living conditions of the participants and their relevance in health and disease. For this reason, Cantabria Cohort works with administrations to facilitate the secondary use of data for research. Among them, the following draw attention:

– National Death Index (IND) from Spanish Ministry of Health: The deaths in the cohort will be updated annually based on the IND.
– Spanish Cadastre from the Spanish Ministry of Finance and Public Administration: based on the cadastral reference, information about real state and housing will be imported (square meters of the state, cadastral value, location, etc.).

Alonso-Peña *et al. BMC Public Health* (2023) 23:2429

Page 7 of 13

- The Urban Vulnerability Atlas developed by the Ministry of Transport, Mobility and Urban Agenda: sociodemographic, socio-economic, housing and subjective perception vulnerabilities.
- Social Security System from the Spanish Ministry of Inclusion, Social Security and Migration: it will allow the incorporation of participants' labor information.
- Data collected by the Spanish National Institute of Statistics.

Time line and statistical power considerations

The projected timeframe for the Cantabria Cohort covers 20 years (Fig. 3) starting 2021. For the calculation of statistical power, we have considered two follow-up time points: at 5 and 20 years, after which we expect to have retained 75% (35,000) and 10% (5,000) of the initial sample, respectively (Fig. 4). Estimating the sample power to detect weak associations (hazard ratios less than 1.2) allows us to evaluate the study's capacity to identify associations in the worst-case scenario, which needs the largest sample size. Furthermore, hazard ratio enables the control of the confounding effect of differential follow-up times, therefore it has been proposed employing Hazard Ratio estimation to calculate the necessary sample size to achieve optimal statistical power at the conclusion of the study [51, 52]. Thus, with 35,000 individuals, the

minimum hazard ratio values that can be detected with an 80% power are 1.1, 1.12, 1.14 and 1.2 for effect ratio values of 10%, 7.5%, 5% and 2.5%, respectively; whereas, with 5,000 individuals, the minimum hazard ratio values that can be detected with an 80% power are 1.29, 1.34, 1.43 and 1.65 for effect ratio values of 10%, 7.5%, 5% and 2.5%, respectively.

Study organization

Central data management

Data will be collected trough web-based, REDCAP standardized data entry forms, while others sources of data (analytical parameters, impedance, webereables, paper-based questionnaires) will be transformed from their original database into csv files to be easily handled in different statistical software packages. These sources of data will be validated and integrated into a SQL database to be readily available upon request of data collection services, ensuring continuous back-ups of the full datasets. Statistical analysis will depend on the project and specific research aims. However, as a general rule, the data will be presented differently based on the type of variable. For qualitative variables, absolute frequencies and percentages will be used. For quantitative variables, measures of central tendency like mean and median, as well as measures of dispersion like standard deviation, interquartile range, minimum and maximum values will
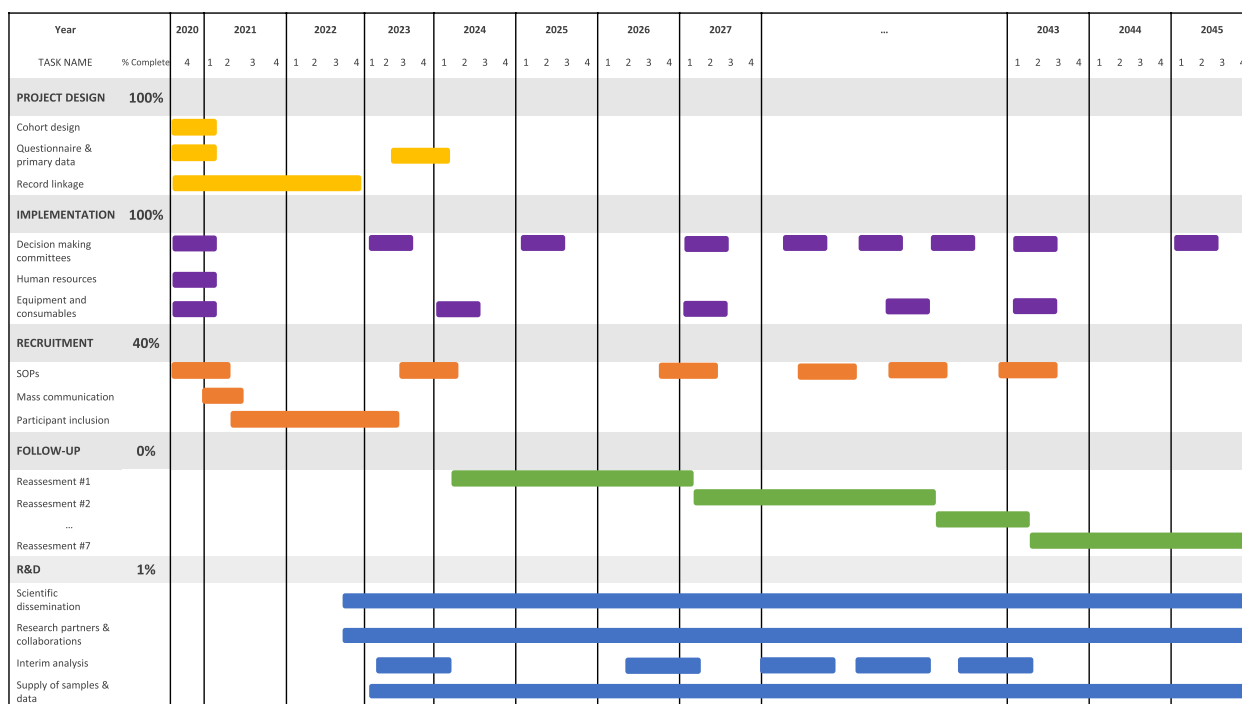

Fig. 3 Cantabria Cohort project execution and future development – timeline. SOPs: standard operating procedure. R&D: Research and development
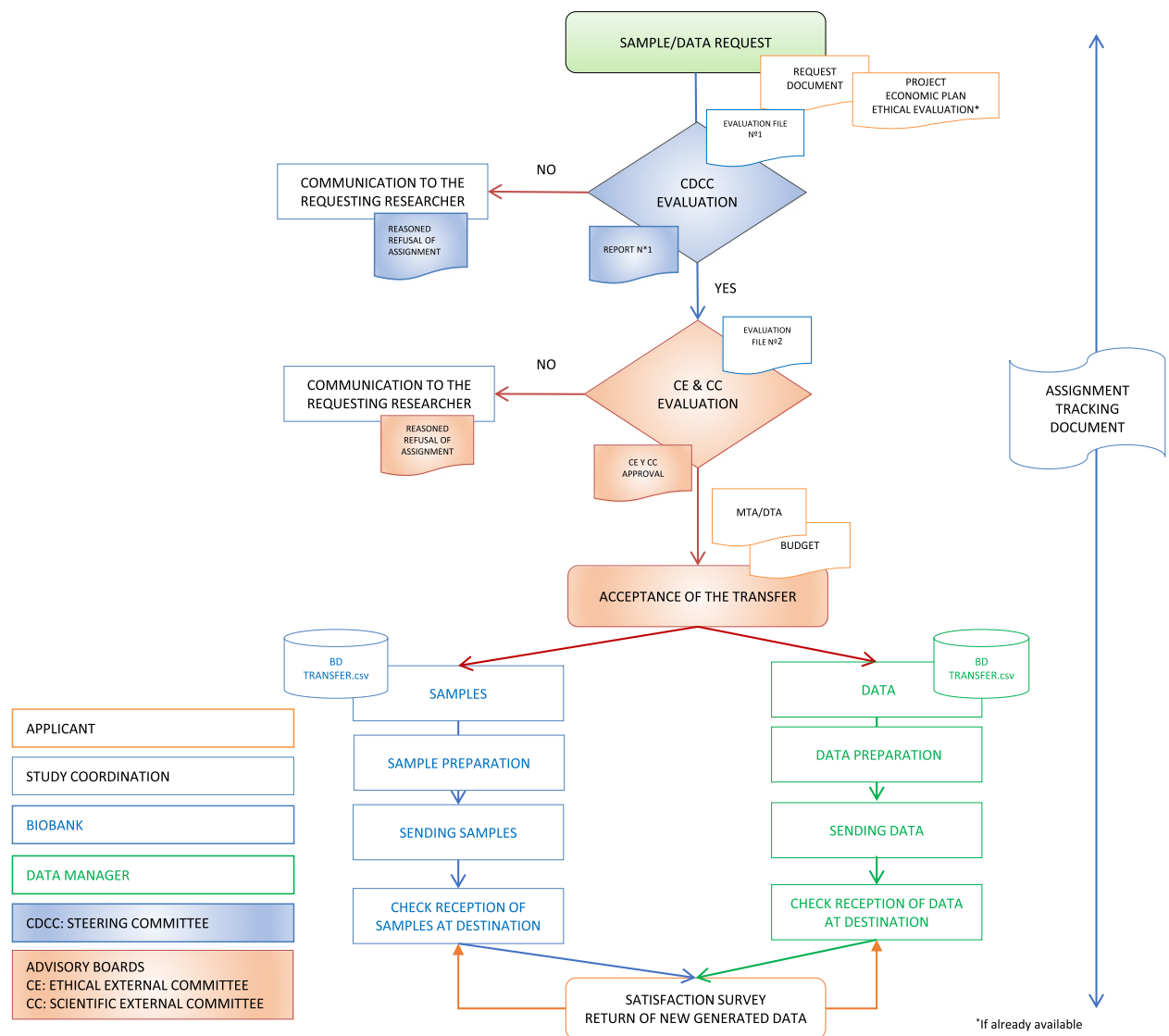
Alonso-Peña *et al. BMC Public Health* (2023) 23:2429

Page 8 of 13

Fig. 4 Statistical power per hazard ratio at 5 (**A**) and 20 (**B**) years of follow-up

be used to provide indicators of the distribution's shape, such as asymmetry indices and kurtosis. In the case of ordinal variables, the description used will depend on the number of categories. Regardless of the variable type, a column will be added to the data presentation table to indicate the number of patients with available data.

All data will be kept in secure servers from the regional health service. Prospective record linkage will be performed by the Digital Transformation team from the Regional Ministry of Health of the Government of Cantabria, a trusted third party separate from the main study database. The trusted third party maintains all linkages to external secondary data sources.

Collection and storage of biological samples

Samples collected for future research are registered, processed and stored at the Valdecilla Biobank, an active member of the Spanish Platform ISCIII Biobanks and Biomodels. For this purpose, two tubes of blood are drawn from all participants who agree to donate a sample to the biobank:

- 10 ml Serum separator clot activator tube
- 10 ml K2EDTA tube.

Samples are processed according to the SOPs of the Spanish Biobanks Network to obtain serum, plasma and buffy coat (BC) or DNA. Handling and storage of

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 9 of 13

the samples is performed in the same location where obtained, ensuring the quality and stability of the biological samples by following the ISO 9001:2015 standard. Plasma and serum samples are aliquoted into six 2D cryotubes each and stored at -80ºC within two hours of collection. BC samples remain at -40 °C until DNA extraction. A magnetic bead nucleic acid extraction system (Chemagic 360, PerkinElmer Inc) is used to isolate DNA from the BC. A biorepository management software is used to efficiently and securely manage biological samples, guaranteeing the complete traceability of sample information and all their associated data (Noray-banks, Noray Bioinformatics, S.L.U.)

Ethics and data confidentiality

All participants in the study receive written information about the project and the Informed Consent document at home and are also informed verbally about it (by telephone call to make an appointment). If they agree to participate, they sign the Informed Consent form on the day of the appointment at the study center where nurses and other healthcare personnel can resolve any other questions or doubts. The Informed Consent has been designed for collection and use of participant data and biological specimens in ancillary studies.

The study protocol (ID 2021.057) was approved by the Ethics Committee for Drug Research of Cantabria (CEIm) on 26 February 2021. The study respects the ethical principles of research with biological samples, the
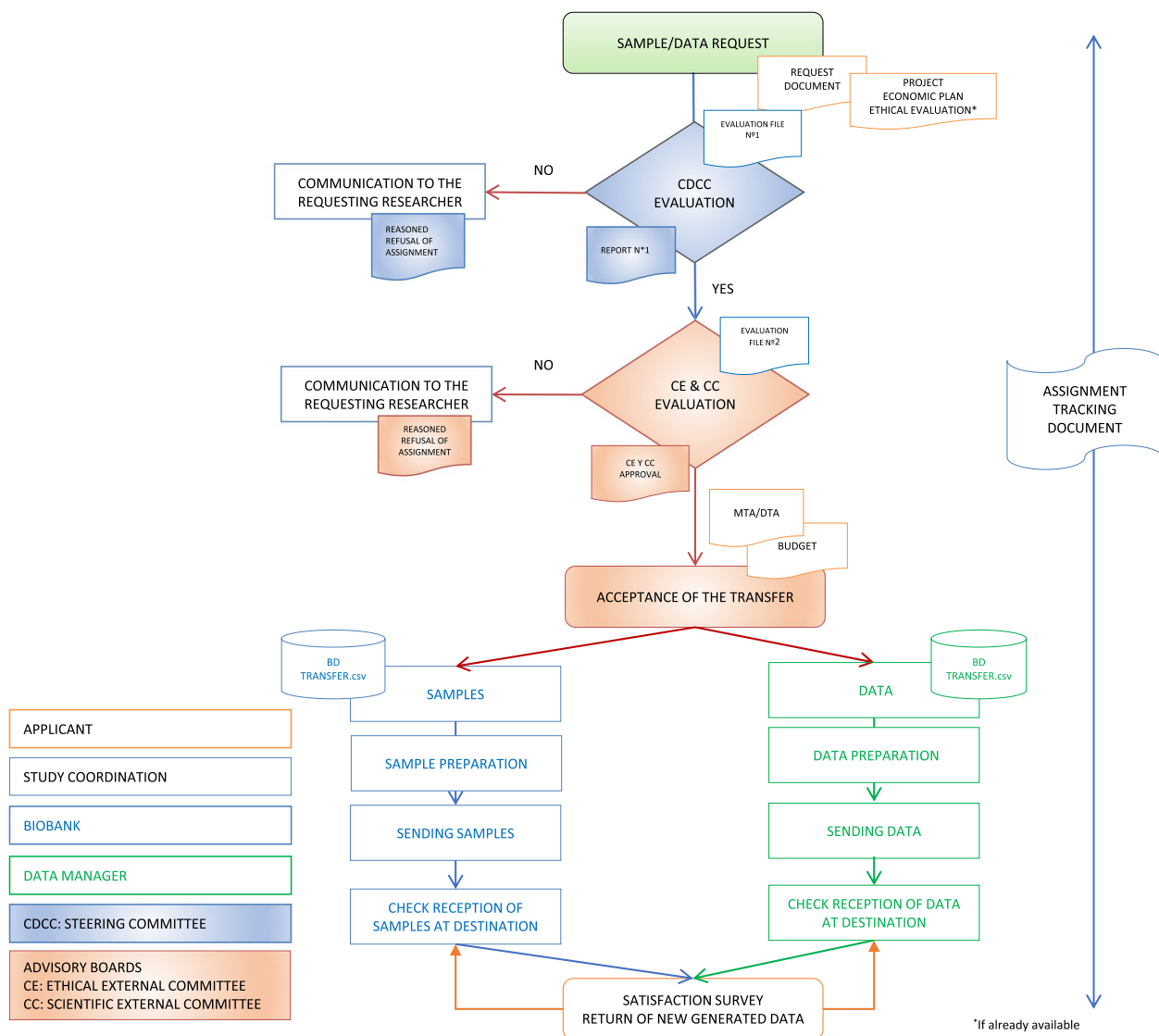
Fig. 5 Flowchart for Cantabria Cohort's data and sample transfer to research groups and institutions
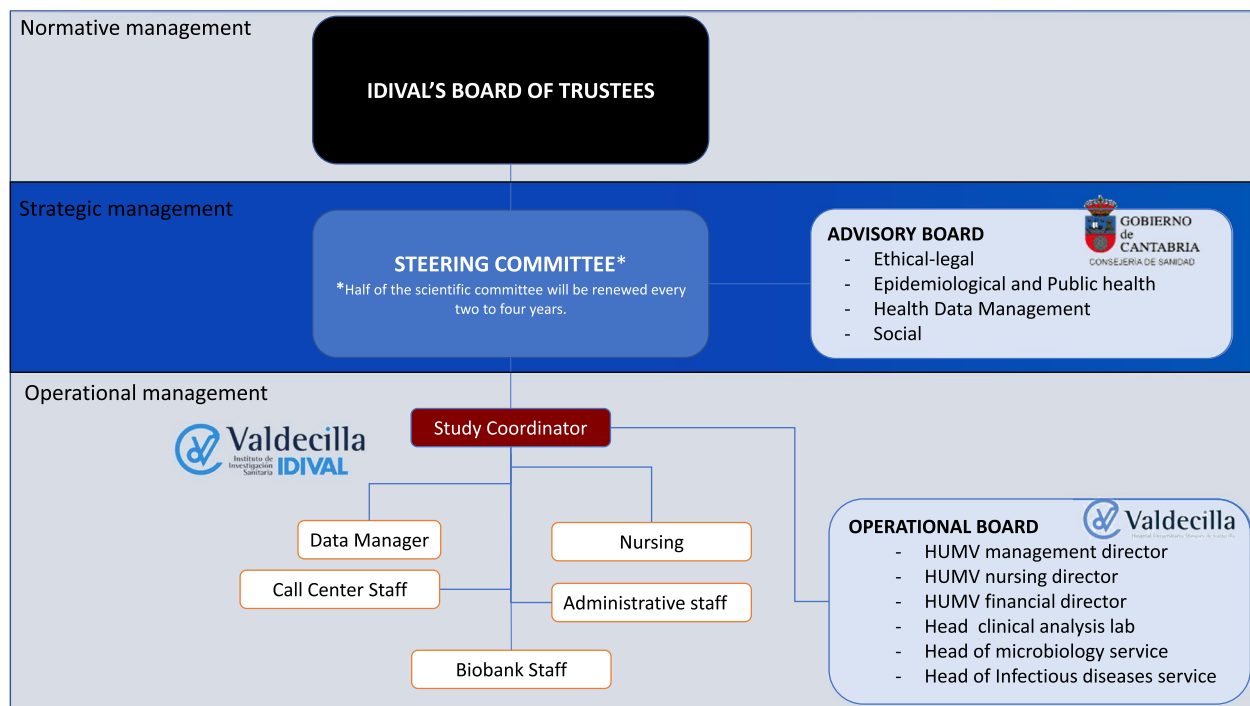
Fig. 6 Scheme of governance and management structures of the Cantabria Cohort

1975 Declaration of Helsinki and Spanish Law 14/2007, of July 3, 2007, on Biomedical Research. Specifically, restrictions to sample and data used given by participants through informed consent will always be revised and respected before any sample or data transfer. Furthermore, the study have been registered in ClinicalTrials.gov as "The Cantabria Cohort, a Protocol for a Population-based Cohort in Northern Spain" (NCT05852678), being the protocol version 3.0 on December 2, 2022.

The processing, communication and transfer of personal data of all participants will comply with the provisions of the applicable regulations: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on Data Protection (RGPD) and the Organic Law 3/2018, of 5 December, on the Protection of Personal Data and guarantee of digital rights [53]. Thus, samples and data accessible to the public will always be transmitted in a pseudo-anonymized form and specific SOPs and security mechanisms will be implemented to avoid re-identification. Specifically, to guarantee the privacy and confidentiality of the information obtained, two sets of data were generated, stored in encrypted form:

1) A main database, only accessible by the study's data manager, with the identification data of the partici-

pants together with their identification codes in the different sources of data and related tables.

2) For every data transfer, an anonymized database in which a numerical code is assigned to each participant will be generated. This dataset gathers all the information collected in this study and requested by researchers.

In compliance of the Organic Law 3/2018 on the Protection of Personal Data and guarantee of digital rights and Spanish Law 14/2007 on Biomedical Research, any important protocol modifications and information regarding third parties accesing to the data will be communicated broadly through the project webpage, www.cohortecantabria.com, and participants will be specifically informed on a quarterly basis through a newsletter.

Sample and data transfer

The Cantabria Cohort has a firm commitment to participants and researchers to establish, develop and promote health research and the advancement of medicine and strategies to improve the well-being of the Cantabrian society. Samples and data from the Cantabria Cohort may be used for biomedical research projects by public

biomedical research centers, universities, or private for-profit and not-for-profit institutions. Due to the limited nature of the samples and the strategic nature of the associated information, which is their main added value, sample and data transfer will be controlled by the Steering Committee (Fig. 5) and will be based on the following principles:

– The Cantabria Cohort will publish on its website its policy and procedures for access to samples and data for use in research.
– The Cantabria Cohort will maintain full control over access to and use of project data and samples in accordance with its commitment to public use.
– The Cantabria Cohort Steering Committee will ensure compliance with the collection access policy to ensure that the resource is used efficiently and for public benefit.

Once the Cantabria Cohort Steering Committee authorizes the transfer of samples and/or data, the request will be sent for evaluation to the Ethical External Committee and the External Scientific Committee (Fig. 5).

Regarding authorship eligibility, the number of authors will depend on the specific requirements of each journal, with the maximum number of authors allowed as a limit. In order to facilitate maximum impact for the greatest number of authors participating in each of the ancilliary studies of Cantabria Cohort, all the papers derived directly from the project will include a first, third and last authors proposed by the principal investigator of the study and a senior co-author proposed by the Steering Committee. Furthermore, the publication must express the acknowledgement to the Cantabria Cohort and IDIVAL.

Funding, governance and management

This project has the institutional support of all the health research organizations in our community. It is led by IDIVAL, with all participating researchers being the heads of the research unit of their respective groups (Javier Crespo, Trinidad Dierssen, Marcos López Hoyos and Pascual Sánchez). In addition, Marcos López Hoyos is the scientific director of IDIVAL; Galo Peralta, the managing director of IDIVAL and María José Marín, the Scientific Director of the Valdecilla Biobank. The Cantabria Cohort has public endorsements from the Regional Minister of Health of the Government of Cantabria, the manager of the Cantabrian Health Service, the rector of the University of Cantabria, the director of IBBTEC and the managing director of HUMV (Fig. 6). This study protocol was not independently peer reviewed as part of the funding

process. The initial budget, which covered the first 8 months of recruitment, amounted to an investment of 1.5 million euros. It is estimated that annual costs will increase to 1.6–1.8 million euros.

Conclusion

The Cantabria Cohort has been designed as a multipurpose prospective cohort and research tool to investigate any acute and chronic illnesses, especially those associated with lifestyle, nutrition, exercise and aging. This project sets extraordinary groundwork for scientific cooperation and networking among epidemiologists and other health scientists, improving our regional health system and advancing the national and global achievements of health-related United Nations' Sustainable Development Goals.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-023-17318-8.

Additional file 1.

Additional file 2.

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 12 of 13

Availability of data and materials

The datasets used and/or analysed during the current study available on reasonable request following standard procedures as described in the manuscript, "Sample and Data Transfer" section. Contact information and rules for accession will be updated on www.cohortecantabria.com. For inquiries related to this article, please contact corresponding author.

Declarations

Ethics approval and consent to participate

This study protocol (ID 2021.057) was approved by the Ethics Committee for Drug Research of Cantabria on 26 February 2021. The study respects the ethical principles of research with biological samples, the 1975 Declaration of Helsinki and Spanish Law 14/2007, of July 3, 2007, on Biomedical Research. Informed consent will be obtained from all subjects included in the study and/or their legal guardian(s). It can be consulted, together with the participant's information sheet, at https://cohortecantabria.com/documentos/.

Consent for publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Nonetheless, they stated that: M.A.P. received a fellowship from Hoffmann-La Roche; J.V.L. reports grants from AbbVie, Gilead Sciences, MSD, and Roche Diagnostics; consulting fees from Novo-Vax; payment or honoraria for lectures, presentations, speakers' bureaus, and educational events from AbbVie, Gilead, Sciences, Intercept, and Janssen; M.L.H. reports grants from Werfen and ThermoFisher; consulting fees from Werfen and ThermoFisher; payment or honoraria for lectures, presentations, speakers' bureaus, and educational events from Werfen, ThermoFisher, Sanofi, GSK, Astra Zeneca, UCB Pharma, Astellas, and Takeda. The rest of the authors have no competing interest to declare.

References

1. Lazarus JV, Ortiz A, Tyrovolas S, Fernández E, Guy D, White TM, et al. A GBD 2019 study of health and Sustainable Development Goal gains and forecasts to 2030 in Spain. Sci Rep. 2022;12(1):21154.
2. Dawber TR, Moore FE, Mann GV. Coronary heart disease in the Framingham study. Am J Public Health Nations Health. 1957;47(4 Pt 2):4–24.
3. Oppenheimer GM. Becoming the Framingham Study 1947–1950. Am J Public Health. 2005;95(4):602–10.
4. Susser M, Susser E. Choosing a future for epidemiology: I. Eras and paradigms. Am J Public Health. 1996;86(5):668–73.
5. Le Reste JY, Nabbe P, Lingner H, Kasuba Lazic D, Assenova R, Munoz M, et al. What research agenda could be generated from the European General Practice Research Network concept of Multimorbidity in Family Practice? BMC Fam Pract. 2015;16:125.
6. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113–24.
7. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.
8. Henny J, Nadif R, Got SL, Lemonnier S, Ozguler A, Ruiz F, et al. The CONSTANCES Cohort Biobank: An Open Tool for Research in Epidemiology and Prevention of Diseases. Front Public Health. 2020;8:605133.
9. Langseth H, Gislefoss RE, Martinsen JI, Dillner J, Ursin G. Cohort Profile: The Janus Serum Bank Cohort in Norway. Int J Epidemiol. 2017;46(2):403–4.
10. Consortium GNCG. The German National Cohort: aims, study design and organization. Eur J Epidemiol. 2014;29(5):371–82.
11. Oreiro-Villar N, Raga AC, Rego-Pérez I, Pértega S, Silva-Diaz M, Freire M, et al. PROCOAC (PROspective COhort of A Coruña) description: Spanish prospective cohort to study osteoarthritis. Reumatol Clin (Engl Ed). 2020;18(2):100–4.
12. Pérez-González A, Araújo-Ameijeiras A, Fernández-Villar A, Crespo M, Poveda E, Institute CC-otGSHR. Long COVID in hospitalized and non-hospitalized patients in a large cohort in Northwest Spain, a prospective cohort study. Sci Rep. 2022;12(1):3369.
13. Rodríguez-Artalejo F, Graciani A, Guallar-Castillón P, León-Muñoz LM, Zuluaga MC, López-García E, et al. Rationale and methods of the study on nutrition and cardiovascular risk in Spain (ENRICA). Rev Esp Cardiol. 2011;64(10):876–82.
14. Martinez-Revuelta D, Irure-Ventura J, López-Hoyos M, Olmos JM, Pariente E, Martín-Millán M, et al. Comparison of ANA testing by indirect immunofluorescence or solid-phase assays in a low pre-test probability population for systemic autoimmune disease: the Camargo Cohort. Clin Chem Lab Med. 2023;61(6):1095–104.
15. Obón-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galván-Femenía I, et al. GCAT|Genomes for life: a prospective cohort study of the genomes of Catalonia. BMJ Open. 2018;8(3):e018324.
16. Crespo J, Cuadrado A, Perelló C, Cabezas J, Llerena S, Llorca J, et al. Epidemiology of hepatitis C virus infection in a country with universal access to direct-acting antiviral agents: data for designing a cost-effective elimination policy in Spain. J Viral Hepat. 2020;27(4):360–70.
17. Rodríguez-Antigüedad Zarranz A, Mendibe Bilbao M, Llarena González C, Audicana C. Mortality and cause of death in multiple sclerosis: findings from a prospective population-based cohort in Bizkaia, Basque Country. Spain Neuroepidemiology. 2014;42(4):219–25.
18. Naves Díaz M, Díaz López JB, Gómez Alonso C, Altadill Arregui A, Rodríguez Rebollar A, Cannata Andía JB. Estudio de incidencia de fracturas osteoporóticas en una cohorte mayor de 50 años durante un período de 6 años de seguimiento. Med Clin. 2000;115(17):650–3.
19. Soriguer F, Rubio-Martín E, Fernández D, Valdés S, García-Escobar E, Martín-Núñez GM, et al. Testosterone, SHBG and risk of type 2 diabetes in the second evaluation of the Pizarra cohort study. Eur J Clin Invest. 2012;42(1):79–85.
20. National Institute of Health Carlos III IMPaCT 2021 [Available from: https://www.isciii.es/QueHacemos/Financiacion/IMPaCT/Paginas/default.aspx.
21. Instituto Nacional de Estadística. INEbase 2021 [Available from: https://www.ine.es/dynInfo/Infografia/Territoriales/galeriaCapitulo.html?capitulo=4334.
22. United Nations. Sustainable Development Goals [Available from: https://sdgs.un.org/.
23. European Health Data & Evidence Network. EHDEN Portal 2022 [Available from: https://www.ehden.eu/.
24. Observational Health Data Sciences and Informatics collaborative. The book of OHDSI 2021. Available from: https://ohdsi.github.io/TheBookOfOhdsi/.
25. Ministerio de Sanidad. Informe anual del Sistema Nacional de Salud 2020–2021. Informes, Estudios E Investigación; 2022. Available from: https://www.sanidad.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm.
26. World Health Organization. Regional Office for Europe. WHO European Regional Obesity Report 2022; 2022. Available from: https://www.who.int/europe/publications/i/item/9789289057738.
27. Spanish Ministry of Health. Principales problemas crónicos de salud, porcentaje de población de 15 y más años que padece determinados problemas crónicos, registrado en atención primaria, por comunidad autónoma. 2021 [Available from: https://www.sanidad.gob.es/estadEstudios/sanidadDatos/tablas/tabla5.htm.
28. Juhn YJ, Beebe TJ, Finnie DM, Sloan J, Wheeler PH, Yawn B, et al. Development and initial testing of a new socioeconomic status measure based on housing data. J Urban Health. 2011;88(5):933–44.

Alonso-Peña *et al. BMC Public Health*   (2023) 23:2429

Page 13 of 13

29. Ministry of Finance and the Civil Service, Secretary of State for Finance, General Directorate of Cadastre. Electronic Cadastre Office [Available from: https://www.sedecatastro.gob.es/.

30. Spanish Ministry of Public Works. Atlas of urban vulnerability in Spain: methodology and contents 2012 [Available from: https://www.yumpu.com/en/document/view/28042109/atlas-of-urban-vulnerability-in-spain-methodology-and-contents.

31. Lay JO, Liyanage R, Borgmann S, Wilkins CL. Problems with the "omics." Trends Analytical Chemistry. 2006;25(11):1046–56.

32. Cohen MS, Gay CL. Treatment to prevent transmission of HIV-1. Clin Infect Dis. 2010;50 Suppl 3(0 3):S85-95.

33. Fraser H, Martin NK, Brummer-Korvenkontio H, Carrieri P, Dalgard O, Dillon J, et al. Model projections on the impact of HCV treatment in the prevention of HCV transmission among people who inject drugs in Europe. J Hepatol. 2018;68(3):402–11.

34. Park JS, Saraf N, Dieterich DT. HBV plus HCV, HCV plus HIV, HBV plus HIV. Curr Gastroenterol Rep. 2006;8(1):67–74.

35. Unidad de vigilancia de VIH ITS y hepatitis. Vigilancia Epidemiológica del VIH y sida en España 2021: Sistema de Información sobre Nuevos Diagnósticos de VIH y Registro Nacional de Casos de Sida. Centro Nacional de Epidemiología. Instituto de Salud Carlos III/División de control de VIH, ITS, Hepatitis virales y tuberculosis. Ministerio de Sanidad; 2022. Available from: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/VIH/informes%20de%20vigilancia%20VIH%20y%20sida%20anteriores/Informe%20VIH_SIDA_2022_CCAA.pdf.

36. Centro Nacional de Epidemiología. Vigilancia epidemiológica de la hepatitis B en España, 2020. Instituto de Salud Carlos III; 2022. Available from: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/archivos%20A-Z/Hepatitis%20B/Vigilancia_HepatitisB_2020.pdf.

37. Centro Nacional de Epidemiología. Vigilancia epidemiológica de la hepatitis C en España, 2020. Instituto de Salud Carlos III; 2022. Available from: https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/archivos%20A-Z/Hepatitis%20C/Vigilancia_HepatitisC_2020.pdf.

38. Chan AW, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586.

39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81.

40. Bosy-Westphal A, Schautz B, Later W, Kehayias JJ, Gallagher D, Müller MJ. What makes a BIA equation unique? Validity of eight-electrode multifrequency BIA to estimate body composition in a healthy adult population. Eur J Clin Nutr. 2013;67(Suppl 1):S14-21.

41. Estado Jd. Ley 14/2007, de 3 de julio, de Investigación biomédica. 2007 [Available from: https://www.boe.es/eli/es/l/2007/07/03/14.

42. Ministerio de Ciencia e Innovación. Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica. 2011 [Available from: https://www.boe.es/eli/es/rd/2011/11/18/1716.

43. Alonso J, Prieto L, Antó JM. The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results. Med Clin (Barc). 1995;104(20):771–6.

44. Martínez-González MA, Fernández-Jarne E, Serrano-Martínez M, Wright M, Gomez-Gracia E. Development of a short dietary intake questionnaire for the quantitative estimation of adherence to a cardioprotective Mediterranean diet. Eur J Clin Nutr. 2004;58(11):1550–2.

45. Martínez-González MA, López-Fontana C, Varo JJ, Sánchez-Villegas A, Martinez JA. Validation of the Spanish version of the physical activity questionnaire used in the Nurses' Health Study and the Health Professionals' Follow-up Study. Public Health Nutr. 2005;8(7):920–7.

46. Karavetian M, Salhab N, Rizk R, Poulia KA. Malnutrition-Inflammation Score VS Phase Angle in the Era of GLIM Criteria: A Cross-Sectional Study among Hemodialysis Patients in UAE. Nutrients. 2019;11(11):2771.

47. Nesvold MB, Jensen JL, Hove LH, Singh PB, Young A, Palm Ø, et al. Dietary Intake, Body Composition, and Oral Health Parameters among Female Patients with Primary Sjögren's Syndrome. Nutrients. 2018;10(7):866.

48. López-García S, Lage C, Pozueta A, García-Martínez M, Kazimierczak M, Fernández-Rodríguez A, et al. Sleep Time Estimated by an Actigraphy Watch Correlates With CSF Tau in Cognitively Unimpaired Elders: The Modulatory Role of APOE. Front Aging Neurosci. 2021;13:663446.

49. European Health Data & Evidence Network. EHDEN Portal. Data Partner Catalogue 2022 [Available from: https://www.ehden.eu/datapartners/.

50. Papez V, Moinat M, Voss EA, Bazakou S, Van Winzum A, Peviani A, et al. Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. J Am Med Inform Assoc. 2022;30(1):103–11.

51. Cantor AB. Sample size calculations for the log rank test: a Gompertz model approach. J Clin Epidemiol. 1992;45(10):1131–6.

52. Van Houwelingen H. Modelling Survival Data in Medical Research. D. Collett, Chapman & Hall, London, 1994. No of pages: XVII + 347. Price: E19.99. ISBN 0–41 2–44890–4. Statistics in Medicine. 1995;14(9):1147–8.

53. European Union. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). 2016 [Available from: https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.