BMJ Open

Cohort profile: the Hortega Study for the evaluation of non-traditional risk factors of cardiometabolic and other chronic diseases in a general population from Spain

Maria Tellez-Plaza,[1,2] Laisa Briongos-Figuero,[3] Gernot Pichler,[2]
Alejandro Dominguez-Lucas,[2] Fernando Simal-Blanco,[3] Francisco J Mena-Martin,[3]
Jesus Bellido-Casado,[3] Delfin Arzua-Mouronte,[3] Felipe Javier Chaves,[4,5]
Josep Redon,[2,6,7] Juan Carlos Martin-Escudero[3]

For numbered affiliations see end of article.

Correspondence to
Dr Maria Tellez-Plaza;
m.tellez@isciii.es

ABSTRACT

Purpose The Hortega Study is a prospective study, which investigates novel determinants of selected chronic conditions with an emphasis on cardiovascular health in a representative sample of a general population from Spain.

Participants In 1997, a mailed survey was sent to a random selection of public health system beneficiaries assigned to the University Hospital Rio Hortega's catchment area in Valladolid (Spain) (n=11 423, phase I), followed by a pilot examination in 1999–2000 of 495 phase I participants (phase II). In 2001–2003, the examination of 1502 individuals constituted the Hortega Study baseline examination visit (phase III, mean age 48.7 years, 49% men, 17% with obesity, 27% current smokers). Follow-up of phase III participants (also termed Hortega Follow-up Study) was obtained as of 30 November 2015 through review of health records (9.5% of participants without follow-up information).

Findings to date The Hortega Study integrates baseline information of traditional and non-traditional factors (metabolomic including lipidomic and oxidative stress metabolites, genetic variants and environmental factors, such as metals), with 14 years of follow-up for the assessment of mortality and incidence of chronic diseases. Preliminary analysis of time to event data shows that well-known cardiovascular risk factors are associated with cardiovascular incidence rates, which add robustness to our cohort.

Future plans In 2020, we will review updated health and mortality records of this ongoing cohort for a 5-year follow-up extension. We will also re-examine elder survivors to evaluate specific aspects of ageing and conduct geolocation to study additional environmental exposures. Stored biological specimens are available for analysis of new biomarkers. The Hortega Study will, thus, enable the identification of novel factors based on time to event data, potentially contributing to the prevention and control of chronic diseases in ageing populations.

Strengths and limitations of this study

► The main strengths of the Hortega Study (n=1502) include the complex sampling design, the 14-year follow-up for the assessment of incident chronic disease, the extended array of readily available baseline measurements of traditional and non-traditional risk factors, and the associated biological specimens that allow ancillary studies.

► About 54 of baseline (phase III) participants moved out of the catchment area during the follow-up and were losses to follow-up because of no access to their health records; they showed somewhat younger and healthier baseline profile compared with the overall, which may bias the association of risk factors with health endpoints against the null.

► About 40 of baseline participants lacked follow-up information on the health records and their administrative profile was inactive (eg, people using private insurances) and were considered losses to follow-up; they showed a worse baseline risk profile compared with the overall, which may bias the association of risk factors with health endpoints towards the null.

► About 42 of baseline participants lacked follow-up information on the health records and were considered as administrative censures because they did not use the public health system, but they had an active administrative profile and were reasonably healthier at baseline compared with the overall, although we cannot discard that they may have prospectively developed undiagnosed risk factors.

► While the statistical power is only moderate, we empirically detected multi-adjusted HRs for cardiovascular incidence associated with most traditional risk factors in sex strata and the power substantially increases for the analysis of continuous endpoints.

INTRODUCTION

Traditional cardiovascular risk factors do not completely explain the development

of cardiovascular disease (CVD).[1] [2] Given the elevated burden of vascular diseases in increasingly growing ageing populations, identifying novel factors and mechanisms for cardiovascular risk is urgently needed. Mounting evidence point to a number of non-traditional factors that may influence cardiovascular risk including oxidative stress and inflammation,[3–5] metabolic,[6] genetic[7] and environmental factors such as metals.[8] [9] However, population-based studies which evaluate the potential cardiometabolic effects of these factors in representative samples of the general population are scarce.

The Hortega Study is a population-based survey among adults residing in the catchment area of the University Hospital Rio Hortega (UHRH) (Health Department of Eastern Valladolid, Spain), which aims to investigate the cross-sectional and prospective association of genetic, metabolomic (including lipidomics and oxidative stress metabolites) and environmental risk factors with selected chronic conditions, with a special interest on CVD. The follow-up of 1502 Hortega Study participants (baseline examination visit and collection of biological specimens in 2001–2003), added information on mortality and incident health endpoints which occurred as of 30 November 2015.

COHORT DESCRIPTION
Design and setting
The Hortega Study was obtained through a multi-stage complex sampling, which allows to make inferences that are representative of the target general population. In Spain, tertiary hospitals, such as UHRH, have specific geographic areas for patient referral and integrate the network of primary care centres in each area. UHRH is a public hospital of 600 beds with a healthcare coverage close to the 95% of the population living in the area. The Hortega Study phase I started in 1997 by randomly selecting ~20% of the general population 14–85 years in the list of beneficiaries of the public health system assigned to the UHRH catchment area (reference population of 179 505 subjects). Accordingly, during 1997–1998 validated questionnaires were sent by mail to 34 742 subjects with a response rate of ~33% (n=11 423). Subsequently, in 1998, a phone interview was conducted to identify whether characteristics of 500 randomly selected non-responders were similar to responders. The prevalence of self-reported hypertension, diabetes and dyslipidaemia was, respectively, 12.5%, 5.1% and 17.2% for respondents and 14.9%, 5.7% and 12.6% for non-respondents (online supplemental material and table 1). In phase II, we randomly selected 1502 subjects from six strata defined by age and sex among the 11 423 phase I participants who responded to the mail survey, and a second extended questionnaire was mailed to them. About 48.2% of the selected population responded to the questionnaires (724 participants). The distribution of participant characteristics comparing respondents and non-respondents was similar (online supplemental material and table

1). We then conducted a pilot examination among 495 respondents, who were appointed for an examination visit in the medical office.

In phase III (2001–2003), the 1502 participants initially selected in phase II were invited for the Hortega Study baseline examination visit, including the 495 phase II pilot examination visit participants. Among those, 62% accepted and 2% passed away in the period between phases II and III. For a given age and sex strata, we randomly replaced non-respondent participants with other phase I participants who had a similar CVD risk factor profile according to a prime number classification system (see online supplemental material, methods and table 2). As a result, the age, sex and cardiovascular risk factors distributions in the final Hortega phase III sample (n=1502), closely reflect those of the target population. Biological samples including urine, plasma and buffy coat for DNA extraction were collected and stored at −80°C from all Hortega phase III participants. In 2015, the Hortega Follow-up Study was launched as the follow-up of phase III participants, by actively searching in the health records, which included records from hospital admissions, medical office visits, pharmacy information and laboratory tests, both from Primary and Specialised Care. Figure 1 illustrates the study design and timeline.

Data collection
Table 1 summarises the timing and specific details of the different data collected in the Hortega Study. In brief, data collected from the mailed questionnaires in phase I included information such as age, sex, address, education level, weight, height, lifestyle habits (tobacco smoking, alcohol drinking), but also other factors associated with cardiometabolic risk, such as medical history of hypertension, diabetes, dyslipidaemia, CVD and related treatments. The extended questionnaires mailed in phase II to 1500 individuals collected information on cardiovascular risk factors, socio-demographic characteristics (civil status, number of family members, any dependency), healthcare use (medication, visits to the family doctor, emergency room, specialists), physical exercise, quality of life (SF36), nutrition (semi-quantitative diet and two 24hour recall). The phase II pilot examination provided the opportunity to validate the mailing interview. In addition, anthropometric data and blood pressure (BP) values were measured, and urine and blood samples were collected. In phase III, the examination visit of phase II participants (including re-examination of those who participated in the pilot examination), integrated information from an extended array of determinations for the assessment of traditional and non-traditional CVD risk factors and biological sample collection including urine, plasma and buffy coat, which together with the follow-up of phase III participants for the assessment of mortality and incident chronic diseases, constitute the basis for the Hortega Follow-up Study. We provide below specific information of phase III data collection including endpoints assessment during follow-up.
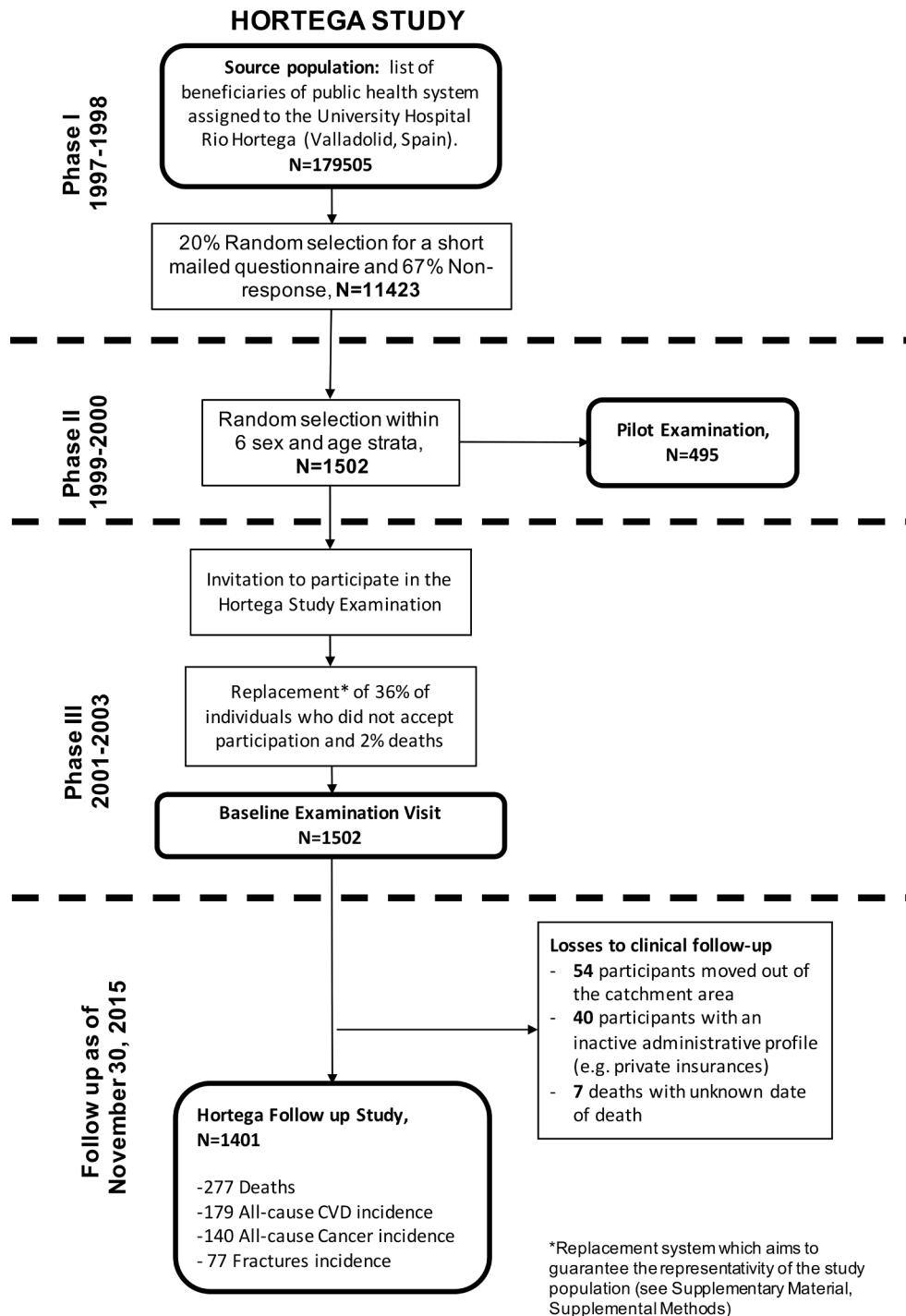
HORTEGA STUDY

Figure 1 Hortega Study design and timeline. The Hortega Study was initiated in 1997 as a mailed survey on a random selection of public health system beneficiaries assigned to the University Hospital Rio Hortega's catchment area in Valladolid (Spain) (n=11 423, phase I), followed by a pilot examination in 1999–2000 of 495 phase I participants (phase II). In 2001–2003, the examination of 1502 individuals (including the re-examination of pilot examination participants) constituted the Hortega Study baseline examination visit (phase III). Follow-up of phase III participants (also termed Hortega Follow-up Study) was obtained as of 30 November 2015 through review of health records. About 9.5% of participants did not have follow-up information (among them 101 participants were considered losses to follow-up and 42 were considered administrative censures because they did not use the public health system, but they had an active administrative profile and were reasonably healthier at baseline compared with the overall).

Cardiovascular risk factors

We had an a priori interest on the prevalence of cardiometabolic factors. Body mass index was calculated using measured height and weight. BP was measured using an OMRON Hem-711C automated device (OMRON Healthcare, USA) following the recommendations of the British Hypertension Society and the Association for the Advancement of Medical Instrumentation. Accordingly, systolic

Table 1	Data items collected in the Hortega follow-up study
Phase	**Measurements**
I (1997)	Questionnaire design and validation, and initial mailing to the ~20% of the beneficiaries of the public health system assigned to the UHRH (n=34 742).
II (1999–2000)	Random selection of 1502 individuals according to age and sex strata. Extended questionnaire by mail and examination and blood samples collection of 495 selected individuals (pilot examination).
III (2001–2003)	Examination of 1502 participants (including re-examination of phase II participants from the pilot examination) and biological sample collection including urine, plasma and buffy coat. Measurements: ► Urine: sodium, potassium, calcium, phosphorus, creatinine, standard (immunogenic) albumin, high-performance liquid chromatography-based (non-immunogenic) albumin, copper, zinc, selenium, lead, cadmium, arsenic, cobalt, chromium, cotinine, GSSG, GSH, 8-oxo-7, 8-dihydroguanine, malondialdehyde, norepinephrine, epinephrine, dopamine, nitrates and nitrites. ► Non-fasting standard plasma biochemistry (information on average hours since the last intake of food was also collected): glucose, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, creatinine, uric acid, sodium, copper, zinc, selenium, aluminium, arsenic, high-sensitivity C reactive protein, thyroid hormones, anti-thyroid antibodies and metabolomics profile. ► Genetic single nucleotide polymorphisms related with cardiovascular risk including the renin–angiotensin–aldosterone system, diabetes, dyslipidaemia and oxidative stress and metal transporters. Targeted next generation sequencing of candidate regions. ► Clinical parameters and medical history: age, gender, weight, height, body mass index, place of residence, waist circumference, hip, thigh and upper arm, social class, education level, profession, marital status, medical services attendance (primary care, emergency room, outpatient clinic, time off work, prescription drug use), bone densitometry (Lunar Pixi), quality of life scales (Euroqol EQ-5D, short-form 36, Profile of Life quality in the Chronically Ill (PLC-PECVEC), Older Americans Resources and Services in subjects older than 65 years), hypertension status, office blood pressure (BP) measurements (two measurements), 24 hours ambulatory monitoring (in 500 subjects with a history of hypertension, newly diagnosed hypertension or in secondary prevention of cardiovascular disease), estimation of physical activity in metabolic equivalents (METS) (during walking and other types of physical activities), Epworth Sleepiness Scale, smoking status (pack-years, time since smoking cessation, age at smoking onset, daily tobacco exposure, Fagerstrom Test for Nicotine Dependence), alcohol consumption and types of alcohol, dietary intake estimation (24 hours food diaries, semi-quantitative survey, nutrition knowledge survey, conversion to nutrients, vitamins and trace elements), Edinburgh Claudication Questionnaire for detecting peripheral arterial disease.
Follow-up study (2015)	13-year follow-up of phase III participants for total mortality and incidence of chronic disease, especially cardiovascular, but also cancer, bone fractures, diabetes, hypertension, dyslipidaemia and chronic kidney disease.

GSH, reduced glutathione; GSSG, oxidised glutathione; UHRH, University Hospital Rio Hortega.

BP (SBP) and diastolic BP (DBP) were the average of two readings measured at 5 min intervals. Hypertension was defined as a mean SBP greater than 140 mm Hg, a mean DBP greater than 90 mm Hg, a recorded physician diagnosis or medication use. Urine cotinine was measured by enzyme-linked immunosorbent assay (Kit 'Análisis DRI Cotinina', Ref. 0395 Microgenics laboratories). Blood samples were obtained after an average of 3 hours since the last intake of food (range 0–17). Non-fasting glucose was measured in the whole study population by the glucose oxidase method with a Hitachi 704 autoanalyzer (Roche Boehringer, Germany). In a second step, fasting glucose determinations were obtained only in participants with non-fasting glucose levels equal or higher than 140 mg/dL. Glycosylated haemoglobin (HbA1c) levels were measured from capilar blood samples with a DCA 2000 HbA1c analyzer (Bayer Diagnostics, USA). Subjects were considered as having diabetes if they had a record of type 2 diabetes diagnosis or a record of use of diabetes medications in the clinical history, if fasting plasma glucose was >126 mg/dL, or if HbA1c was ≥6.5%. Non-fasting total, high-density lipoprotein cholesterol and triglicerides concentrations were determined with an Hitachi 917 autoanalyzer (Roche Boheringer). Low-density lipoprotein (LDL)-cholesterol levels were estimated with the Friedewald formula. Dyslipidaemia was defined as a total cholesterol >200 mg/dL or medication use. Urine and serum creatinine were measured by the modified kinetic Jaffé method by isotope dilution mass spectrometry on a Hitachi 917 analyzer (Roche Boheringer). The glomerular filtration rate was estimated based on serum creatinine determinations (eGFR) by the CKD-EPI abbreviated.[10] Urine albumin was measured by automated nephelometric immunochemistry (Behring, Germany). The ratio of urinary albumin to urinary creatinine (ACR) was reported in milligrams per gram. We

defined abnormal albuminuria as an ACR greater than or equal to 30 mg/g. Chronic kidney disease was defined as the presence of a glomerular filtration rate lower than 60 mL/min/1.73 m² or abnormal albuminuria.

Biomarkers of exposure to trace elements

Total plasma metal and metalloid levels (including arsenic, zinc, selenium and copper) were determined by atomic absorption spectrometry with graphite furnace on a Varian AAS 240 Zeeman (Varian, USA) at Cerba International Laboratories. Total urine metals and metalloids (cadmium, cobalt, copper, molybdenum, zinc, antimony, barium, cadmium, chromium, vanadium, arsenic) and arsenic species concentrations were measured by inductively coupled-plasma mass spectrometry (ICPMS) on an Agilent 7500CEx ICPMS equipped with an octapole collision cell (Agilent Technologies, Japan) and by an ICPMS Thermo XSeries2 (Thermo Scientific, Germany) equipped with an octapole reaction cell, coupled to an anion exchange liquid chromatography (IEC-HPLC) system on a Agilent 1100 (Agilent, USA), respectively, following a standardised protocol in the Laboratory of Environmental Chemistry and Bioanalysis of Huelva University (Spain). The determination of urine arsenic species, including arsenite, arsenate, methylarsonate, dimethylarsinate and arsenobetaine occurred only in a random subsample of 295 individuals, which allows the implementation of multiple imputation strategies for values missing completely at random in the remaining cohort as done by our group in previous studies.[11–13]

Oxidative stress biomarkers

We measured commonly used biomarkers of oxidative stress including the oxidised (GSSG) to reduced (GSH) glutathione ratio (GSSG/GSH), malondialdehyde (MDA) and damaged base 8-oxo-2'-deoxyguanosine (8-oxo-dG). GSSG/GSH, MDA and 8-oxo-dG were measured in urine. GSSG and GSH were analysed by high-performance liquid chromatography (HPLC). MDA was analysed by HPLC and spectrophotometric quantification of MDA-thiobarbituric acid at 532 nm. The amount of 8-oxo-dG in urine was measured by HPLC with electrochemical detection (HPLC-EC). The equipment for oxidative stress biomarkers determinations included a Waters 600 controller including a Waters 515 HPLC pump and two Waters 2487 dual absorbance detectors (Waters, USA).

NMR spectroscopy

We determined the metabolomics spectrum using nuclear magnetic resonance (NMR). Eighty-two microliters of D2O were added to 418 μL of blood plasma and placed in a 5-mm NMR tube. About 1 hour NMR spectra were recorded using a Bruker Avance DRX 600 spectrometer (Bruker GmbH, Germany). Nominal temperature of the sample was kept at 37°C. A single-pulse pre-saturation experiment was acquired in all samples. The spectra were referenced using the doublet of alanine at 1478 ppm.

The chemical shift region, including resonances between 0.50 and 4.70 parts per million (ppm) of spectrometer frequency, was investigated. The spectra were binned into 0.01 ppm buckets and normalised to total aliphatic spectral area to eliminate differences in metabolite total concentration. Signals belonging to selected metabolites were quantified using semi-automated in-house MATLAB V.6.5 integration and peak-fitting routines. Reproducibility of NMR spectroscopy was tested by superposition of normalised spectra of blood serum. Chenomx NMR Suite V.4.5 software and two-dimensional (2D) NMR methods including homonuclear correlation spectroscopy and heteronuclear single quantum correlation spectroscopy were used to identify and subsequently confirm the assessment of metabolites.

In parallel, we also assessed lipid particles profiles using the LIPOSCALE methods for analysis of NMR spectra. About 500 μL of blood plasma samples were shipped on dry ice to Biosfer Teslab (Reus, Spain) for lipoprotein analysis by NMR on a Bruker Avance III 600 spectrometer (Bruker GmbH) operated at proton frequency of 600.20 MHz. Particle concentration and the diffusion coefficients were obtained from the measured amplitudes and attenuation of their spectroscopically distinct lipid methyl group NMR signals using the 2D diffusion-ordered 1 hour NMR spectroscopy pulse. The methyl signal was surface fitted with three Lorentzian functions associated with each lipoprotein subtypes: large, medium and small of the VLDL, LDL and HDL. The area of each Lorentzian function was related to the lipid concentration of each lipoprotein subtype, and the size of each subtype was calculated from their diffusion coefficient. The particle numbers of each lipoprotein subtype were calculated by dividing the lipid volume by the particle volume of a given class. The lipid volumes were determined by using common conversion factors to convert concentration units into volume units.

Candidate single nucleotide polymorphisms

We isolated DNA from peripheral blood cells using Chemagic System (Chemagen, USA), and quantified it with PicoGreen dsDNA Quantification Reagent (Invitrogen, USA). DNA was diluted to a final concentration of 100 ng/μL. We used bibliography searches and the SYSNPS programme, based on public data sources including Ensembl and HapMap (GRCh36), to identify 1251 single nucleotide polymorphisms (SNPs) from 416 candidate genes implicated cardiometabolic pathways, such as inflammation, redox homeostasis, response to insulin or others (see complete list of genes and SNPs in online supplemental material and table 3). We included genes with available preliminary data in support of their association with health outcomes in humans, including genes reported in the literature. The SNPs were genotyped using an oligo-ligation-assay (SNPlex, Applied Biosystems, USA) following the manufacturer's guidelines. Moreover, we have available funding for next

Table 2 Age-adjusted participants' baseline characteristics overall and in subgroups lost to follow-up

	Overall (n=1502)	Patients who did not use the public health system with an active administrative profile (n=42)	Change of catchment area (n=54)	Patients who did not use the public health system with inactive administrative profile (n=40)
Age, years; mean	48.7	36.6	41.2	47.9
Male; %	49.0	62.2	52.7	54.0
Obesity; %*	16.7	6.3	11.4	25.2
Former smoking; %†	28.6	38.2	17.4	38.3
Current smoking; %†	27.0	31.3	22.6	34.0
Diabetes; %	6.3	0.0	3.3	3.0
Glucose lowering medication; %	3.1	0.0	2.2	3.0
Hypertension; %	36.1	7.1	24.9	39.7
Antihypertensive medication; %	14.7	2.2	7.4	12.9
eGFR<60 mL/min/1.72 m^2; %	6.7	0.0	2.1	7.0
Dyslipidaemia; %‡	52.7	40.1	45.4	58.6
Lipid lowering medication; %‡	6.0	0.0	5.3	12.4
Former alcohol intake, %§	9.0	14.6	13.2	3.2
Current alcohol intake, %§	53.1	61.1	56.1	47.1
Prevalent CVD, %	6.3	0.0	0.0	0.0

*58 participants missing obesity.
†20 participants missing smoking.
‡Five participants missing lipids.
§Eight participants missing alcohol.
CVD, cardiovascular disease; eGFR, estimated glomerular filtration rate.

generation sequencing of additional ~200 candidate regions associated to cardiometabolic risk.

Linkage to morbidity, mortality and other routine data sources

In the Hortega Study, the in-person interview and examination visit was only conducted at baseline (phase III). Information for the 14-year follow-up was adjudicated by reviews of clinical records. Patient data was saved from various occasions and locations where they had received care, such as in the primary care physician's office, as well as other physician specialists, nurses' offices, pharmacies, hospitals and emergency departments. We thus extracted patient medications, vital status, medical history and laboratory data, among others, which was linked to the diagnostics and mortality reports generated during hospitalisations. Table 2 shows the distribution of participant characteristics by subgroups with different reasons for not having follow-up information in the health records.

The primary outcome of this study was cardiovascular incidence, including both fatal and non-fatal events. Detailed definitions of specific cardiovascular events follow the guidelines published by the 2014 American College of Cardiology and American Heart Association Task Force of Clinical Data Standards Committee.[14] Participant were considered as having incident diabetes, hypertension, dyslipidaemia or chronic kidney disease, if they were not prevalent cases of such conditions at the time of the baseline examination and they fulfilled the corresponding standard definitions during the follow-up based on the health records. Time to event was calculated as the difference between the date of the baseline examination and the date of the event, the date of death or 30 November 2015 (administrative censoring), whichever occurred first. Records from all phase-III Hortega participants were reviewed by a committee composed of two physician reviewers who assigned incident events with no a priori knowledge of who may or may have not developed CVD events. We additionally collected information on incidence of other chronic diseases. For instance, incident cancer cases and associated time to event variables were defined based on the date of the first positive pathology report in the health record. Incident fractures (mainly, hip, humerus, vertebral and Colles fractures) and associated time to event were defined based on the date of the first pathological imaging technique (including X-ray, CT scan or NMR).

Patient and public involvement

Patients were not involved in the Hortega Follow-up Study as the follow-up data of Hortega Phase III participants was obtained by review of electronic health records.

Table 3 Age-adjusted rates (cases/10 000 person-years) of mortality and combined CVD incidence by risk factors in CVD-free participants*

	Men					Women				
		Total mortality		Combined CVD			Total mortality		Combined CVD	
	n (%)	Cases	Rate	Cases	Rate	n (%)	Cases	Rate	Cases	Rate
Overall	599 (100)	121	100.4	82	84.4	630 (100)	75	49.5	65	52.9
Age group (years)										
<50	300 (58.5)	6	14.8	8	20.9	329 (57.2)	2	4.6	5	12.2
50–65	112 (22.4)	8	54.3	10	70.2	96 (17.4)	4	31.1	7	55.6
65–75	87 (10.2)	33	283.1	29	315.0	96 (13.3)	20	134.9	21	166.1
≥75	100 (8.9)	74	835.3	35	476.6	109 (12.1)	49	418.5	32	305.8
BMI										
<30	502 (84.2)	97	98.3	69	89.7	517 (83.2)	50	48.2	42	49.4
≥30	97 (15.8)	24	113.4	13	66.0	113 (16.4)	25	54.4	23	62.8
Smoking										
Never	196 (33.0)	29	66.8	20	56.6	378 (56.4)	68	49.5	55	42.8
Former	241 (36.7)	79	118.8	49	75.3	111 (19.1)	4	36.9	5	46.4
Current	162 (30.3)	13	109.9	13	106.4	141 (24.5)	3	65.2	5	96.5
Hypertension										
No	352 (64.8)	28	79.4	23	77.5	384 (64.8)	16	49.8	10	31.7
Yes	247 (35.2)	93	132.5	59	118.0	246 (35.2)	59	59.2	55	83.4
Diabetes mellitus										
No	545 (92.7)	91	96.5	61	77.6	598 (95.8)	61	47.4	54	51.1
Yes	54 (7.3)	30	157.1	21	179.9	32 (4.2)	14	109.3	11	140.6
Dyslipidaemia										
No	395 (66.4)	83	106.9	46	71.9	402 (65.3)	38	51.0	28	48.8
Yes	204 (33.6)	38	87.0	36	111.0	228 (34.7)	37	47.0	37	59.8

Combined CVD definition included fatal and non-fatal CHD, stroke, heart failure and peripheral arterial disease.
*We excluded 81 participants missing baseline obesity, smoking or lipid determinations, 89 participants lost to follow-up, 7 participants who died but the date of death was unknown and 96 participants with prevalent coronary heart disease, stroke, heart failure or peripheral arterial disease at baseline, leaving 1229 individuals for the prospective analysis. Summary age-adjusted rates for all-cause mortality and combined cardiovascular incidence were estimated from Poisson regression models for individual mortality and cardiovascular incidence time-to-event data (endpoint status at the end of follow-up as dichotomous outcome and log-transformed individual follow-up time as offset). All analyses were weighted to the underlying population. Average and absolute amount of follow-up was, respectively, 12.5 and 15 402.8 person-year for mortality and 12.0 and 14 784.4 person-year for combined CVD.
BMI, body mass index; CVD, cardiovascular disease.

Future plans

Starting in 2020, we will review updated health records to extend the follow-up for additional 5 years. We will also be periodically obtaining updated mortality information of the Hortega Study participants from the Spanish National Institute for Statistics. Moreover, the geocoding of the participants' home addresses (which will be conducted in collaboration with the Department of Chronic Diseases Epidemiology at the National Center of Epidemiology, Instituto de Salud Carlos III, Madrid) will allow to evaluate additional environmental factors that may influence health. In addition, there is a plan to re-contact and re-examine survivors older than 65-year-old (n=600) with the objective to evaluate specific aspects of ageing. Another prospective goal is to enrich already-collected data through the addition of '-omics' data (ie, DNA methylation or exome sequencing) as well as new phenotypic data.

FINDINGS TO DATE

The phase III Hortega Examination visit has generated numerous cross-sectional studies of traditional (including diabetes,[15] hypertension[16] and dyslipidaemia[17]) and non-traditional (genetic,[15 17–20] metabolomics[21 22] and environmental[23–27]) cardiovascular risk factors. The follow-up data from the Hortega Study will allow to conduct time to event survival analysis and other prospective studies which will provide novel evidence on the potential role of genetic, environmental and metabolomic factors in

Table 4 HR (95% CI) for combined CVD incidence by traditional risk factors in CVD-free study participants (N=1229*).

	Overall	Men	Women
Current smoker	1.55 (1.09 to 2.23)	1.29 (0.82 to 2.03)	2.25 (0.96 to 5.30)
Obesity	0.95 (0.77 to 1.17)	0.71 (0.48 to 1.05)	1.17 (0.84 to 1.63)
Hypertension	1.83 (1.40 to 2.40)	1.61 (1.17 to 2.23)	2.30 (1.22 to 4.33)
Diabetes	2.09 (1.55 to 2.81)	2.39 (1.44 to 3.96)	1.67 (0.81 to 3.42)
Dyslipidaemia	1.35 (1.12 to 1.63)	1.17 (0.87 to 1.57)	1.60 (1.09 to 2.35)

Combined CVD definition included fatal and non-fatal CHD, stroke, heart failure and peripheral arterial disease. [a]We excluded 81 participants missing baseline obesity, smoking or lipid determinations, 89 participants lost to follow-up, 7 participants who died but the date of death was unknown and 96 participants with prevalent coronary heart disease, stroke, heart failure or peripheral arterial disease at baseline, leaving 1229 individuals for the prospective analysis. Rate ratios for cardiovascular incidence associated with the presence of traditional cardiovascular risk factors were estimated from Poisson regression models of individual time-to-event data adjusting for age at follow-up, sex and traditional cardiovascular risk factors including smoking status (never, former, current), obesity (no, yes), dyslipidaemia (total cholesterol ≥200 mg/dL or lipid lowering medication), diabetes (no, yes), hypertension (no, yes), low-estimated glomerular filtration rate (no, yes), anti-hypertensive medication (no, yes), glucose lowering medication (no, yes) and lipid lowering medication (no, yes). All analyses were weighted to the underlying population.
CVD, cardiovascular disease.

the pathogenesis of chronic disease, mainly cardiovascular, but also other such as cancer, and bone disease. So far, we have identified 277 deaths and several incident endpoints, such as all-cause CVD (n=179), all-cancer (n=140) and fractures (n=77). Table 3 shows age-adjusted total mortality and cardiovascular incidence rates by subgroups defined by traditional cardiovascular risk factor, which are consistent with published data from other cardiovascular cohorts and add robustness to our cohort.

Strengths and limitations

The main strength of the Hortega Study is that it is a representative sample of a general population from Spain, with associated biological specimens and follow-up for assessment of incident chronic diseases, with a special interest in traditional and non-traditional factors associated with cardiometabolic risk, and an extended array of readily available measurements including cardiovascular risk factors and treatments, dietary and physical activity habits, anthropometric measures, standard plasma biochemistry profile, urine oxidative stress biomarkers, genetic polymorphisms, quality of life and lifestyle-related questionnaires, biomarkers of environmental exposures, such as urinary metals and metabolic compounds from metabolomics platforms. Men and women and a wide age range are represented, including participants from rural and urban areas. Another strength is provided by the consent form, which was approved by the corresponding Ethics committees and allow ancillary studies. Thus, future studies making use of the Hortega Study data and samples are possible.

The Hortega Follow-up Study is not exempt of limitations. In phase III, only 62% of individuals who answered the questionnaires in initial stages of the sampling process, agreed to participate in the physical examination. However, we implemented a complex stratified system based on prime numbers to randomly replace non-participating individuals by other who did belong to the same poli-strata defined by cardiovascular risk profile, sex and age (see online supplemental maerial, methods and table 1). Consequently, the distribution of cardiovascular risk factors in the Hortega Study baseline examination visit (phase III) is the same as in the preliminary random selection of participants. However, during the follow-up period 7 participants died with unknown death date, 54 participants were assigned to another catchment area and 82 did not use the public health system (among them 42 with and 40 without, active administrative profile in file). Participants with an inactive administrative profile show a worse cardiovascular risk profile at baseline compared with the overall (table 2) and could have been out-of-catchment area deaths, people that changed address without having communicated it to the public health system or people using private insurances, and were considered loses to follow-up, which could have biased estimated relative risks towards the null. Participants lost to follow-up due to a confirmed change of catchment area show a somewhat healthier cardiovascular profile compared with the overall, which could bias the estimated relative risks against the null. Participants with active administrative profile in file but who did not contact the health system showed a markedly healthier cardiovascular profile at baseline, and it is likely that they did not undergo any hospitalisation. Consequently, we considered individuals with active administrative profile as reasonably 'healthy' and censored them at the end of the follow-up. We cannot discard, however, that some of these participants may have developed cardiovascular risk factors during the follow-up that have remained undiagnosed.

Other weaknesses include the lack of fasting plasma glucose determinations in the whole study population. Only individuals with suggestive evidence of altered fasting glucose levels underwent a second measurement in fasting conditions. This strategy allows for analysis of diabetes status but not for associations with glycaemia

as a continuous measure. An additional limitation is the moderate sample size. However, we have enough power to detect multi-adjusted HRs for cardiovascular incidence associated with most traditional risk factors in men and women (table 4). Table 4 empirically shows that we have enough power to detect a minimum of 60% change in cardiovascular risk associated with dichotomous variables in sex-specific analyses, when the number of cases among exposed is as low as 37 (eg, the association of dyslipaemia and combined CVD in women shown in tables 3 and 4). For analysis using continuous explanatory variables the power increases substantially.

Finally, the data are not publicly available as the steering committee and the participants have not approved unrestricted data sharing at the time of approval of the study by the corresponding ethics committee and unrestricted data sharing was not included in the consent form. There are however opportunities for collaboration and for ancillary studies proposals as data are available on reasonable request (see the Data availability statement section).

COLLABORATIONS

The available biological specimens (plasma, buffy coat, DNA) will add value to our cohort as potentially novel biomarkers of disease are identified and measured through prospective collaborations. The Hortega Study, a cohort from a south European population with elevated prevalence of traditional cardiovascular risk factor but low burden of disease, thus, offers substantial opportunities for collaborative proposals aiming to the identification of novel biological mechanisms for disease and risk factors, which in turn could contribute to the prevention and control of chronic diseases in ageing populations.

Author affiliations
[1]Department of Chronic Diseases Epidemiology, National Center for Epidemiology, Instituto de Salud Carlos III (ISCIII), Madrid, Madrid, Spain
[2]Area of Cardiometabolic and Renal Risk, Biomedical Research Institute Hospital Clinic de Valencia (INCLIVA), Valencia, Valencia, Spain
[3]Department of Internal Medicine, Hospital Universitario Rio Hortega, Valladolid, Valladolid, Spain
[4]Genomic and Genetic Diagnosis Unit, INCLIVA, Valencia, Valencia, Spain
[5]CIBER of Diabetes and Associated Metabolic Diseases (CIBERDEM), ISCIII, Madrid, Madrid, Spain
[6]Deparment of Internal Medicine, Hospital Universitario Clinic de Valencia, Valencia, Valencia, Spain
[7]CIBER Physiopathology of Obesity and Nutrition (CIBEROBN), ISCIII, Madrid, Madrid, Spain

REFERENCES

1. Laclaustra M, Casasnovas JA, Fernández-Ortiz A, et al. Femoral and carotid subclinical atherosclerosis association with risk factors and coronary calcium: the AWHS study. *J Am Coll Cardiol* 2016;67:1263–74.
2. Patel MR, Peterson ED, Dai D, et al. Low diagnostic yield of elective coronary angiography. *N Engl J Med* 2010;362:886–95.

3. Ridker PM, MacFadyen JG, Everett BM, *et al*. Relationship of C-reactive protein reduction to cardiovascular event reduction following treatment with canakinumab: a secondary analysis from the CANTOS randomised controlled trial. *Lancet* 2018;391:319–28.

4. Shah A, Gray K, Figg N, *et al*. Defective base excision repair of oxidative dna damage in vascular smooth muscle cells promotes atherosclerosis. *Circulation* 2018;138:1446–62.

5. Patel RS, Ghasemzadeh N, Eapen DJ, *et al*. Novel biomarker of oxidative stress is associated with risk of death in patients with coronary artery disease. *Circulation* 2016;133:361–9.

6. Cheng S, Shah SH, Corwin EJ, *et al*. Potential impact and study considerations of metabolomics in cardiovascular health and disease: a scientific statement from the American heart association. *Circ Cardiovasc Genet* 2017;10.

7. Shah SH, Newgard CB. Integrated metabolomics and genomics. *Circulation* 2015;8:410–9.

8. Tellez-Plaza M, Guallar E, Navas-Acien A. Environmental metals and cardiovascular disease. *BMJ* 2018:k3435.

9. Chowdhury R, Ramond A, O'Keeffe LM, *et al*. Environmental toxic metal contaminants and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* 2018;362:k3310.

10. Levey AS, Stevens LA, Schmid CH, *et al*. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–12.

11. Tellez-Plaza M, Navas-Acien A, Crainiceanu CM, *et al*. Cadmium and peripheral arterial disease: gender differences in the 1999-2004 US National Health and Nutrition Examination Survey. *Am J Epidemiol* 2010;172:671–81.

12. Tellez-Plaza M, Navas-Acien A, Menke A, *et al*. Cadmium exposure and all-cause and cardiovascular mortality in the U.S. general population. *Environ Health Perspect* 2012;120:1017–22.

13. Grau-Perez M, Navas-Acien A, Galan-Chilet I, *et al*. Arsenic exposure, diabetes-related genes and diabetes prevalence in a general population from Spain. *Environ Pollut* 2018;235:948–55.

14. Hicks KA, Tcheng JE, Bozkurt B, *et al*. 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials: A Report of the American College of Cardiology/ American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Cardiovascular Endpoints Data Standards). *J Am Coll Cardiol* 2015;66:403–69.

15. Mansego ML, Martínez F, Martínez-Larrad MT, *et al*. Common variants of the liver fatty acid binding protein gene influence the risk of type 2 diabetes and insulin resistance in Spanish population. *PLoS One* 2012;7:e31853.

16. Llopis-González A, Rubio-López N, Pineda-Alonso M, *et al*. Hypertension and the fat-soluble vitamins A, D and E. *Int J Environ Res Public Health* 2015;12:2793–809.

17. Abellán R, Mansego ML, Martínez-Hervás S, *et al*. Dietary polyunsaturated fatty acids may increase plasma LDL-cholesterol and plasma cholesterol concentrations in carriers of an ABCG1 gene single nucleotide polymorphism: study in two Spanish populations. *Atherosclerosis* 2011;219:900–6.

18. Abellán R, Mansego ML, Martínez-Hervás S, *et al*. Association of selected ABC gene family single nucleotide polymorphisms with postprandial lipoproteins: results from the population-based Hortega study. *Atherosclerosis* 2010;211:203–9.

19. Martinez-Hervas S, Mansego ML, de Marco G, *et al*. Polymorphisms of the UCP2 gene are associated with body fat distribution and risk of abdominal obesity in Spanish population. *Eur J Clin Invest* 2012;42:171–8.

20. Martínez-García F, Mansego ML, Rojo-Martínez G, *et al*. Impact of obesity-related genes in Spanish population. *BMC Genet* 2013;14:111.

21. Marrachelli VG, Monleon D, Rentero P, *et al*. Genomic and metabolomic profile associated to microalbuminuria. *PLoS One* 2014;9:e98227.

22. Marrachelli VG, Rentero P, Mansego ML, *et al*. Genomic and metabolomic profile associated to clustering of cardio-metabolic risk factors. *PLoS One* 2016;11:e0160656.

23. Galan-Chilet I, Tellez-Plaza M, Guallar E, *et al*. Plasma selenium levels and oxidative stress biomarkers: a gene-environment interaction population-based study. *Free Radic Biol Med* 2014;74:229–36.

24. Galan-Chilet I, Guallar E, Martin-Escudero JC, *et al*. Do genes modify the association of selenium and lipid levels? *Antioxid Redox Signal* 2015;22:1352–62.

25. Martínez-Barquero V, de Marco G, Martínez-Hervás S, *et al*. Polymorphisms in endothelin system genes, arsenic levels and obesity risk. *PLoS One* 2015;10:e0118471.

26. Grau-Perez M, Pichler G, Galan-Chilet I, *et al*. Urine cadmium levels and albuminuria in a general population from Spain: A gene-environment interaction analysis. *Environ Int* 2017;106:27–36.

27. Galan-Chilet I, Grau-Perez M, De Marco G, *et al*. A gene-environment interaction analysis of plasma selenium with prevalent and incident diabetes: The Hortega study. *Redox Biol* 2017;12:798–805.