Special article

Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design

Gemma Castaño-Vinyals a, b, c, q, s, u, Nuria Aragonés d, q, r, u, Beatriz Pérez-Gómez d, q, r, Vicente Martín e, q, Javier Llorca f, q, Victor Moreno g, h, q, Jone M. Altzibar i, q, Eva Ardanaz j, q, Silvia de Sanjóse d, q, José Juan Jiménez-Moleón k, q, Adonina Tardón l, q, Juan Alguacil m, q, Rosana Peiró n, q, Rafael Marcos-Gradera o, q, Carmen Navarro o, q, s, Marina Pollán d, q, r, u, Manolis Kogevinas a, b, c, q, t, u, MCC-Spain Study Group

a Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
b IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
c Universitat Pompeu Fabra (UPF), Barcelona, Spain
d Environmental and Cancer Epidemiology Unit, National Center of Epidemiology, Instituto de Salud Carlos III, Madrid, Spain
e Universidad de León, León, Spain
f Universidad de Cantabria, Santander, Spain
g DIBELL-Catalan Institute of Oncology, L’Hospitalet de Llobregat, Spain
h Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain
i Subdirección de Salud Pública de Cipuzkoa, Donostia, Spain
j Instituto de Salud Pública de Navarra, Pamplona, Navarra
k Instituto de Investigación Biosanitaria de Granada (ibs.CRANADA), Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain
l Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Asturias, Spain
m Centro de Investigación en Salud y Medio Ambiente (CYSMA), Universidad de Huelva, Huelva, Spain
n Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana FISABIO–Salud Pública, Valencia, Spain
o Epidemiology Unit and Girona Cancer Registry, Oncology Coordination Plan, Department of Health, Autonomous Government of Catalonia, Catalan Institute of Oncology, Girona Biomedical Research Institute (idBGI), Girona, Spain
p Department of Epidemiology, Murcia Regional Health Authority, Murcia, Spain
q CIBER Epidemiologia y Salud Pública (CIBERESP), Madrid, Spain
t IIS Puerta de Hierro, Majadahonda, Spain
ut Department of Health and Social Sciences, Universidad de Murcia, Murcia, Spain
u School of Public Health, Athens, Greece

A R T I C L E   I N F O

Article history:
Received 31 July 2014
Accepted 12 December 2014
Available online 19 January 2015

Keywords:
Case-control
Epidemiology
Colorectal cancer
Prostate cancer
Breast cancer
Gastric cancer
Chronic lymphocytic leukemia

A B S T R A C T

Introduction: We present the protocol of a large population-based case-control study of 5 common tumors in Spain (MCC-Spain) that evaluates environmental exposures and genetic factors.

Methods: Between 2008-2013, 10,183 persons aged 20-85 years were enrolled in 23 hospitals and primary care centres in 12 Spanish provinces including 1,115 cases of a new diagnosis of prostate cancer, 1,750 of breast cancer, 2,171 of colorectal cancer, 492 of gastro-oesophageal cancer, 554 of chronic lymphocytic leukaemia (CLL) and 4,101 population-based controls matched by frequency to cases by age, sex and region of residence. Participation rates ranged from 57% (stomach cancer) to 87% (CLL cases) and from 30% to 77% in controls. Participants completed a face-to-face computerized interview on sociodemographic factors, environmental exposures, occupation, medication, lifestyle, and personal and family medical history. In addition, participants completed a self-administered food-frequency questionnaire and telephone interviews. Blood samples were collected from 76% of participants while saliva samples were collected in CLL cases and participants refusing blood extractions. Clinical information was recorded for cases and paraffin blocks and/or fresh tumor samples are available in most collaborating hospitals. Genotyping was done through an exome array enriched with genetic markers in specific pathways. Multiple analyses are planned to assess the association of environmental, personal and genetic risk factors for each tumor and to identify pleiotropic effects.

Discussion: This study, conducted within the Spanish Consortium for Biomedical Research in Epidemiology & Public Health (CIBERESP), is a unique initiative to evaluate etiological factors for common cancers and will promote cancer research and prevention in Spain.

© 2014 SESPAS. Published by Elsevier España, S.L.U. All rights reserved.
Estudio multicaso-control de base poblacional de tumores comunes en España (MCC-Spain): razón y diseño del estudio

RESUMEN

Introducción: Presentamos el protocolo del estudio caso-control de base poblacional de 5 tumores comunes en España (MCC-Spain) que evalúa factores ambientales y genéticos.

Métodos: Durante 2008-2013, se reclutaron 10.183 sujetos entre 20-85 años en 23 hospitales de 12 provincias españolas, incluyendo 1.115 casos de cáncer de próstata, 1.750 de mama, 2.171 colorrectal, 492 gastro-esofágicos, 554 de leucemia linfática crónica (LLC) y 4.101 controles poblacionales emparejados por frecuencia por edad, sexo y región de residencia. Las tasas de participación varían del 57% (cáncer de estómago) al 87% (casos de LLC) y del 30% al 77% en controles. Los participantes respondieron una entrevista personal informatizada sobre factores socio-demográficos, exposiciones ambientales, ocupación, medicación, estilos de vida, e historia médica personal y familiar. Además, cumplimentaron un cuestionario alimentario y realizaron entrevistas telefónicas. Se recogió sangre del 76% de los participantes y saliva para los casos de LLC y participantes que rechazaron la donación de sangre. En los casos, se recogió información clínica y se dispone de muestras de tumor fresco o parafinado a través de los biobancos de los hospitales. Se realizó el genotipado con un array de exoma suplementado con marcadores en rutas específicas. Se han planificado diversos análisis para evaluar la asociación de factores genéticos, personales y ambientales para cada tumor e identificar efectos pletóricos.

Discusión: Este estudio, desarrollado en el Consorcio de Investigación Biomédica de Epidemiología y Salud Pública (CIBERESP), es una iniciativa única para evaluar factores etiológicos de tumores comunes y promoverá la investigación en cáncer y prevención en España.

© 2014 SESPAS. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.
similar age, sex and hospital catchment area were randomly selected from the general practitioner lists. If contact with the first person of this list was not possible after a minimum of five tries at different times of the day, if he/she refused to participate, the following person of the list was approached.

In Table 2 we present the main characteristics of the study population.

Response rates

Response rates were calculated using subjects interviewed in the numerator, and all subjects including refusals in the denominator. For cases, these response rates were 68% for colorectal cancer cases, 71% for breast, 72% for prostate, 57% for gastric and 87% for CLL. In controls, mean participation rate of controls was 53% and varied by region. For 22% of the subjects the phone contact was not possible due wrong phone number or no-answer.

Questionnaires, biological samples, hospital records and anthropometric measurements

A structured computerized epidemiological questionnaire was administered by trained personnel in a face-to-face interview (http://www.mccspain.org). The average duration of the interview was 70 minutes (range 30–130). Information was collected on socio-demographic factors, residential history, lifelong retrospective environmental exposures, including water consumption and use (showering, bathing, swimming in pools), occupational history—including night shift—, medication, lifestyles—smoking, alcohol consumption, physical activity, use of cosmetic products—, sun-bathing and sleeping habits, personal/family medical history and quality of the interview. Missing values on key variables and specific questions on additional study objectives (e.g. questions on disruption of the circadian rhythm) were completed through subsequent telephone contact. The main characteristics of the study population are presented in Table 2. After the interview, biological samples and anthropometric data were obtained following the study protocol. Height and weight at different ages were self-reported and waist and hip circumference were measured with a tape. Subjects were provided a semi-quantitative Food Frequency Questionnaire (FFQ), which was a modified version from a previously validated instrument in Spain to include regional products. It included 140 food items, and assessed usual dietary intake during the previous year. Portion sizes were specified for each item, and photographs were used to define degrees of doneness. Information on consumption of vitamin and mineral supplements and on important changes on dietary habits in the past 5 years were also collected. The FFQ was self-administered and returned by mail or filled out face to face (global response rate 88%).

When feasible, 27 ml of peripheral blood was drawn from participants, which were aliquoted in whole blood, plasma, cellular fraction for DNA extraction, and serum and stored at −80 °C. Saliva was collected for subjects refusing to donate blood and for all CLL cases, with the Oragene® DNA Kit and stored at room temperature until DNA extraction. We collected biological samples for DNA extraction for 96% of participants with interview (76% blood and 27% saliva) as well as toenail and hair samples were taken from participants (79% and 84% respectively). In 4 centres (Madrid, Cantabria, Asturias and Huelva, which include approximately 1/3 of the study participants) cases and controls also donated urine samples (60 ml) that were aliquoted and frozen at − 80 ° C. Fresh tumour biopsies or paraffin embedded samples are available in all participating hospitals.

Standardised basic clinical and pathological information on the diagnosis and treatment of tumours was collected from hospital records using a predefined format.

Sociodemographic, lifestyle and environmental factors

MCC-Spain will examine the Socioeconomic status will be examined using multilevel approaches that allow the evaluation of the role of structural socioeconomic factors on health. Environmental justice, proximity to green spaces and environmental pollution will be assessed through an evaluation of exposures proximate to the place of residence.

Lifestyle exposures are one of the main objectives of this study. Diet is examined through summary intakes of relevant food groups based on reported intake frequencies and portion size information. Food composition tables were developed and will be combined with reported intake frequencies and cooking method preferences to estimate intakes of nutrients, food contaminants (e.g. polycyclic aromatic hydrocarbons) and food properties (e.g. total antioxidant capacity). An a priori Mediterranean diet score, alternative diet patterns and a factor-analysis derived diet pattern will be examined. General and central obesity is examined together with leisure time physical activity and sedentary lifestyle. Numerous other potential risk factors that could be associated with the cancers investigated are examined. These include, among others, smoking habits, exposure to medical radiation, use of cosmetics, use of tight clothes and belts, exposure to sun or sleeping patterns.

Use of medical drugs was collected through personal interviews, mainly by indication. Information was coded (Anatomical Therapeutic Chemical-ATC code) to assess individual exposure to different drugs including statins and anti-inflammatory drugs, analgesics, hormones, antihypertensives, beta-blockers, bisphosphonates and corticosteroids.

Hormonal factors and endocrine disruption are also examined. Sex dimorphic phenotypes (finger ratio and anogenital
distance) will be evaluated in relation to the development of breast and prostate cancer. The ratio of the length of the index finger and middle finger of both hands (2D: 4D) were measured using callipers with a resolution of 0.05 mm. A validation study has shown a good repeatability of finger measurements. Anogenital distance was assessed in a subgroup of prostate cancer cases and controls. The influence of reproductive history, hormonal treatments (contraceptives, hormone replacement therapy) and influence of hormonal development (pattern of fat distribution at different ages, and height) will be examined. Endocrine disruptors (xenoestrogens and other persistent organic pollutants) will be measured in serum through a determination of the global xenoestrogenic burden (TEXB).

Among environmental pollutants, evaluation of drinking water contaminants focuses on disinfection by products (such as trihalomethanes and haloacetic acids), nitrates and metals. Exposure data from water companies and municipalities, national surveillance data and new water analyses has been gathered and modelled to estimate historical levels of pollutants in drinking water and combined with individual data from the questionnaire. Urinary trichloroacetic acid was measured in a subset of controls. MCC-Spain will also study environmental exposure to different metals, including Cd, Ni, Cr, As, Pb, Se, and Zn in relation to the five combining biomarker-based estimations with information based on the epidemiological information.

Occupational exposures will also be studied. All jobs conducted for more than one year were recorded with information regarding specific tasks, exposures and timing of the job. Jobs were coded by two experts following the Spanish National Classification of Occupations (CNO-94). The Spanish JEM, MatEmEsp, will be applied. Detailed information on work shift (rotating and night) and disruption of the circadian rhythm was also collected.

Other possible etiological factors have also been included in the project. Among them, several infections will be evaluated in relation to colorectal and gastric tumours and CLL. The role of *H. pylori* infection will be estimated using seroprevalence against several virulence antigens. In relation to CLL, seroprevalence of several polymaviruses, herpesviruses, and *Chlamydia trachomatis* will be evaluated. Additionally the antibody response pattern to EBV will also be measured.

Extensive information on family history was collected to identify familial cases. This information will allow describing the typical family structure of study participants and, if genetic effects are identified, estimating their penetrance using Kin Cohort methods. In addition, genetic analyses will be carried out within MCC-Spain and also through participation in international consortia such as the prostate cancer consortium PRACTICAL (http://ccge.medschl.cam.ac.uk/consortia/practical/). The Infinium HumanExome BeadChip from Illumina was used to genotype >200,000 coding markers plus 6,000 additional custom variants on the pathways of interest such as inflammation, circadian rhythm or detoxification.

**Ethics and availability of data**

The protocol of MCC-Spain was approved by the Ethics committees of the participating institutions. All participants were informed about the study objectives and signed an informed consent. Confidentiality of data is secured removing personal identifiers in the datasets. The database was registered in the Spanish Agency for Data Protection, number 2102672171. Permission to use the study database will be granted to researchers outside the study group, after revision and approval of each request by the Steering Committee. More details on the organization of the project can be found online at http://www.mccspain.org.

**DISCUSSION**

Cancer locations and hypotheses examined in MCC-Spain were selected with a public health perspective to provide information
useful for cancer prevention. MCC-Spain also aims to foster a network of research in cancer epidemiology in Spain.

The option of a single set of population controls and a single questionnaire for all tumours has the main benefit of the optimization of the economic resources. This approach has been successfully used previously. On the other hand, the main drawback is the need to use the same tools to gather the information regarding risk factors for all types of cases.

The advantages and problems of the selection of population versus hospital controls have been extensively discussed. The evaluation of a variety of exposures makes hospital controls less suitable given the potential association between multiple diseases and exposures of interest. As expected from other population-based studies, participation rates of cases were higher than those of controls. Selecting controls through lists of general practitioners provides a representative sampling frame given the almost universal public coverage of the national health system in Spain. However, errors in these lists concerning personal data resulted in a lower than expected response rate.

The study is population-based since we intended to recruit all cases with a first diagnosis of the studied tumors in the selected health areas, using for this purpose the reference hospital/hospitals in each area and identifying every new diagnosis of the studied cancers. We could not use population cancer registries to ascertain the number of cases lost since in most of the regions included in the study there was not any such registry, but we can certainly assume to be few. Potential misclassification of exposure is a major limitation of case-control studies. The implementation of a computerized questionnaire, training and continuous feedback to interviewers, and repeated interviews to complete missing values is likely to reduce errors.

A challenge of current cancer epidemiology research is to accurately define the molecular phenotype of tumours so that specific risk factors can be identified for each molecular subtype. All tumours have pathology slides in the reference hospitals that can be retrieved and some hospitals also have tumour banks that have collected fresh tumour tissue for some cases.

Finally, networking is among the major achievements of the study. MCC-Spain includes 17 different centres and has followed organizational procedures to promote the exchange of knowledge and experiences between centres.

Editor in charge

Alberto Ruano-Ravina.

Statement of authorship

All authors have contributed to the conception and design of the study, and have acquired data, have been involved in drafting the manuscript. All authors read and approved the final manuscript.

Funding


Samples: Biological samples were stored at the Parc de SalutMAR Biobank (MARBiobanc; Barcelona) which is supported by Instituto de Salud Carlos III FEDER (RD09/0076/00036). Also at the Public Health Laboratory to Gipuzkoa and the Basque Biobank. Also sample collection was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncologia de Catalunya (XBTC). Biological samples were stored at the “Biobanco La Fe” which is supported by Instituto de Salud Carlos III (RD 09 0076/00021) and FISABIO biobanking, which is supported by Instituto de Salud Carlos III (RD09 0076/00058).

Genotyping: SNP genotyping services were provided by the Spanish “Centro Nacional de Genotipado” (CEGEN-ISCI) and the Basque Biobank.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We thank all the subjects who participated in the study and all MCC-Spain collaborators (the lists can be found below).

Anex 1.

MCC–Spain study group: Maria Teresa Alonso, Pilar Amiano, Cristina Ares, Mikel Azpiri, Yolanda Benavente, Elena Boldo, Aurora Bueno, Mariona Bustamante, Francisco Javier Caballero, Elías Campo, Rafael Cantón, Rocío Capelo, Carme Carmona, Delphine Casabonne, María Dolores Chirlaque, Judith Cirac, Juan Clofent, Enrique Colado, Laura Costas, Marta Crous, Rosa del Campo, Marian Díaz Santos, Trinidad Dierssen-Sotos, María Ederra, Ana Espinosa, Marieta Fernández Cabrera, Ana Fernández Somoano, Tania Fernández Villa, Esther García-García-Espinosa, Paloma García Martín, Inés Gómez-Acebo, Cristina González Puga, Esther Gracia, Marcela Guevara Eslava, Elisabet Guínó, José María Huerta, Virginia Lope, Gonzalo López-Abente, Carlos Lopez-Otín, Beguña Martínez Argüelles, Sergio Merino Salas, Benito Mirón Pozo, Antonio José Molina de la Torre, Eduardo Moreno, Concepción Moreno Iribas, Nicolás Olea, Gemma Osca Gelas, Laia Paré, Miquel Porta, Montse Puig, Manuel Rivas del Fresno, Claudia Robles, Marta María Rodríguez Suarez, Beatriz Romero, Ana Isabel Sáez Castillo, María Sala Serra, Dolores Salas Trejo, Ana Santaballa, Miguel Santibáñez, Ángeles Sierra, Ana Souto, Cristina M Villanueva.

Annex 2.

BARCELONA

Marín; Biobanco: Manuel Posada, Juan Cosmen, Ana Villanueva; Centro Nacional de Sanidad Ambiental: Argelia Cañado, José Antonio Jiménez, Carmen Navarro.

Demográfica: Ana Rin, Gema Díaz, Marta Herreros, Virginia Pedrazá, Patricia López, Miguel de la Fuente.


Atención Primaria: Gerencia Área 5; Ricardo Rodríguez Barrientos; Centro de Salud Barrio del Pilar: Gerardo López, Alberto Fernández, Ana B. García, Ana Noriega, M. Rosario Campo, Elisa Varona, Fernando López, M. José Montero, María Teresa Gómez, Pilar Bartibás, Raquel Sanz, Tomasas Montes, Gerencia Área 4; Miguel Ángel Salinero; Centro de Salud Mar Baltico: Margarita Herrero, Mónica Igea, Carmen Calvo, M. Lorena Rodríguez, Carmen Pérez-Pellón, Marta Maestre; Centro de Salud Los Alpes: María Ayuso Agora, María García Martín, Dolores Velázquez, Begoña Fernández, Miriam Castro, Encarnación Ayuso, Raquel Masa, Concha Antelo, Soledad García, Susana Herrero, Ascensión Delgado. Agradecimiento general a todos el personal involucrado en estos centros sanitarios, con especial énfasis al personal administrativo de los mismos

GIRONA


LEÓN


Referencias


