

Cantabria Cohort — A biomedical research tool open to the world

The contribution of large population-based study cohorts to the knowledge of disease risk factors and causes is unquestionable. The Framingham Cohort, a paradigm for the population-based studies that changed the history of medicine, emerged after the antibiotic revolution, the decrease in infection-related mortality, and the boom in cardiovascular morbidity and mortality.

It began in 1948 with 5,209 volunteers aged from 28 to 62 years of age, residents of the City of Framingham, Massachusetts, and the project is still ongoing with the third-generation study involving the grandchildren of the original cohort participants (1). Over the past few decades the number of large prospective cohorts has notably increased, including the European Prospective Investigation into Cancer and Nutrition cohort, UK Biobank, CONSTANCES Cohort Biobank, Janus Serum Bank Cohort, German National Cohort (2-6), and China Kadoorie Biobank (7). In Cantabria there are previous experiences of developing much more modest population-based cohort studies such as Cohorte Camargo, initiated in 2006 to study bone metabolism and cardiovascular disease in postmenopausal women and men older than 50 years seen at a Cantabrian health center (8,9); Cohorte ETHON, made up of subjects between 20 and 79 years of age, selected from among the general population attending health clinics at three Spanish university hospitals (Madrid, Santander, Valencia), which allowed to assess the prevalence of HCV and metabolic-associated fatty liver disease (10,11); and Cohorte Valdecilla, started in 2018 with volunteers aged over 55 years, healthy or with cognitive decline, aimed at studying risk predictors for dementia development (12). It was in late 2020 that a unique study, a first in Spain, was initiated in Cantabria with a multipurpose population-based cohort as an effort towards precision medicine: the Cantabria Cohort (13).

Cantabria Cohort was conceived as a prospective, multipurpose, population-based cohort aiming to collect as much information as possible on the lifestyles and the demographic, educational, socioeconomic and health aspects of a significant portion of the Cantabrian population aged 40 to 70 years, relating these variables to the health events recorded during follow-up, which is expected to last 20 years. It is a cohort conceived as a tool to drive collaborative research at the local, national and international levels, integrating, validating and analyzing data from various sources, including: a) identification of lifestyle-associated risk factors and their implication in chronic diseases, even monitoring physical activity; b) assessment of geographic and socioeconomic inequalities regarding health and health care; c) impact of noncare-related health determinants through the construction of an individual socioeconomic status index derived from housing characteristics (*HOUSES index*) (14), and the Spanish Cadaster primary questionnaires and Atlas of Urban Vulnerability (15); d) use of health services, medical interventions, and drugs and other healthcare products; e) identification and assessment of biomarkers for early disease/risk detection; and finally, f) monitoring of viral hepatitis and human immunodeficiency virus (HIV) infection. Furthermore, collection of biological samples from all participants will allow different omic studies for the advancement of disease knowledge, and identification of novel biomarkers paving the way towards precision medicine. In order to analyze and optimize the huge data set being collected, we shall use the best artificial intelligence tools available, which will make possible the development of various clinical decision-making support systems and advanced predictive models, among others, in order to ensure progress in precision medicine at both the individual and the population levels (16) (Fig. 1).

Cantabria Cohort is now a formidable fact, having enrolled over 33,000 citizens in our region within two years (17). These are individuals for whom we collect sociodemographic, educational, physical activity, quality of life, dietary habits, and work history data, as well as, obviously, any information relating to their health, whether at the primary care or the hospital level, and full prescription history. Probably this cohort already possesses the most complete dataset of its kind in our country, a dataset that will grow with multiple additional data (including genetic data) and any events developing in the next few decades. This dataset is already available to accredited researchers for the development of studies to improve our understanding of the prevention, diagnosis and treatment of a wide array of health conditions. Volunteers' generosity allows the setting of highly ambitious goals by sharing information, facilitating national and international collaboration, and speeding up innovation. This cohort is not only intended to plot our community's genetic map; it also aspires to radically transform today's medicine and our health. Advances will be decisively made towards future medicine, that is, precision medicine. A medicine that provides patients with tailored (*personalized*) care; that is capable of developing new models to anticipate the risk for certain diseases (*predictive*); that may suggest a given lifestyle to avert illness and a way to detect

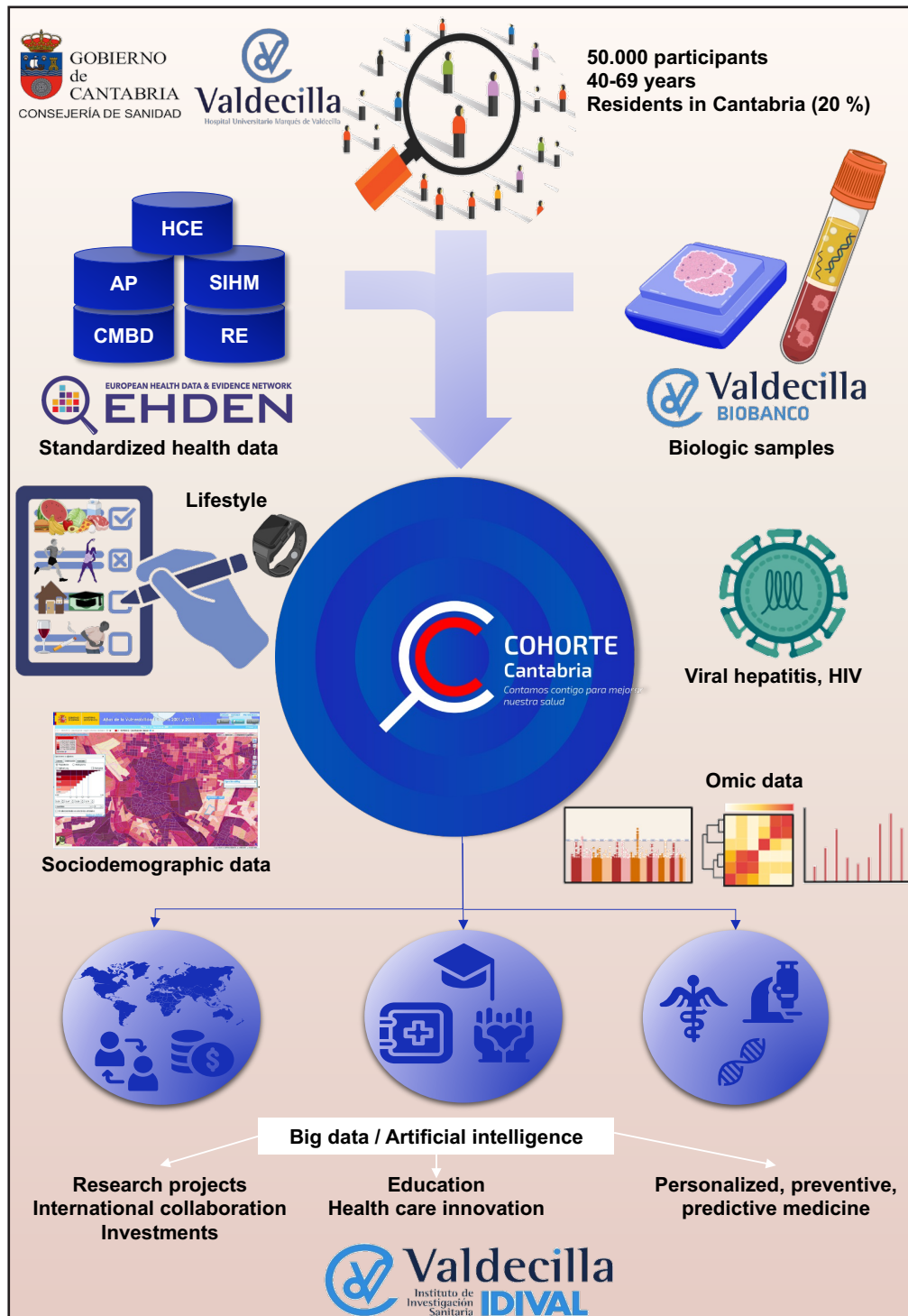


Fig. 1. Value generation process of the Cantabria Cohort Project. Inputs: participants (20 % of target population), clinical, sociodemographic, lifestyle, etc., data, and biological samples. Outputs: research projects, international collaboration, investments, innovation, education, personalized, predictive, preventive medicine, new insights, treatments, diagnostic biomarkers, etc. (HCE: electronic medical records; AP: primary care; SIHM: multicenter hospital information system; CMBD: minimum basic data set; RE: electronic prescription; HIV: human immunodeficiency virus).

illness early (*preventive*); and finally, that takes into account and values a patient's view and experience (*participative*). We expect to meet our baseline goal of recruiting 50,000 volunteers during the first semester of 2025. This would be a terrific landmark that, added to participant commitment, will allow — even as we are taking our first steps — setting ambitious goals for the short, mid, and long terms.

Cantabria Cohort is intended to reach beyond individual health, is intended to become a *paradigm of precision public health*. Its goal is to apply the appropriate intervention to the appropriate population at the appropriate time aiming to improve the health of the whole population, incorporating the social and environmental determinants of health. And this precision public health differs from the conventional approach in the use of big data, advanced analysis, artificial intelligence and new technologies to achieve more precise intervention goals. The integration and use of information from

different settings, and its ultimate translation into applied interventions, constitutes a formidable challenge for public medicine, a challenge for our Cantabria Cohort, a challenge that will allow improvements in risk prediction and health outcomes. Collaboration with both public and private major research institutes in international research consortia will allow new data to be generated. And in the immediate future, as a niche for both national and international investigator training, it will prevent the figure of the, tentatively called, “clinician-scientist”, namely that health professional who spends part of his/her time caring for patients (or populations) and devotes a considerable time to research, from disappearing (18). An open science collaboration in the personalized medicine setting that will allow participation in various European multipurpose cohort initiatives, naturally including IMPaCT project (19). And, as a result of this collaboration, we hope to speed up innovation and make ourselves a place at the leading edge of health science both at present and in the future. Innovation by developing technological health solutions to be implemented within the health system itself, contributing to the ecological digital transition. Since the health sector is one of the economic drivers of society, we expect —why not?— that Cantabria Cohort, by invigorating biomedical research and generating both knowledge and innovation, will draw investment to the cohort’s environment in addition to improving health in our setting.

To conclude, Cantabria Cohort was conceived as a prospective multipurpose cohort, a research instrument to gain further insight into both acute and chronic diseases. It is a project that lays extraordinary foundations for cooperation and networking between health scientists, thus improving our regional health system and the promotion of health-related national and international achievements. A project capable of contributing to the improvement of health due to the quantity, quality and management of the information made available to investigators worldwide.

A Cantabrian project open to the world.

Authors’ contribution: Writing-original draft: J. C.; writing-review and editing: M. A.-P.

Conflict of interest: the authors declare no conflict of interest.

Artificial intelligence: the authors declare that they did not use artificial intelligence (AI) or any AI-assisted technologies in the elaboration of the article.

Marta Alonso-Peña, Javier Crespo
Gastroenterology and Hepatology Department. Clinical and Translational Research in Digestive Diseases. Valdecilla Research Institute (IDIVAL). Hospital Universitario Marqués de Valdecilla. Santander, Spain

DOI: 10.17235/reed.2024.10400/2024

REFERENCES

1. Oppenheimer GM. Becoming the Framingham Study 1947-1950. *Am J Public Health*. 2005;95(4):602-10. DOI: 10.2105/AJPH.2003.026419
2. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113-24. DOI: 10.1079/PHN2002394
3. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203-9. DOI: 10.1038/s41586-018-0579-z
4. Henny J, Nadif R, Got SL, et al. The CONSTANCES Cohort Biobank: An Open Tool for Research in Epidemiology and Prevention of Diseases. *Front Public Health* 2020;8:605133. DOI: 10.3389/fpubh.2020.605133
5. Langseth H, Gislefoss RE, Martinsen JI. Cohort Profile: The Janus Serum Bank Cohort in Norway. *Int J Epidemiol* 2017;46:403-4. DOI: 10.1093/ije/dyw027
6. Consortium GNCG. The German National Cohort: aims, study design and organization. *Eur J Epidemiol* 2014;29:371-82. DOI: 10.1007/s10654-014-9890-7
7. China Kadoorie Biobank. Available from: <https://www.ckbiobank.org>
8. Hernández JL, Olmos JM, Ramos C, et al. Serum lipids and bone metabolism in Spanish men: the Camargo cohort study. *Endocr J* 2010;57:51-60. DOI: 10.1507/endocrj.k09e-228
9. Pini SF, Pariente E, Olmos JM, et al. Diffuse idiopathic skeletal hyperostosis (DISH) and trabecular bone score (TBS) in postmenopausal women: The Camargo cohort. *Semin Arthritis Rheum* 2023;61:152217. DOI: 10.1016/j.semarthrit.2023.152217
10. Crespo J, Cuadrado A, Perelló C, 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;27:360-70. DOI: 10.1111/jvh.13238
11. Calleja JL, Rivera-Esteban J, Aller R, et al. Prevalence estimation of significant fibrosis because of NASH in Spain combining transient elastography and histology. *Liver Int* 2022;42:1783-92. DOI: 10.1111/liv.15323
12. Riancho J, Vázquez-Higuera JL, Pozueta A, et al. MicroRNA Profile in Patients with Alzheimer’s Disease: Analysis of miR-9-5p and miR-598 in Raw and Exosome Enriched Cerebrospinal Fluid Samples. *J Alzheimers Dis* 2017;57:483-91. DOI: 10.3233/JAD-161179

13. Alonso-Peña M, Dierssen T, Marín MJ, et al.; Cantabria Cohort Collaborators. The Cantabria Cohort, a protocol for a population-based cohort in northern Spain. *BMC Public Health* 2023;23:2429. DOI: 10.1186/s12889-023-17318-8
14. Juhn YJ, Beebe TJ, Finnie DM, et al. Development and initial testing of a new socioeconomic status measure based on housing data. *J Urban Health* 2011;88:933-44. DOI: 10.1007/s11524-011-9572-7
15. 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>
16. Banerji CRS, Chakraborti T, Harbron C, et al. Clinical AI tools must convey predictive uncertainty for each individual patient. *Nat Med* 2023;29:2996-8. DOI: 10.1038/s41591-023-02562-7
17. Cohorte Cantabria. Contamos contigo para mejorar nuestra salud. Disponible en: <https://cohortecantabria.com>
18. O’Rahilly S. Academic clinician-scientists risk becoming an endangered species. *Nat Med* 2023;29:2989. DOI: 10.1038/s41591-023-02626-8
19. Instituto de Salud Carlos III. Disponible en: <https://www.isciii.es>