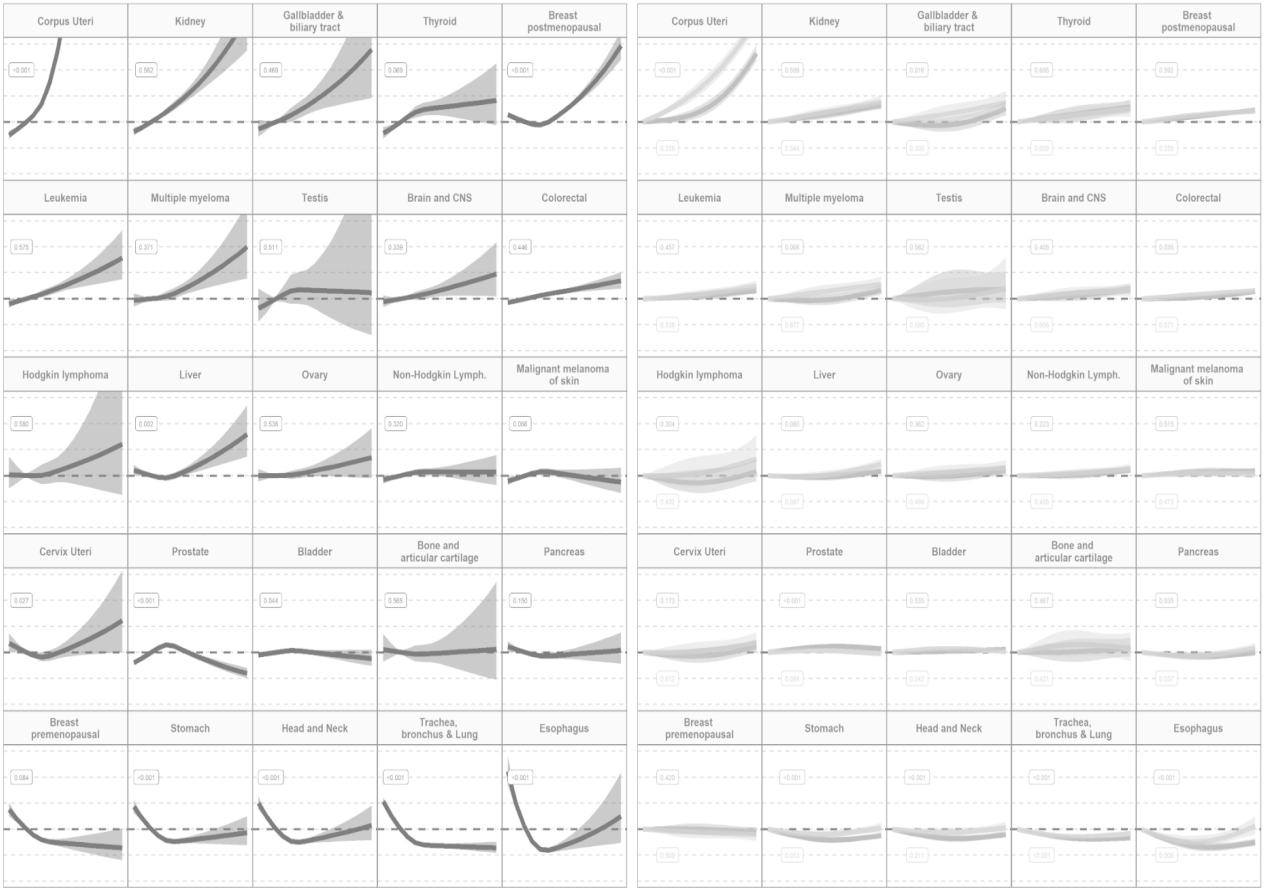
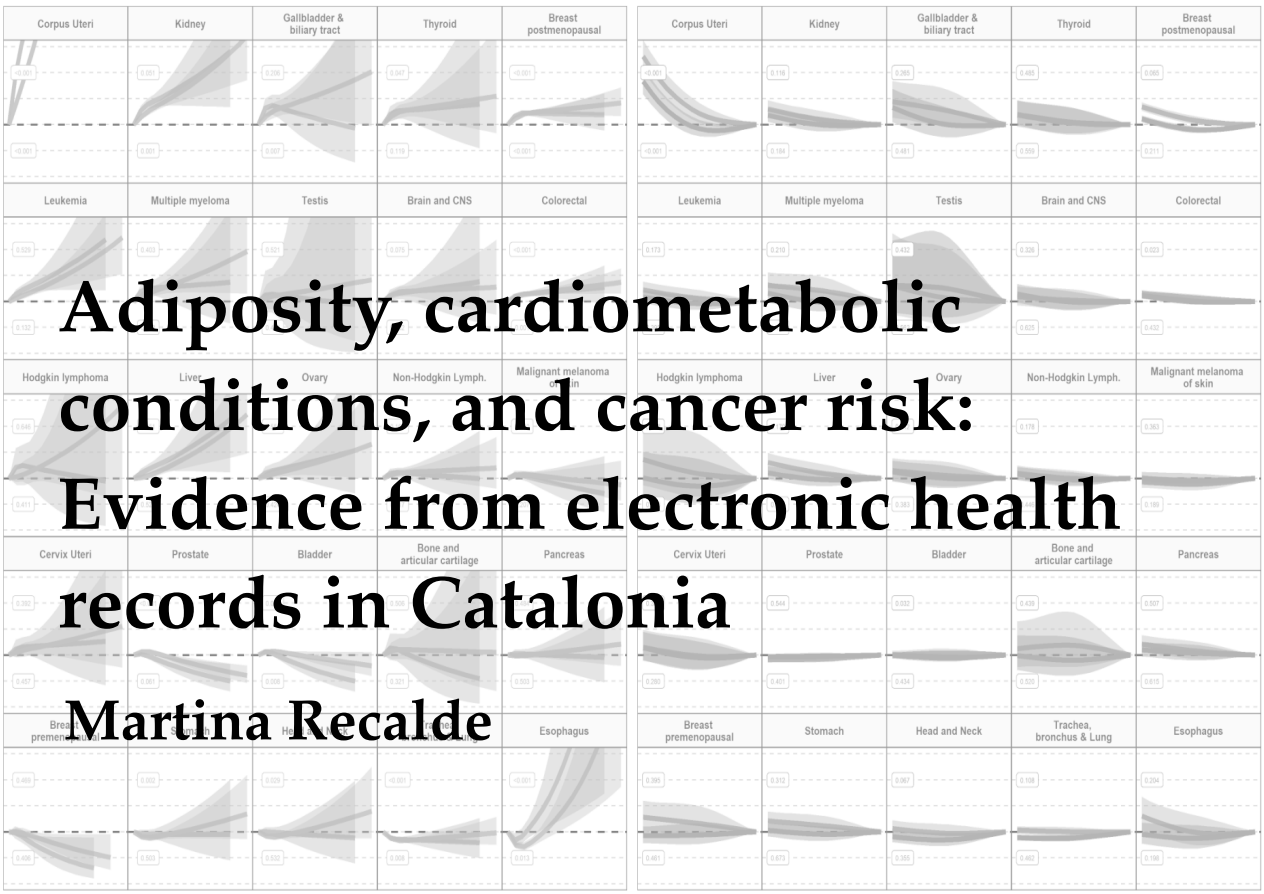
Universitat Autònoma de Barcelona

Adiposity, cardiometabolic conditions, and cancer risk: Evidence from electronic health records in Catalonia

Martina Recalde





Adiposity, cardiometabolic conditions, and cancer risk: Evidence from electronic health records in Catalonia

Martina Recalde

Doctoral program in Methodology of Biomedical Research and Public Health
Department of Paediatrics, Obstetrics & Gynaecology, and Preventive Medicine and Public Health
Universitat Autònoma de Barcelona, 2022.

Doctoral Thesis

Adiposity, cardiometabolic conditions, and cancer risk: Evidence from electronic health records in Catalonia

By Martina Recalde

Thesis supervisors:

Dr. Talita Duarte Salles
Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)

Dr. Heinz Freisling
International Agency for Research on Cancer (IARC-WHO)

Thesis tutor:

Dr. Xavier Castells
Department of Epidemiology and Evaluation, Hospital del Mar Institute for Medical Research, Barcelona, Barcelona, Spain

A mi papá

This doctoral Thesis, presented as a compendium of publications, has been conducted in the Real World Epidemiology (RWEpi) group of the *Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina* (IDIAP Jordi Gol) in Barcelona, Spain as well as in the Nutrition, Cancer, and Multimorbidity (NCM) team of the International Agency for Research on Cancer (IARC-WHO) in Lyon, France, under the co-supervision of Dr. Talita Duarte-Salles and Dr. Heinz Freisling.

This Thesis was intended to be composed of five studies. However, only three studies are included in the Results section of this Thesis (which are published in scientific journals). For the sake of completeness, the last two studies (which have been submitted to scientific journals) are included in the Appendix.

Martina Recalde was funded by Wereld Kanker Onderzoek Fonds (WKOF, grant number: 2017/1630) and World Cancer Research Fund (UK) (grant number: IIG_2019_1978), as part of the World Cancer Research Fund International grant program.

Where authors are identified as personnel of the International Agency for Research on Cancer and World Health Organization, the authors alone are responsible for the views expressed in the articles of this Thesis and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer and World Health Organization.

Barcelona, March 2022

Acknowledgments

Este camino empezó hace 4 años cuando Jeroen me comentó que su co-supervisora de Tesis estaba buscando un(a) "predoc" para un proyecto sobre obesidad y cáncer. Si bien él sabía que en principio no me interesaba la idea de hacer un doctorado (y que de cáncer no sabía nada), después de remarcar algunas de las muchas virtudes de Talita, no me pude resistir a escribirle un correo para saber más sobre ella y este proyecto. Luego de reunirme con Talita no sólo me atrapó el proyecto y las posibles estancias en el extranjero, sino que me di cuenta de que Talita era la mejor persona que podía imaginar para guiarme y acompañarme en esta aventura.

Ahora que se acerca el final de esta etapa, miro para atrás y me doy cuenta de lo mucho que crecí en estos 3 años y medio, tanto a nivel personal como profesional. Sinceramente no logro imaginarme como sería (ni qué estaría haciendo) si no hubiera emprendido este viaje. Quería aprovechar este espacio para agradecer a algunas de las personas que me ayudaron y acompañaron en esta etapa.

Talita, muchísimas gracias por haberme dado esta oportunidad, fue un placer trabajar y aprender de vos, no sólo de tus habilidades como investigadora sino también de tu liderazgo. Pensar que empezamos compartiendo el despacho y ahora liderás un grupo de 12 personas (pero sé que esto es sólo el comienzo para vos ☺)… Gracias *for speaking up* cuando estás en desacuerdo o percibís injusticias, por tu paciencia (sobre todo el primer año cuando los borradores de manuscritos o presentaciones necesitaban muchas idas y vueltas), por motivarme a alcanzar metas pero siempre dándome espacio y autonomía para gestionarme, también gracias por tu cercanía, empatía, y humildad. Gracias por entender mi situación personal (sobre todo durante la pandemia) y por haber luchado para que pudiera estar acompañada de mi familia en esos momentos difíciles. Gracias por ayudarme y empujarme a crecer, siempre haciendo lo imposible para tener la posibilidad de ir a congresos, cursos, hacer estancias, etc. Si no me asusta cerrar esta etapa y empezar la siguiente con bastante incertidumbre es, en gran parte, porque me cargaste de herramientas para poder aspirar a diferentes tipos de trabajos. Seguro me estoy dejando muchas cualidades y cosas por agradecer, pero espero que sirva al menos como una pincelada para que sepas que te estoy muy agradecida. Sin dudas, la buena corazonada que tuve sobre vos cuando te conocí superó hasta mis mejores expectativas.

Heinz, thanks a lot for everything you taught me during this time as my supervisor, especially regarding methodological issues and scientific writing. Your love for science and attention to

detail will never cease to inspire me. I also want to thank you for your relaxed and calm style of supervision. Whenever I was stressed because of timelines you always transmitted me this calmness and made me look at things from another perspective. Particularly, I think of the moment when I told you that I did not think I could finish the thesis in the timeline we had in mind and you did not have to think twice about extending my contract and told me I should not stress about things I cannot control. Thank you for your patience in reviewing manuscripts (or even paragraphs) over and over, for giving me space to work autonomously, and help me grow and gain confidence in having a leading role. Finally, I also want to thank you for understanding my personal situation and being so flexible about it in the last few months that I have been working with you at IARC.

Gracias a toda la gente del IDIAP (Constanza -companya!-, Anna B, Laura, Diana, Mariona, María García, María Aragón, Clara, Boni, Albert, Tomàs, etc) por vuestra ayuda y momentos compartidos. En especial quiero agradecer a algunos miembros de nuestro grupo. Elena, qué alegría me dio que decidieras hacer la tesis con Talita! Gracias por tus consejos y charlas (ya sean metodológicas o personales), siempre me encanta escuchar tu visión crítica y tu perspectiva de las cosas. Me encantó tener la oportunidad de escribir un paper juntas (sobre obesidad y COVID), aprendí y disfruté un montón. También me abrió los ojos sobre lo enriquecedor (y eficiente) que es trabajar en equipo (de verdad). No tengo ninguna duda que te espera un futuro brillante. Andrea, fue un placer compartir una parte de esta etapa con vos ya sea a la distancia o en Lyon. Un *special shout-out* por tu paciencia en revisar errores en mis scripts de R :). Pienso en el primer y último paper que trabajamos juntos (4 papers en total en un poco más de un año :o?) y me parece increíble todo lo que creciste como investigador. Ahora empezás vos este camino, y no me cabe ninguna duda de que será un éxito. Alicia, que bien compartir tiempo juntas en el IDIAP, y lamento que no hubiera sido más. Sobre todo te quería agradecer toda tu ayuda (y mucha paciencia!) en este último tiempo para poder navegar los procesos administrativos relacionados con el depósito de la tesis. Ed, I am so lucky to have had the opportunity to work with you over the last 2+ years. Your capacity to understand information, analyze data, write manuscripts, but more especially your modesty about it will never cease to amaze me. Jeroen, gracias por haber sido el mejor supervisor de prácticas que podía pedir. Gracias por tu ayuda y consejos a lo largo del doctorado. Muchas veces pienso en el importante papel que terminaste teniendo en mi vida, si no fuera porque me pusiste en contacto con Talita, no estaría hoy aquí. También quería agradecer a Cyntia por acogerme cuando llegué al IDIAP y por ayudarme a dar los primeros pasos en el doctorado y a Yesika

por escucharme, por tu paciencia ayudándome a descifrar errores en mis inicios en R, por tu compañía y tu amistad. Gracias Miguel-Ángel por haber sido mi tutor durante tres años. Gracias también Anna B. y Boni por vuestra empatía, sin vuestro apoyo durante la pandemia no sé si esta etapa hubiera tenido el buen cierre que está teniendo ahora. Isaura y Caroline, ha sido un placer poder co-supervisar vuestros trabajos de final de máster. Es increíble ver cómo progresaron en menos de un año. Muchas gracias por vuestra paciencia con mi inexperiencia de supervisión y por vuestra motivación de cada día (que dio y dará muy buenos frutos).

I am also thankful for the people at IARC (especially from the NME branch and NCM team) and for the ECVSs without whom going to work and being in Lyon would not be as enjoyable as it is. Thank you, Emma, for meeting with me every time I doubted the analysis of BMI, cardiometabolic conditions, and cancer risk, and for trying to figure out how to implement additive interactions in RStudio with me ☺. Gracias Laia, por ser una super compañera de oficina, por escucharme y animarme, y sobre todo por nuestras eternas caminatas y charlas. Mucha suerte en esta etapa, estoy segura de que serás una investigadora top. Gracias Vero por ayudarme a integrarme cuando llegué a IARC y por tu buen humor de siempre. Gracias Anna por toda tu ayuda con la burocracia que representa depositar y defender una tesis. Thank you, Vivian, for your tireless critical thinking, your feedback and questions always make me approach things differently. Gracias Isabelle porque sin tu hospitalidad y generosidad, venir a IARC no hubiera sido posible (o al menos mucho más complicado).

Gracias a mi familia y especialmente a mamá. La vida no te tocó fácil, pero pudiste sacarnos adelante. Sin tu ejemplo de luchadora incansable y tu priorización de nuestra educación seguro no estaría acá. Gracias también a Butty, Alain, Micky, y Nico por haberme adoptado (y nunca hacerme sentir que estaba de más en esa casa) durante mi *bachelor* en Ginebra, su generosidad no tiene límites. Merci Tante Jacqueline de m'avoir aidé lors de mon déménagement en Europe et pour tout ton soutien lors des premières années. Domi, je pense que je ne vais jamais pouvoir te remercier suffisamment tout ce que tu as fait pour moi, sans toi, je pense pas que je serai arrivée si loin. Je profite cet espace pour remercier tout ton soutien.

Sebi, gracias por tu paciencia y apoyo en todos estos años. Paciencia no solo para escuchar mis quejas, dudas, y angustias sino también para revisar mis artículos, posters, presentaciones (cuyo feedback fui aprendiendo a recibir sin sentirme tan atacada jaja). Gracias por tus sabios consejos que siempre me ayudan a ver las cosas desde otra perspectiva. Vos fuiste el primero que me dijo que me imaginaba haciendo un doctorado y que me animó a emprender este

camino. Acá estamos, 4 años más tarde, los dos a punto de ser doctores y por empezar otra nueva etapa juntos que me llena de ilusión.

“Our greatest weakness lies in giving up.
The most certain way to succeed is always to try just one more time.”— Thomas Edison

“Talent wins games, but teamwork and intelligence win championships.” — Michael Jordan

Abstract

Cancer is one of the leading causes of morbidity and mortality worldwide and in Spain. Body mass index (BMI), the most common indicator of general adiposity, has been associated with the risk of several cancer types. The use of databases of routinely collected electronic health records (EHR) has become more common in cancer epidemiology over the past decades and could be useful to fill in gaps in the adiposity-cancer literature. The main aim of this Thesis was to investigate the association between adiposity and cancer risk as well as the role of cardiometabolic conditions in this relationship. As a prior step, we aimed to evaluate the suitability of a large EHR database from Catalonia, Spain for research and, more specifically, for cancer-related research.

In this Thesis, we provide an extensive characterization of the Information System for Research in Primary Care (SIDIAP) database, we validate 25 types of incident cancer cases in the SIDIAP using regional cancer registries as the gold standard and we investigate the association between adiposity and the risk of 26 types of cancer accounting for potential non-linearity, different adiposity indicators, and individual-level factors such as smoking status and incident cardiometabolic conditions.

The findings of this Thesis revealed that SIDIAP is a suitable database to conduct health- and cancer-related research. SIDIAP includes 76% of the cancer diagnoses in the population-based cancer registries of Catalonia but includes a considerable number of cases that are not in the registries. Furthermore, adiposity is associated with an increased risk of several cancer types. We confirmed associations previously reported in studies focusing on baseline BMI and we provide novel evidence that higher and longer exposure to adiposity increases the risk of four hematological as well as head and neck and bladder (among never smokers) cancers. The BMI-cancer association is similar among individuals free of cardiometabolic conditions and those with incident hypertension and/or cardiovascular disease but it is attenuated among individuals with type 2 diabetes mellitus. BMI and waist circumference result in comparable estimates of cancer risk associated with adiposity at a population level.

The findings of this Thesis reinforce the need for public health strategies to reduce and prevent overweight and obesity. The findings also highlight the usefulness of EHRs for conducting health-related research and providing evidence for public health action.

Resumen

El cáncer es una de las principales causas de morbilidad y mortalidad en el mundo y en España. El Índice de masa corporal (IMC), el indicador más común de adiposidad general, ha sido asociado con el riesgo de varios tipos de cáncer. El uso de bases de datos de historias clínicas electrónicas (EHR, del inglés) recolectadas de forma rutinaria se ha vuelto más común en las últimas décadas en la epidemiología del cáncer y podría ser útil para colmar lagunas en la literatura de la obesidad y el cáncer. El principal objetivo de esta Tesis era investigar la asociación entre la adiposidad y el riesgo de cáncer, así como también el rol de condiciones cardiometabólicas en esta relación. Como paso previo, teníamos el objetivo de evaluar la adecuación de una gran base de datos de EHR de Cataluña, España para la investigación, y más específicamente, para la investigación relacionada con el cáncer.

En esta Tesis, proporcionamos una extensa caracterización de la base de datos del Sistema de Información para el Desarrollo de la Investigación en Atención Primaria (SIDIAP), validamos 25 tipos de casos de cánceres incidentes en el SIDIAP utilizando registros regionales de cáncer como criterio de referencia e investigamos la asociación entre la adiposidad y el riesgo de 26 tipos de cáncer teniendo en cuenta posibles asociaciones no lineales, diferentes indicadores de adiposidad, y factores a nivel individual como el hábito tabáquico y condiciones cardiometabólicas incidentes.

Los hallazgos de esta Tesis revelaron que el SIDIAP es una base de datos adecuada para realizar investigación relacionada con la salud y el cáncer. El SIDIAP incluye el 76% de los diagnósticos de cáncer de los registros de cáncer poblacionales de Cataluña, pero incluye un considerable número de casos que no figuran en los registros. Además, la adiposidad se asocia con un mayor riesgo de diversos tipos de cáncer. Confirmamos asociaciones previamente encontradas en estudios focalizados en medidas basales de IMC y proporcionamos novedosa evidencia de que una mayor y más larga exposición a la adiposidad incrementa el riesgo de cuatro cánceres hematológicos asi como del cáncer de cabeza y cuello y de véjiga [únicamente entre los nunca fumadores]). La asociación entre adiposidad y cáncer es similar en individuos sin condiciones cardiometabólicas y en aquellos con hipertensión y enfermedad cardiovascular incidentes, pero la asociación se atenúa en personas con diabetes de tipo 2. Tanto el IMC como el perímetro de cintura resultan en estimadores de riesgo de cáncer asociado con la adiposidad similares a nivel poblacional.

Los hallazgos de esta Tesis refuerzan la necesidad de estrategias de salud pública para reducir y prevenir el sobrepeso y la obesidad. Los hallazgos también resaltan la utilidad de las EHR para realizar investigación en salud y proporcionar evidencia para la acción en el ámbito de la salud pública.

Scientific Work

About the author

Martina Recalde studied Psychology (Bachelor Degree) at the University of Geneva (2013-2016) and Public Health (Master Degree) at the Pompeu Fabra University and Autonomous University of Barcelona (2016-2018). In September 2018, she joined the Real World Epidemiology (RWEpi) group led by Dr Talita Duarte-Salles at *Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina* (IDIAP Jordi Gol) in Barcelona, Spain. In October 2021, she joined the Nutrition, Cancer, and Multimorbidity (NCM) team led by Dr Heinz Freisling at the Nutrition and Metabolism (NME) branch of the International Agency for Research on Cancer (IARC-WHO) in Lyon, France. A summary of the scientific work conducted by the author during the period of her Doctoral Thesis is provided below.

Scientific articles

Published articles

Recalde, M., Rodríguez, C., Burn, E., Far, M., Manuel-García, D., Carrere-Molina, J., Benítez, M., Moleras, A., Pistillo, A., Bolíbar, B., Aragón, M., Duarte-Salles T. Data Resource Profile: the Information System for Research in Primary Care (SIDIAP). 2022. *International Journal of Epidemiology, dyac068.* https://doi.org/10.1093/ije/dyac068

Recalde M., Manzano-Salgado C.B., Díaz Y., Puente, D., Garcia-Gil, M.d.M., Marcos-Gragera, R., Ribes-Puig, J., Galceran, J., Posso, M., Macià, F., Duarte-Salles, T. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. 2019. *Clinical Epidemiology. Volume 11:1015-1024.* https://doi.org/10.2147/CLEP.S225568

Recalde, M., Davila-Batista, V., Díaz, Y., Leitzmann, M., Romieu, I., Freisling, H., Duarte-Salles, T. Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain. 2021. *BMC Medicine. 19, 10.* https://doi.org/10.1186/s12916-020-01877-3

Recalde, M., Pistillo, A., Fernandez-Bertolin, S., Roel, E., Aragon, M., Freisling, H., Prieto-Alhambra, D., Burn, E., Duarte-Salles, T. Body mass index and risk of COVID-19 diagnosis, hospitalisation, and death: a cohort study of 2 524 926 Catalans. 2021. *The Journal of Clinical Endocrinology & Metabolism.* Volume 106 (12):e5030–e5042. https://doi.org/10.1210/clinem/dgab546

Recalde, M., Roel, E., Pistillo, A., Sena, A.G., Prats-Uribe, A., Ahmed, W.U.R., [...], Duarte-Salles, T. Characteristics and outcomes of 627 044 COVID-19 patients living with and without obesity in the United States, Spain, and the United Kingdom. 2021. *International Journal of Obesity.* 45:2347–2357. https://doi.org/10.1038/s41366-021-00893-4

Recalde, M., Peralta, A., Oliveras, L., Tirado-Herrero, S., Borrell, C., Palència, L., Gotsens, M., Artazcoz, L., Marí-Dell'Olmo, M. Structural energy poverty vulnerability and excess winter mortality in the European Union: Exploring the association between structural determinants and health. 2019. *Energy Policy*. 133. https://doi.org/10.1016/j.enpol.2019.07.005

Roel, E., Pistillo, A., **Recalde, M.**, Sena, A.G., Fernández-Bertolín, S., Aragón, M., [...], Duarte-Salles, T. Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain. 2021. *Cancer Epidemiology, Biomarkers, & Prevention.* Volume 30 (10): 1884-1894. 10.1158/1055-9965.EPI-21-0266

Roel, E., Pistillo, A., **Recalde, M.**, Fernández-Bertolín, S., Aragón, M., Soerjomataram, I., Jenab, M., Puente, D., Prieto-Alhambra, D., Burn, E., Duarte-Salles, T. Cancer and the risk of coronavirus disease 2019 diagnosis, hospitalisation and death: A population-based multistate cohort study including 4 618 377 adults in Catalonia, Spain. 2022. *International Journal of Cancer.* Volume 150(5): 782- 794. https://doi.org/10.1002/ijc.33846

Reyes, C., Pistillo, A., Fernández-Bertolín, S., **Recalde, M.**, Roel, E., Puente, D., [...], Duarte-Salles, T. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. 2021. *BMJ Open.* 11:e057632. doi: 10.1136/bmjopen-2021-057632

López-Jiménez, T., Duarte-Salles T., Plana-Ripoll, O., **Recalde, M**., Xavier-Cos, F., Puente, D. Association between Metabolic Syndrome and 13 types of Cancer in Catalonia: a matched case-control study. 2022. *Plos One.* Volume 17(3): e0264634. https://doi.org/10.1371/journal.pone.0264634

Burn, E., Tebé, C., Fernandez-Bertolin, S., Aragon, M., **Recalde, M.,** Roel, E., Prats-Uribe, A., Prieto-Alhambra, D., Duarte-Salles, T.. The natural history of symptomatic COVID-19 during the first wave in Catalonia. 2021. *Nature Communications.* Volume 12, 777. https://doi.org/10.1038/s41467-021-21100-y

Prats-Uribe, A., Sena, A.G., Lai, L.Y.H., Ahmed, W., Alghoul, H., Alser, O., [...], **Recalde, M.**, [...], Prieto-Alhambra, D. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. 2021. *The BMJ.* 373 :n1038. https://doi.org/10.1136/bmj.n1038

Tan, E.H., Sena, A.G., Prats-Uribe, A., You, S.C., Ahmed, W.U.R, Kostka, K., [...] **Recalde, M.**, [...], Daniel Prieto-Alhambra. COVID-19 in patients with autoimmune diseases: characteristics and outcomes in a multinational network of cohorts across three countries. 2021. *Rheumatology.* Volume 60 (SI):SI37–SI50. https://doi.org/10.1093/rheumatology/keab250

Duarte-Salles, T., Vizcaya, D., Pistillo, A., Casajust, P., Sena, A.G., Lai, L.Y.H., [...], **Recalde, M.**, [...], Prieto-Alhambra, D. Thirty-Day Outcomes of Children and Adolescents With COVID-19: An International Experience. 2021. *Pediatrics.* Volume 148 (3): e2020042929. 10.1542/peds.2020-042929

Morales, D., Ostropolets, A., Lai, L.Y.H., Sena, A.G., Duvall, S., Suchard, M., [...], **Recalde, M.,** [...], Kostka, K. Characteristics and outcomes of COVID-19 patients with and without asthma from the United States, South Korea, and Europe. 2022. *Journal of Asthma*. 10.1080/02770903.2021.2025392

Kostka, K., Duarte-Salles, T., Prats-Uribe, A., Sena, A.G., Pistillo, A., Khalid, S., [...], **Recalde, M.,** [...], Prieto-Alhambra, D. Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using CHARYBDIS. 2022. *Clinical Epidemiology.* Volume 14: 369-384. https://doi.org/10.2147/CLEP.S323292

Under review or submitted articles

Recalde, M., Pistillo, A., Davila-Batista, V., Leitzmann, M., Romieu, I., Viallon, V., Freisling, H, Duarte-Salles, T. Longitudinal body mass index-derived exposures and risk of 26 types of cancer: a cohort study of 2.6 million adults in Catalonia, Spain. 2022. *Submitted to a scientific journal.*

Recalde, M., Pistillo, A., Fontvieille, E., Viallon, V., Freisling, H., Duarte-Salles, T. Body mass index and incident cardiometabolic conditions in relation to cancer risk: a population-based cohort study in Catalonia, Spain. 2022. *Submitted to a scientific journal. Available as preprint in:* http://dx.doi.org/10.2139/ssrn.4046665

Terre-Torras, I., **Recalde, M.**, Díaz, Y., de Bont, J., Bennett, M., Aragón, M., Cirach, M., O'Callaghan-Gordo, C., Nieuwenhuijsen, M.J., Duarte-Salles, T. Air pollution and green spaces in relation to the risk of breast cancer among pre- and postmenopausal women in Catalonia: a mega cohort. 2022. *Under Review in Environmental Research.*

Bridges, M.C., **Recalde, M.**, Pistillo, A., Aragon, M., de Bont, J., Cirach, M., Nieuwenhuijsen, M.J., Duarte-Salles, T. Availability of green space and the risk of anxiety and depression in a large population-based cohort study in Catalonia, Spain. 2022. *Under Review in Environmental Health Perspectives.*

Scientific conferences

Recalde, M., Pistillo, A., Fernandez-Bertolin, S., et al. Body mass index and risk of COVID-19 diagnosis, hospitalisation, and death: a multi-state model of 2,524,926 adults in Catalonia, Spain, 2021, *Reunión Anual de la Sociedad Española de Epidemiología (SEE)* (oral presentation)

Recalde, M., Davila-Batista, V., Díaz, Y., et al. Comparison between body mass index and waist circumference in relation to cancer risk: preliminary results from a large population-based cohort study of Spanish adults, 2020, *European and International Congress on Obesity* (poster).

Recalde, M., Davila-Batista, V., Díaz, Y., et al. General (Body Mass Index) and central (Waist Circumference) obesity in relation to cancer risk: preliminary results from a large Catalan

population-based cohort study, 2020, *Reunión Anual de la Sociedad Española de Epidemiología (SEE)* (video without defense)

Recalde, M., Burn, E., Díaz, Y., et al. Effect of bariatric surgery on cancer risk: identifying appropriate non exposed controls for a cohort study, 2020, *OHDSI European Symposium* (poster).

Recalde, M., Roel, E., Pistillo, A., et al. Characteristics and outcomes of COVID-19 obese patients: preliminary results from 122,058 patients in Catalonia, 2020, *Reunión Anual de la Sociedad Española de Epidemiología (SEE)* (video without defense)

Recalde, M., Roel, E., Pistillo, A., et al. Characteristics and outcomes of COVID-19 patients with obesity: preliminary results of an international network study, 2020, *OHDSI Symposium*, (poster)

Recalde, M., Manzano-Salgado, C.B., Díaz, Y., et al. Validation of cancer diagnoses in electronic health records in Catalonia: preliminary results from the Information System for Research in Primary Care (SIDIAP), 2019, *The International Society for Pharmacoepidemiology* (spotlight poster)

Recalde, M., Manzano-Salgado, C.B., Díaz, Y., et al. Validation of cancer diagnoses in the Information System for Research in Primary Care (SIDIAP): the importance of including cases from a hospital discharge database, 2019, *Reunión Anual de la Sociedad Española de Epidemiología (SEE)* (oral communication)

Grants and awards

Best Oral Communications prize awarded to newcomers at the Spanish Society of Epidemiology Conference, 2019.

Scholarship to attend The International Conference for Pharmacoepidemiology (ICPE) in Philadelphia, USA, 2019.

Academic training

International Society for Clinical Biostatistics (2021), Joint Models for Longitudinal and Survival Data.

33rd Residential Summer Course of the European Educational Program in Epidemiology (2021), Triangulation of genetic instrumental variables and other causal methods.

University College of London (2020), Masterclass: Multilevel multiple imputation of missing data.

John Hopkins Bloomberg School of Public Health, Fall Institute (2019), The 100 Million Brazilians Cohort, Assessing the Impact of Social Protection Policies on Health: Current Issues in Policy Analysis.

University of Bern (2019), Applied Bayesian Statistics in Medical Research.

Oxford University (2019), Real World Data Epidemiology: Oxford Summer School.

Harvard University-edX (2019), Causal Diagrams: Draw your Assumptions before Your Conclusions.

Education and outreach activities

September 2020 - June 2021: Co-supervision (with Dr. Talita Duarte-Salles) of Bridges, M.C., a student of the Master of Public Health (University of Pompeu Fabra-Autonomous University of Barcelona, Spain). During this time the student worked on her Master Thesis entitled “Risk of anxiety and depression in relation to availability of green spaces: a population-based cohort study in Catalonia, Spain". Final grade of 8.3 out of 10.

September 2019 - June 2020: Co-supervision (with Dr. Anna Berenguera) of Terre-Torras, I. a student of the Master of Public Health (University of Pompeu Fabra-Autonomous University of Barcelona, Spain). During this time the student worked on her Master Thesis entitled “Contaminación del aire, espacios verdes y riesgo de cáncer de mama en mujeres pre y posmenopáusicas de Cataluña: Una mega cohorte". Final grade of 9.8 out of 10 (with honours).

July 2019 - August 2019: Co-supervision (with Dr. Talita Duarte-Salles) of Patel, S., a Medicine student from the CUNY School of Medicine / Sophie Davis Biomedical Education, USA. The student elaborated a research protocol about a study on the association between bariatric surgery and cancer risk during her stay.

The content of the COVID-19 articles published in the International Journal of Obesity (2021) and the Journal of Clinical Endocrinology & Metabolism (2021) was mentioned in several newspapers and other online resources (El Mundo, La Vanguardia, El Español, El Día, El Punt Avui, Regió 7 [link to 2nd article], elDiario.es, Diari de Girona, Diari més, Diari Segre, Estrella Digital, El Correo de Andalucía, Social.cat, CLM24, Noticia expreso, Alnavío, IM Médico Hospitalario and EFE España). MR also gave interviews in three radio (Cadena Ser, Cadena COPE, Catalunya Radio) and a television channel (TV3).

Abbreviations

BMI: Body Mass Index;
CI: Confidence Interval;
CMBD: Minimum Basic Dataset;
CNS: Central Nervous System;
CVD: Cardiovascular Disease;
DNA: Deoxyribonucleic acid;
EHR: Electronic Health Record;
GP: General Practitioner;
HR: Hazard Ratio;
HTN: Hypertension;
IARC: International Agency for Research on Cancer;
ICCC: International Classification for Childhood Cancer;
ICD: International Classification of Diseases (O: Oncology);
IGF: Insulin-like Growth Factors;
MedDRA: Medical Dictionary for Regulatory Activities;
MEDEA; *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales*;
OHDSI: Observational Health Data Sciences and Informatics;
PPV: Positive Predictive Value;
RERI: Relative excess risk due to interaction;
SD: Standard Deviation;
SES: Socioeconomic Status;
SIDIAP: Information System for Research in Primary Care;
SNOMED: Systematized Nomenclature Of Medicine;
SOC: System Organ Class;
T2DM: Type 2 Diabetes Mellitus;
UK: United Kingdom;
US: United States;
WC: Waist Circumference;
WHO: World Health Organization

Table of Contents

Acknowledgments **I**

Abstract **VII**

Resumen **VIII**

Scientific Work **X**

Abbreviations **XV**

1. Introduction **22**

1.1. Cancer **22**

1.1.1. Definition and descriptive epidemiology 22

1.1.2. International Classifications 24

1.1.3. Risk factors 25

1.2. General adiposity **26**

1.2.1. Definition and descriptive epidemiology 26

1.2.2. Measurement of adiposity in epidemiological studies 27

1.2.3. Risk factors 29

1.3. Adiposity and cancer risk association **31**

1.3.1. State-of-the-art 31

1.3.2. Possible mechanisms 35

1.4. Cardiometabolic conditions as modifiers of the adiposity-cancer association **37**

1.4.1. Hypertension 37

1.4.2. Type 2 diabetes mellitus 38

1.4.3. Cardiovascular diseases 39

1.4.4. Possible mechanisms 40

2. Rationale **43**

3. Objectives **45**

4. Methods **47**

4.1. Study designs, setting, and data sources **47**

4.2. Study population **47**

4.3. Variables **48**

4.3.1. Exposures 48

4.3.2. Cancer definition 49

4.3.3. Outcomes 50

4.4. Statistical analyses **50**

5. Results **54**

5.1. Study I **55**

5.2. Study II **69**

5.3. Study III **80**

6. Discussion **96**

6.1. Main findings **96**

6.1.1. The SIDIAP database for epidemiological research 96

6.1.2. Adiposity and cancer risk association 97

6.2. Strengths and limitations **99**

6.2.1. Data source 99

6.2.2. Exposure assessment 100

6.2.3. Outcome assessment 101

6.2.4. Covariate assessment 102

6.3. Contributions to the current knowledge **102**

6.3.1. EHR databases for epidemiological research 102

6.3.2. Adiposity and cancer risk association 104

6.4. Implications of the findings for public health **108**

6.5. Recommendations for future research **109**

6.5.1. Data sources 109

6.5.2. Exposure assessment 110

6.5.3. Outcome assessment 111

6.5.4. Covariate assessment 111

6.5.5. Biological mechanisms 112

7. Conclusions 114

8. References 116

9. Appendices 140

9.1. Appendix 1: Study IV 141

9.2. Appendix 2: Study V 173

Table of Contents: Figures

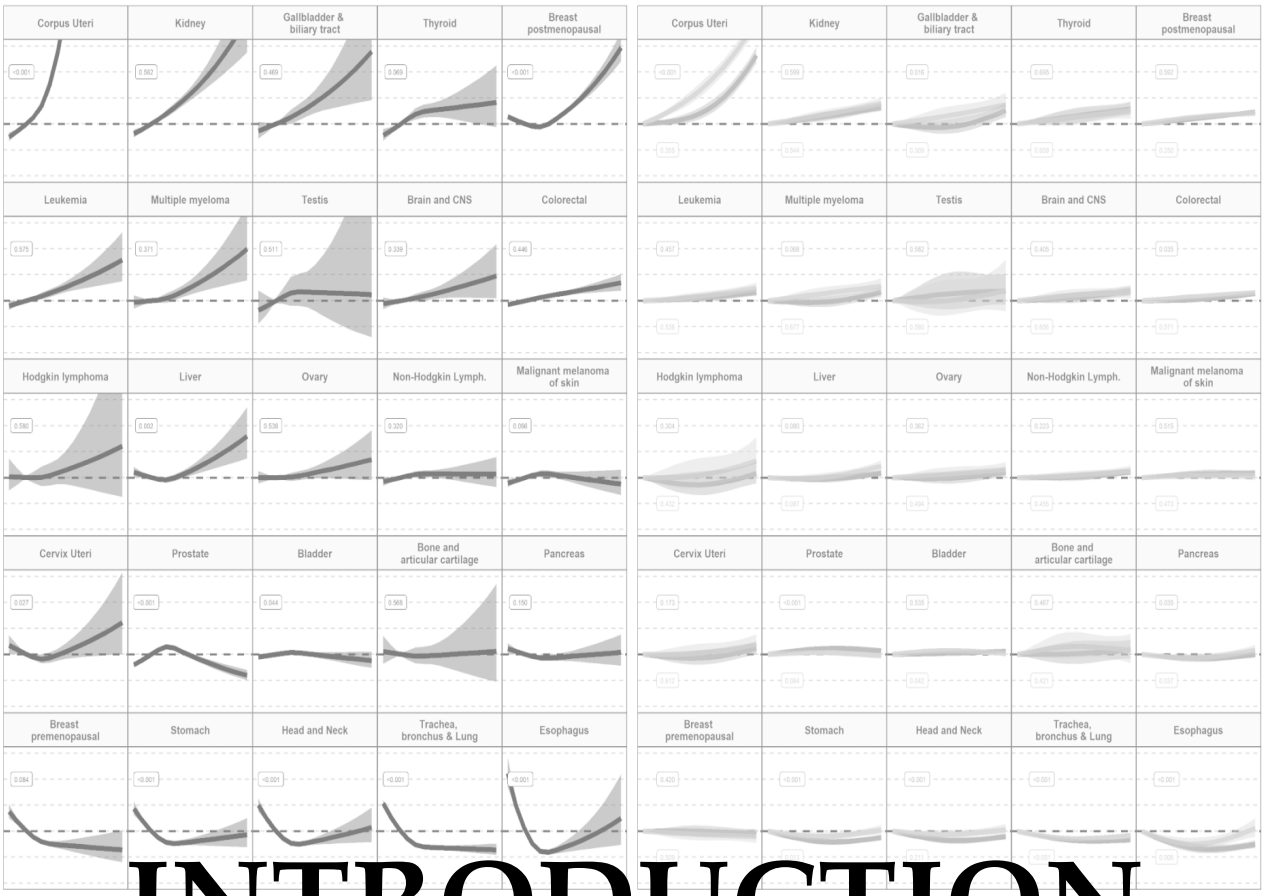
Figure 1. Estimated age-standardized incidence rates of all cancer types combined worldwide in 2020.

Figure 2. Prevalence of obesity among adults (aged 18 years or older) for both sexes worldwide in 2014.

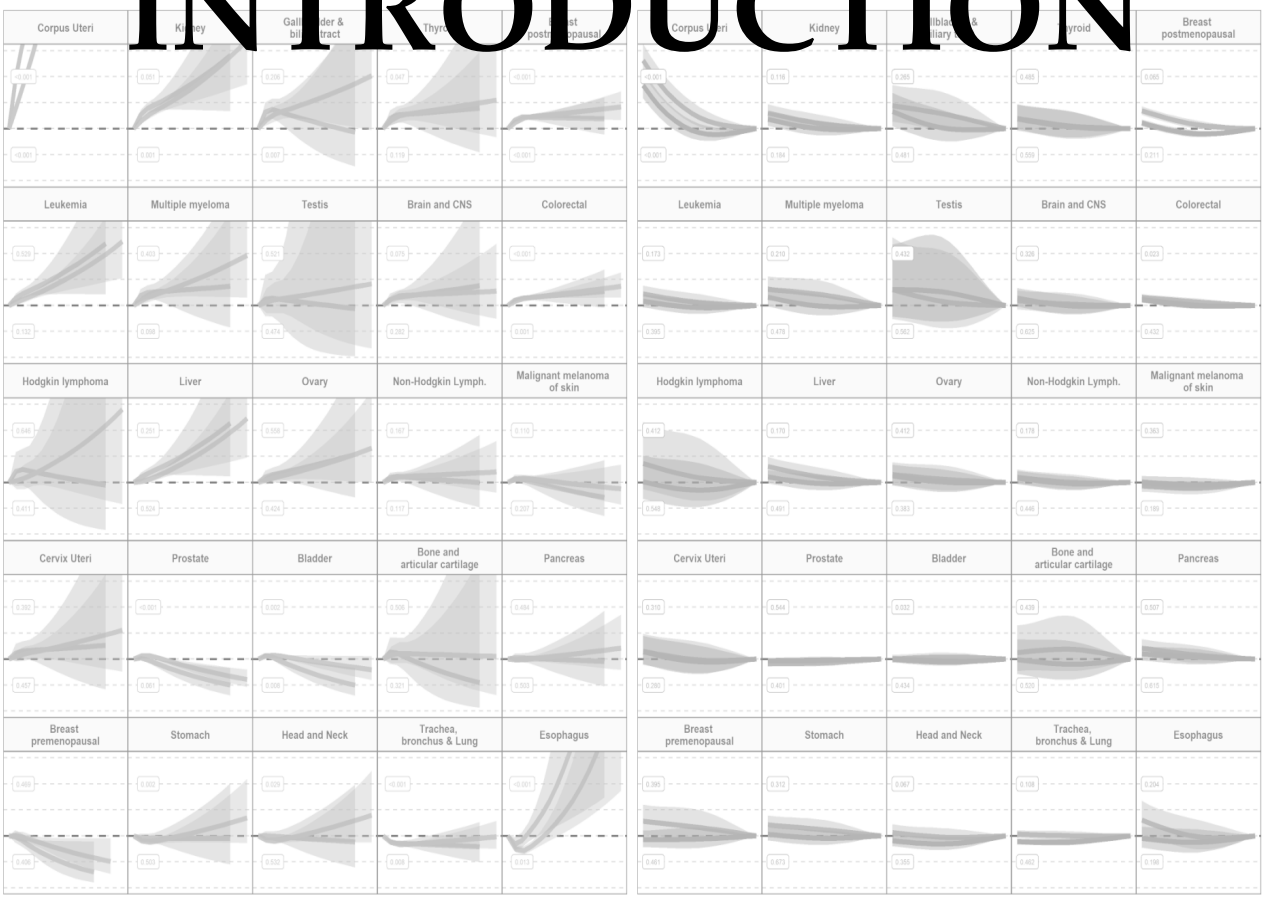
Figure 3. A framework to categorize obesity determinants and solutions.

Figure 4. Summary of the evidence relating adiposity to the risk of specific cancer types (in linear models) in landmark reviews and large cohort studies.

Figure 5. Summary of the findings of this Thesis relating different adiposity exposures to the risk of specific cancer types.



INTRODUCTION



1. Introduction

1.1. Cancer

1.1.1. Definition and descriptive epidemiology

Cancer encompasses a large group of diseases that are characterized by an uncontrollable and abnormal growth of cells.(1) Cancer (or malignant neoplasm) can start at almost every organ or tissue of the body and occurs when these uncontrollably and abnormally grown cells transcend their usual limits to invade adjoining parts of the body and/or spread to other organs.(1)

There are numerous types of cancer such as carcinoma, sarcoma, melanoma, leukemia, lymphoma and multiple myeloma, and central nervous system (CNS) cancers.(2) Carcinoma is the most common type of cancer and begins in the skin or in tissues that line or cover internal organs. Sarcoma starts in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissue. Melanoma arises in the cells that make the pigment in the skin. Leukemia begins in blood-forming tissue, such as the bone marrow, and causes excess creation of abnormal blood cells. Lymphoma and multiple myeloma originate in the cells of the immune system. Finally, CNS cancers initiate in the tissues of the brain and spinal cord.

Cancer is one of the leading causes of morbidity and mortality worldwide.(3) In 2020, there were 19.3 million new cancer cases and 10.0 million cancer deaths.(4) As shown in Figure 1, the regions with the highest age-standardized incidence rates of cancer in 2020 were North America, Europe, and Oceania while the lowest ones were Africa, the Middle East, and South Asia.(5) In Spain, cancer was responsible for more than 277,000 new cancer cases in 2020 and approximately 110,000 deaths in 2019.(6,7) Both at the global and Spanish levels, cancer is the second most frequent overall cause of death.(7,8)

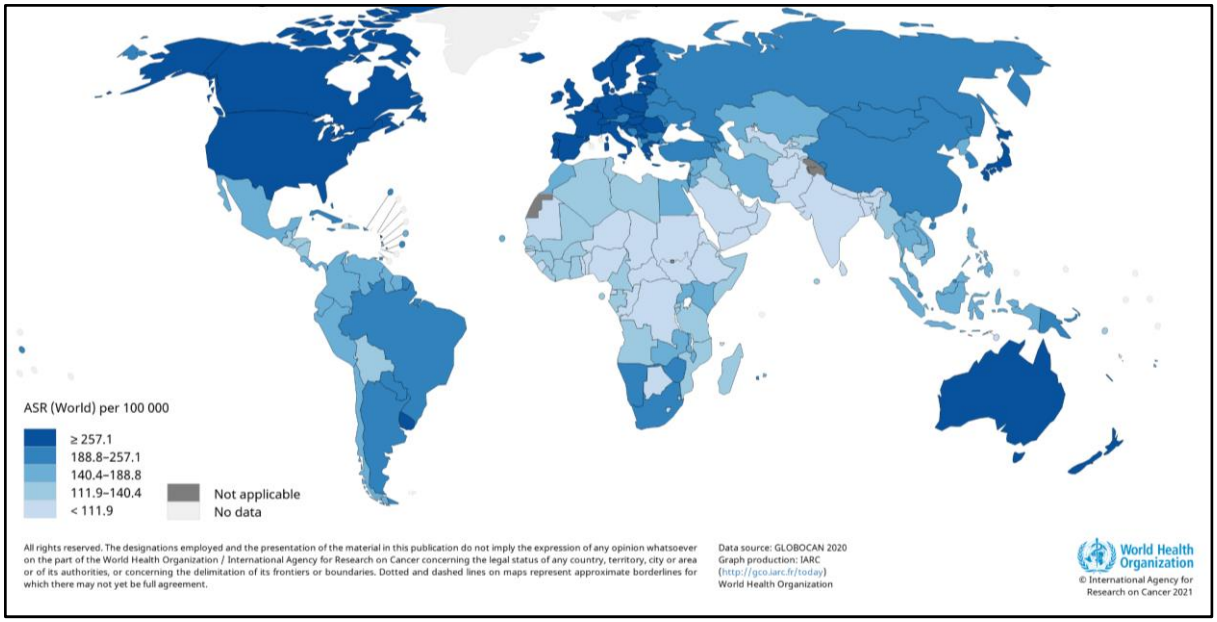


Figure 1. Estimated age-standardized incidence rates of all cancer types combined worldwide in 2020.

Notes: The source is the International Agency for Research on Cancer - World Health Organization (IARC-WHO). GLOBOCAN 2020. 2020 [cited 2021 Oct 04]. Available from: https://gco.iarc.fr/today/home.
Abbreviations: ASR: age-standardized rate.

Among males in 2020, the most common cancer types worldwide included lung (age-standardized incidence of 32 per 100 000), prostate (31 per 100 000), colorectum (23 per 100 000), stomach (16 per 100 000), and liver (14 per 100 000) cancers.(9) In Spain, these were prostate (age-standardized incidence of 71 per 100 000), colorectum (48 per 100 000), lung (incidence of 44 per 100 000), bladder (27 per 100 000), and kidney (13 per 100 000).(10) The most notable differences in the age-standardized incidence rates between Spain and the world were that the incidence rates of prostate, colorectal, and bladder cancers were more than twice as high in Spain, while for stomach and esophagus cancers, the opposite was observed.(9,10) Among females in 2020, breast cancer was the most frequent cancer worldwide (incidence of 48 per 100 000), followed by colorectum (16 per 100 000), lung (15 per 100 000), cervix uteri (13 per 100 000), and thyroid (10 per 100 000).(11) In Spain, the most frequent cancers were breast cancer (incidence of 78 per 100 000), followed by colorectum (25 per 100 000), lung (15 per 100 000), corpus uteri (13 per 100 000), and thyroid (12 per 100 000) cancers.(12) In Spain, the age-standardized incidence rates were considerably higher for breast and colorectal cancers compared to the world ones, while the contrary was seen for cervix and stomach cancers.(11,12)

1.1.2. International Classifications

There are different systems for classifying diseases that can be used to classify cancers such as the International Classification of Diseases (ICD), ICD for Oncology (ICD-O), Systematized Nomenclature Of Medicine (SNOMED), the International Classification for Childhood Cancer (ICCC), or (until 2018) Read codes in the United Kingdom (UK).(13–17) Cancer cases can also be captured in certain databases under vocabularies serving other purposes (eg, Medical Dictionary for Regulatory Activities [MedDRA] in the System Organ Class [SOC]).(18) However, the ICD remains the principal system for classifying diseases worldwide and ICD-O is the principal source for coding neoplasms in cancer registries.

In 1948, the 6th edition of the ICD incorporated for the first time a nomenclature and coding of neoplasms.(13,14) Since then, the ICD series has mainly focused on the topography (eg, lung or breast) and behavior (ie, malignant, benign, in situ, or not specified) of neoplasms.(14) The ICD-O is a manual that also includes information on the morphology of neoplasms.(14) While ICD-O is the principal coding system to categorize neoplasms in cancer registries, the ICD is still widely used in other settings such as primary care practices (which is the main focus of this Thesis).

The ICD-10th edition includes information on the topography and the behavior (contained in the same topographical code) of neoplasms.(19) There are five broad categories for the behavior of the neoplasm: benign neoplasms (D10-D36), neoplasms of uncertain and unknown behavior (D37-D48), in situ neoplasms (D00-D09), malignant neoplasms stated or presumed to be primary (C00-C76, C80-C97), and malignant neoplasms, stated or presumed to be secondary (C77-C79). Thus, primary incident cancers are coded using the malignant neoplasms stated or presumed to be primary (C00-C76, C80-C97) category which is subdivided according to the topography of the neoplasm: malignant neoplasms of the lip, oral cavity and pharynx (C00-C14), digestive organs (C15-C26), respiratory and intrathoracic organs (C30-C39), bone and articular cartilage (C40-C41), melanoma and other malignant neoplasms of skin (C43-C44), mesothelial and soft tissue (C45-C49), breast (C50), female genital organs (C51-C58), male genital organs (C60-C63), urinary tract (C64-C68), eye, brain and other parts of the CNS (C69-C72), thyroid and other endocrine glands (C73-C75), without specification of site (C80), lymphoid, hematopoietic and related tissue (C81-C96), and independent (primary) multiple sites (C97).

1.1.3. Risk factors

Cancer is a disease mainly caused by abnormal changes in the genes of the cells affecting their functions, also known as mutations.(20,21) Certain mutations can be inherited; however, most genetic changes occurring in the cells are acquired.(21)

Inherited gene mutations are linked to specific mutations inherited from an individual's parents (germ-line mutations) and therefore are present in every cell in the body that has a nucleus.(21,22) It has been estimated that inherited genetic mutations play a major role in 5% to 10% of all cancers.(21)

Acquired genetic mutations can occur during a lifespan due to random mistakes during cell replication or from unrepaired deoxyribonucleic acid (DNA) damage. Mutations are more likely to happen as people get older, and some lifestyle and environmental exposures can also increase the risk of mutations.(22,23) The exposures that have been associated with the highest proportion of cancer cases include smoking, overweight and obesity, ultraviolet radiation, occupational exposures, and infections.(24) However, other factors increasing the risk of mutations include high alcohol intake, insufficient fiber intake, exposure to ionizing radiation, intake of processed meat, air pollution, insuficient physical activity, high levels of female sex hormones (eg, estrogens or progesterone) (22–24) A better understanding of the role of modifiable lifestyle factors such as obesity (which is the main focus of this Thesis) in cancer risk is essential for the implementation of preventive strategies of cancer at the individual and population level.

KEY MESSAGES

Cancer

- Cancer is one of the **leading causes of morbidity** and mortality worldwide and in Spain.
- Cancer is a disease caused by abnormal changes in the genes of the cells affecting their functions (ie, mutations). **Acquired genetic mutations** can occur during a life span. **Exposure** to certain **lifestyle factors** can increase the risk of mutations.

- A better understanding of the **role of** modifiable lifestyle factors (eg, **obesity)** in cancer risk is **essential** for the **implementation of preventive strategies** of cancer at the individual and population levels.

1.2. General adiposity

1.2.1. Definition and descriptive epidemiology

Adiposity is defined as abnormal or excessive fat accumulation that presents a health risk.(25) Obesity (or general adiposity) has been associated with the risk of non-communicable diseases, such as cardiovascular diseases, type 2 diabetes mellitus (T2DM), and some cancers.(26) The worldwide prevalence of obesity has doubled between 1980 and 2015, and projections show that the prevalence will continue to rise in the next years.(27,28) In 2015 it was estimated that 15% and 11% of the female and male worldwide population, respectively, were living with obesity (which accounts for more than 600 million adults).(25,27) In Spain, it has been estimated that obesity affects between 17% and 28% (depending on the source) of the adult population, and higher levels among males have been reported.(29–34) There are also substantial socioeconomic, gender, and racial disparities in obesity.(35–42) At a global scale, the prevalence of obesity is higher in high- and middle-income countries; although over the past decades, obesity has also become a large problem in low-income countries.(35,36) As shown in Figure 2, the Americas, Europe and Russia, the North of Africa, the Middle East, and Oceania had prevalences of obesity above 20% in 2014, while most of the African continent and South Asia had prevalences below 10%. At a national or urban scale, in high- and middle-income countries, low socioeconomic status (SES) groups are more likely to have obesity compared to those from higher-SES, whereas in low-income countries, high-SES groups are more likely to live with obesity.(37–40) In high-income countries, obesity tends to affect more men than women, while in low- and middle-income countries, the opposite is observed.(41) There are also important racial disparities in obesity, for example, in the United States (US) obesity rates are higher among non-Hispanic Black, Hispanic, and Mexican American adults compared to non-Hispanic White adults.(42) The high prevalence of obesity, its health disparities along its concomitant health risks make obesity a major global health challenge.

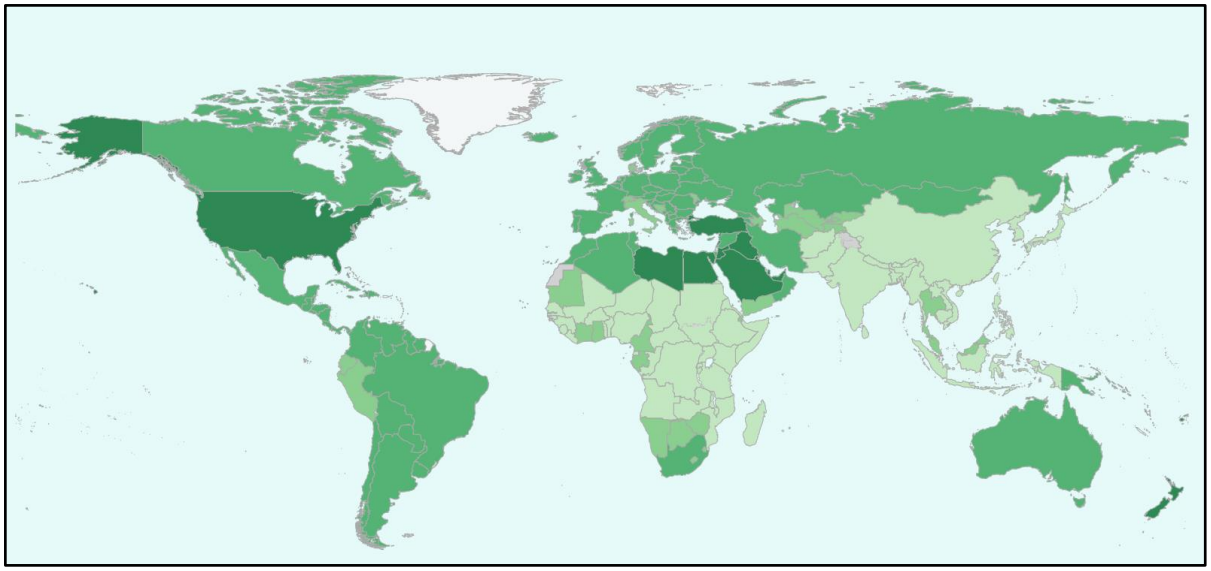


Figure 2. Prevalence of obesity among adults (aged 18 years or older) for both sexes worldwide in 2014.

Notes: The source is the World Health Organization. Obesity atlas. 2017 [cited 2021 Oct 06]. Available from: http://gamapserver.who.int/gho/interactive_charts/ncd/risk_factors/obesity/atlas.html. There are six colors in this map, very dark green stands for a prevalence of obesity ≥30.0%, medium-dark green for one of 20.0%-29.9%, medium-light green for one of 10.0%-19.9%, light green for one of <10.0%, gray means not applicable and white that there is no data.

1.2.2. Measurement of adiposity in epidemiological studies

Body composition can be assessed using different methods which can be largely categorized into reference (also known as direct and criterion) and field (or indirect) methods.(37,43,44) Reference methods include underwater weighing, air-displacement plethysmography, dilution method, whole-body potassium counting, dual-energy x-ray absorptiometry, computed tomography, and magnetic resonance imaging technologies. While these methods are characterized by their accuracy and reproducibility, they have important limitations such as being timely, expensive, hard to transport, and technically complex.(37,43,44) Therefore, in practice, reference methods are used in small-scale studies or as validation tools for field methods. Field methods include bioelectrical impedance analysis and anthropometric measurements such as body mass index (BMI), waist circumference (WC), waist-to-hip ratio, and skinfold thickness. These methods can be implemented in large clinical or epidemiological studies given that they are easy to measure, relatively cheap, and validated against reference methods.(37,44,45) BMI and WC were the two measurements of adiposity used in the analyses of this Thesis.

BMI (weight/height^2) is the most commonly used adiposity index in epidemiological studies.(37) BMI was introduced in 1835 by Lambert Adolphe Jacques Quetelet, a Flemish astronomer, and statistician.(46) Quetelet discovered that the relationship between body mass and height in normal young adults was least affected by height when the ratio of weight to height was squared.(47) Squaring the height reduces the effect of the variance in height in the relationship of weight to height (before, the ratio of the weight to height was used, which overestimated body mass in taller individuals). But it was not until 1972 that Keys et al. popularized the Quetelet index under the name of BMI.(47) In 1995, the World Health Organization (WHO) published a classification of body weight for height, based on the BMI, and BMI became widely adopted.(46,48) BMI was categorized into underweight (BMI < 18.5 kg/m^2), normal weight (BMI between ≥ 18.5 and < 25 kg/m^2), overweight (BMI ≥ 25 and < 30 kg/m^2), and obesity (BMI ≥ 30 kg/m^2). The use of BMI for epidemiological studies has several advantages. This indicator is widely used in clinical practice, inexpensive, easy to calculate and interpret given its standardized categorization, highly correlated with body fat (assessed with reference methods), and strongly related to multiple adverse health outcomes.(26,37,43,44,49–51) However, this indicator also has limitations. BMI cannot distinguish between body fatness and lean body mass and there are differences in the accuracy to capture body fatness by sex, age, and ethnicity (for the same BMI, different levels of body fat and lean body mass are observed, depending on these demographics).(37,52) At an equal BMI, the percentage of body fat is higher in women than in men, higher in older individuals compared to younger ones (loss of lean body mass and gain of fat mass increases with age), higher in Asians and lower in Blacks compared to Whites.(49,53,54)

WC is an indicator of central obesity which is measured at the umbilical level, midway between the anterior superior iliac spine and the inferior border of the rib while participants are standing.(37,43,44) The main strengths of this indicator are its simplicity, inexpensiveness, high correlation with reference methods, standardization of cut-offs, and association with cardiovascular disease (CVD) and mortality.(37,43,44) Cut-points of WC have been established based on their correspondence to a BMI of ≥ 25 kg/m^2 (80cm for women and 94cm for men) and ≥ 30 kg/m^2: (88cm for women and 102cm for men).(55) WC also has its limitations; it is an imperfect indicator of intra-abdominal adipose tissue (it includes both subcutaneous fat deposition and visceral adipose tissue), it is not fully adopted in clinical practice, the exact point of measurement varies between centers or measurers, cut-offs vary

according to age and ethnicity, and measuring WC in individuals with very high BMIs is difficult.(37,43,44,56–59)

1.2.3. Risk factors

Obesity is caused by an energy imbalance between calories consumed and calories expended.(25) However, more broadly, the causes of obesity are complex and multifaceted, which different frameworks of the determinants of obesity have attempted to summarize.(60–63) Although each framework contributes with interesting insights on the matter, we adhere to the framework of the determinants of obesity, published by Swinburn et al. in 2011.(60) This framework states that there are distal and proximal determinants of obesity (Figure 3).

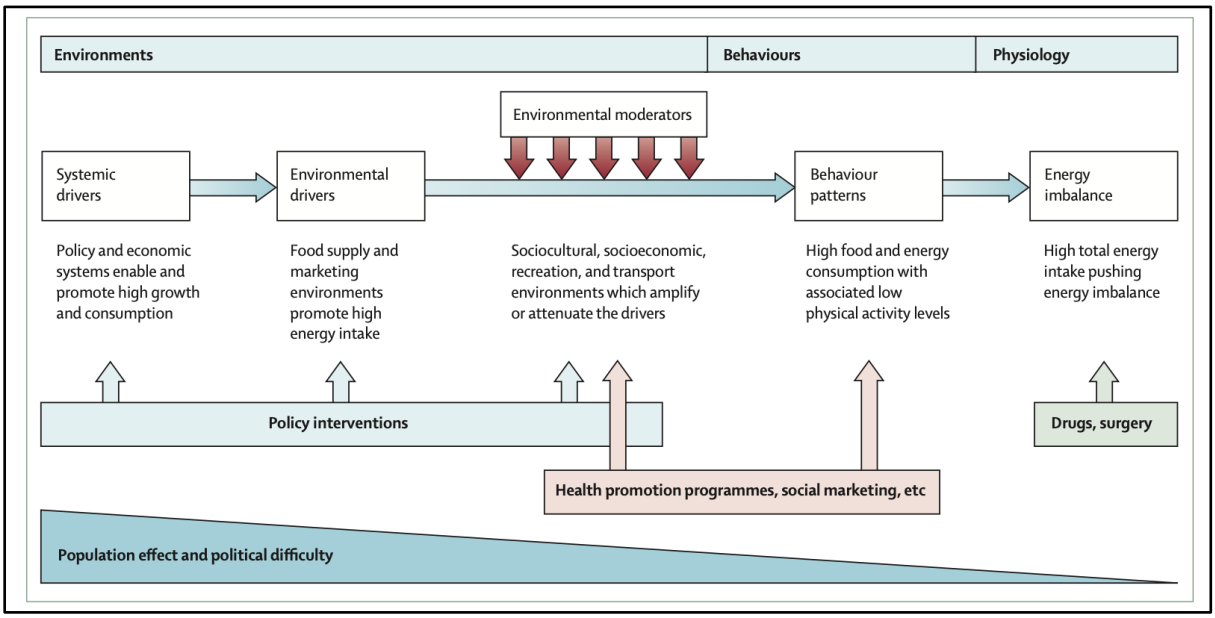


Figure 3. A framework to categorize obesity determinants and solutions.

Notes: The source is Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, Gortmaker SL. The global obesity pandemic: shaped by global drivers and local environments. Lancet. 2011 Aug 27;378(9793):804-14. doi: 10.1016/S0140-6736(11)60813-1. PMID: 21872749.

At a distal level, the changes in the global food system (ie, increased production of processed, palatable, energy-dense, affordable, and effectively marketed food) seem to be the main drivers of the worldwide growth of obesity in the last decades.(60,64) For these changes to affect a country, countries must also have sufficient wealth as a precondition to developing obesity.(60) Furthermore, policy and economic systems can be considered as systematic drivers of obesity, these systems enable and promote high growth and consumption (eg, environmental drivers can be reinforced through laws and regulations).(60) The differences in the prevalence of

obesity at the national or local level can be explained by environmental moderators which might amplify or attenuate the changes in the global food system. These moderators can be sociocultural (eg, large body size preference or valuation of energy-dense food/sedentary activities), socioeconomic (eg, low proportion of manual occupations or a high proportion of car owners), recreational (eg, preference for passive leisures or lack of spaces favoring active recreation such as parks and sporting facilities), and transport environments (eg, prioritization for car transportation or lack of availability of cycle/footpaths, public transport, and accessible stairs in buildings).(60,62,63,65–67) The environmental drivers and moderators (amplifiers of the drivers) of obesity described in this framework have been labeled with the term "obesogenic environment" elsewhere.(68) This term, coined in the 1990s, was proposed to refer to the role that environmental factors may play in determining both energy intake and expenditure.(62,68) More precisely, an obesogenic environment is the "sum of the influences that the surroundings, opportunities or conditions of life have on promoting obesity in individuals and populations".(68)

The proximal determinants of obesity can explain why individuals have different susceptibility to obesogenic environments. While this framework highlights the role of behavioral patterns (high energy consumption and low levels of physical activity) as proximal determinants of obesity, there are other important factors. These include socio-demographics such as age (ie, individuals tend to gain weight as they age), gender (eg, more women are affected by obesity in low- and middle-income countries), individual SES (robust evidence links SES with obesity, although there is a differential effect of SES according to the countries' wealth), race/ethnicity (due to differential SES or social norms regarding eating and physical activity), and nativity (eg, migrants tend to adopt the obesity prevalence and/or lifestyle practices of the country of immigration over time).(37,69–71) But also other relevant variables, for example, biological factors (eg, several genes have been associated with obesity in genome-wide association studies), prenatal exposures (eg, maternal smoking during pregnancy, gestational weight gain, or gestational diabetes), post-natal/early-life exposures (eg, breastfeeding and its duration, early adiposity rebounds during childhood, childhood overweight, and lack of sleep), sleep deprivation (some potential mechanisms include altered thermoregulation, increased feeling of fatigue which can lead to reduced physical activity and increased caloric intake), and psychological factors (eg, habits, beliefs, stress, lack of social support, and depression).(37,62,72–89)

KEY MESSAGES

Adiposity

- The **prevalence** of obesity worldwide has more than **doubled** over the past three decades, reaching **>600 million adults** in 2016. In Spain, it has been estimated that obesity affects between 17% and 28% (depending on the source) of adults.
- In epidemiological studies, **general adiposity** is commonly assessed using the **body mass index** while **central adiposity** is measured using the **waist circumference** indicator.
- Obesity is a complex public health problem, thought to be caused by **distal** (eg, changes in the global food system) and **proximal** (eg, age, gender, SES, race) **determinants.**

1.3. Adiposity and cancer risk association

1.3.1. State-of-the-art

BMI has been convincingly associated with risk of cancers of the esophagus (adenocarcinoma), gastric cardia, colon and rectum, liver, gallbladder, pancreas, breast postmenopausal, corpus uteri, ovary, kidney (renal cell), meningioma, thyroid, and multiple myeloma in an analysis of more than 100 epidemiologic studies by the International Agency for Research on Cancer (IARC) Viewpoint Working Group (Figure 4).(90) For other types of cancers such as non-Hodgkin lymphoma, leukemia, lung, bladder, or testis, the evidence is currently limited or considered inadequate.(90) Data are also limited for rare cancers (incidence <6/100,000/year), including Hodgkin-lymphoma (incidence in Spain in 2020: 2.6/100.000 men and women), larynx (3.0/100.000 men and women), or cervix uteri (5.4/100.000 women), given that less-frequent cancers are difficult to study in traditional cohort studies due to their low incidence. Conducting research on rare cancers is important as these cancers combined contribute to more than 20% of all cancers diagnosed annually in Europe and as five-year relative survival is worse for individuals diagnosed with rare cancers compared to common cancers (47% vs 65%).(91,92)

To this date, large reviews and meta-analyses have been essential to synthesize the data reported in numerous small studies investigating the association between BMI and risk of cancer.(22,90,93) Despite this wealth of knowledge, these reviews and meta-analyses have included heterogeneous studies with data recorded in different settings, analyzed with diverse methodologies, missing important covariates in their models, and on self-reported BMIs. More importantly, data on rare cancers are limited in these landmark publications given that these cancers have seldom been analyzed in smaller studies.

Large cohort studies (>1 million participants) using prospectively collected data have emerged in the last years as a great opportunity to study the association between BMI and the risk of different cancer types (including less-frequently occurring ones) using systematic methodologies.(94–105) Engeland et al. and Børge et al. included data on 2 million men and women from the general population of Norway and analyzed 15 cancer types across 10 different publications using a similar methodology.(94–103) They found positive linear associations between BMI and risk of 12 cancer sites [esophageal adenocarcinoma, prostate, renal cell carcinoma, colorectal, gallbladder, small intestine, thyroid (only in women), non-Hodgkin lymphoma (only men), Hodgkin lymphoma (only women), leukemia (only for acute lymphatic type and men), plasma cells, and corpus uteri], and a negative association with esophageal squamous cell carcinoma (Figure 4). However, in five of these studies, they investigated potential non-linear associations and found that the association between BMI and renal cell carcinoma, rectum (men), and gallbladder (men) might be non-linear (although the confidence intervals [CIs] were very wide).(97,98,100,101,103) Reeves et al. conducted a study in the UK with 1.2 million women recruited into the Million Women Study and examined 17 cancer types.(104) They found positive associations between BMI and the risk ok the following cancers: esophageal adenocarcinoma, colorectal (only in premenopausal women), pancreatic, breast cancer in postmenopausal women, corpus uteri, ovarian, kidney, non-Hodgkin's lymphoma, multiple myeloma, and leukemia; and negative associations with oesophageal squamous cell carcinoma and lung (Figure 4). Finally, Bhaskaran et al. gathered data from primary care electronic health records (EHRs) of 5.25 million individuals living in the UK.(105) Their findings revealed associations between BMI and the risk of 17 out of 22 analyzed cancers (Figure 4 only summarizes associations from linear models). They reported non-linear associations for the risk of 10 cancers (oral cavity, esophagus, stomach, colon, liver, lung, malignant melanoma, breast premenopausal, beast postmenopausal, and prostate) and

positive linear associations with the risk of cancers of the gallbladder, cervix, corpus uteri, ovaries, kidney, thyroid, and leukemia.

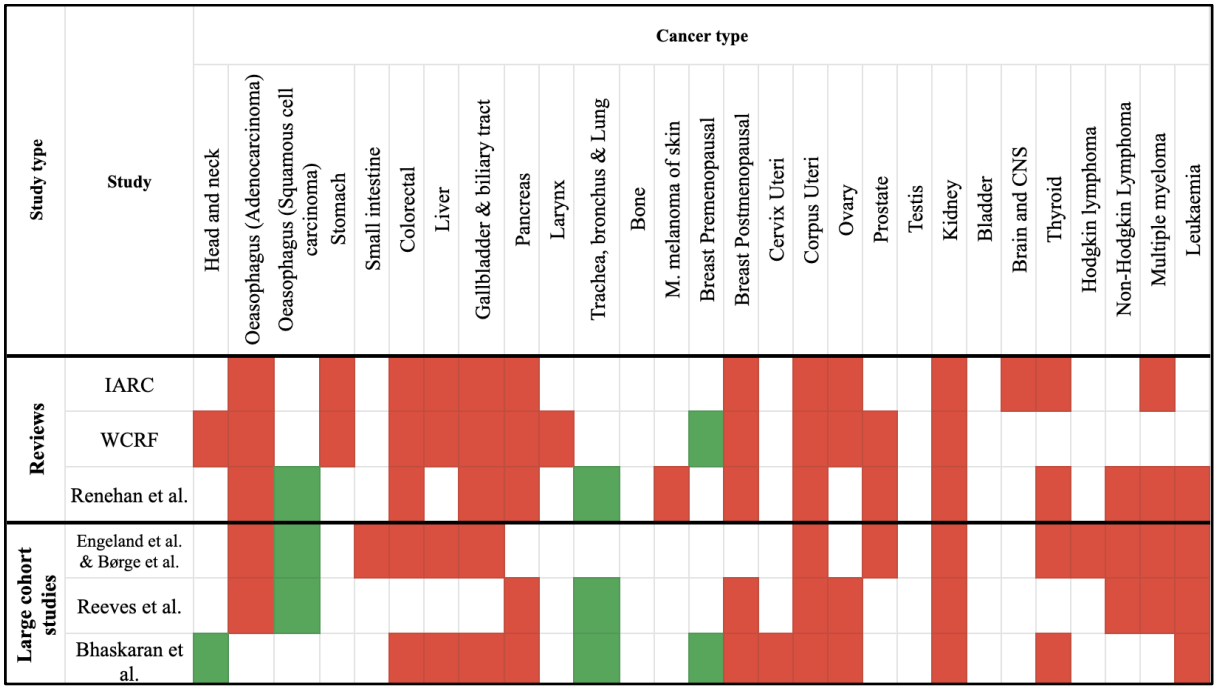


Figure 4. Summary of the evidence relating adiposity to the risk of specific cancer types (in linear models) in landmark reviews and large cohort studies.

Notes: Own elaboration with data from 15 publications.(22,90,93–105). Red boxes represent positive associations and green boxes negative associations between BMI and cancer risk. In the IARC Viewpoint Working Group study "Gastric cardia" was considered instead of "stomach", "kidney: renal cell" instead of "kidney", "meningioma" instead of "brain and CNS". In the WCRF report "Mouth, pharynx, larynx" was considered instead of "head and neck" and “larynx”. For cancers of the mouth, pharynx, larynx, stomach, gallbladder, ovary and prostate, the evidence was considered "probable", while for the rest it was “convincing”. In Renehan et al. the results were only provided stratifying by sex, therefore, we reported the associations between adiposity and risk of each cancer type, that were so for at least one sex. Engeland: We reported associations for obesity-related cancers for at least one sex. "Renal cell carcinoma" was considered instead of "kidney", "acute lymphatic type of leukemia” instead of "leukemia", "plasma cells" instead of "multiple myeloma".
Abbreviations: CNS: central nervous system; IARC: International Agency for Research on Cancer; WCRF: World Cancer Research Fund.

These studies suggested that BMI is associated with the risk of a larger number of cancer types than currently recognized in landmark reviews and meta-analyses and that some of those associations may be non-linear.(22,90,93,97,98,100,101,103,105) However, the main limitations of available studies include limited adjustment for potential confounding, reliance on self-reported weight and height, and lack of generalizability to different populations (ie, these studies were all conducted in Northwestern European countries). Furthermore, conducting analyses stratified by smoking status is critical to provide unbiased estimates of the impact of obesity on cancer risk. From a biological perspective, BMI probably represents

different pathophysiologic alterations in smokers and non-smokers, given that smoking can help maintain a lower BMI (with lower lean body mass compared to a non-smoker), and can promote visceral adiposity and insulin resistance.(106–113) From an epidemiologic perspective, there can also be residual confounding by smoking (BMI values have different interpretations according to smoking status, for example, low BMI in non-smokers likely indicate higher lean body mass and lower adiposity and metabolic consequences of obesity, while in smokers the opposite might be true).(106) Finally, differentiating the risk by smoking status could help targeted recommendations in clinical practice.(106) Despite all these reasons, many of the above-mentioned studies failed to present results stratified by smoking status. (94,96,99–103)

Another question that remains unanswered is whether BMI as a sole indicator of general adiposity fully captures the complex association between adiposity and cancer risk.(114) Fat distribution, which is not accounted for in BMI measurements, has emerged as an important factor for metabolic and cardiovascular disease risk.(114,115) Nevertheless, the evidence in the cancer field is still limited and contradictory. On the one hand, central adiposity, typically assessed using WC, has been related to the risk of several cancer types and a systematic review of 13 studies found that WC (in comparison to BMI) better discriminates risk associated with obesity for gastro-oesophageal, leukemia, liver, biliary, and renal (only in females) cancer risk.(116) Other studies have also suggested WC is more strongly related to colorectal and breast post-menopausal cancers.(117–120) On the other hand, several pooled analyses and a multi-national study have found comparable positive associations of BMI and WC with the risk of pancreatic, endometrial, ovarian, advanced prostate, colorectal, and obesity-related cancers.(121–125) All in all, definitive evidence is lacking and only few studies have systematically compared the effect estimates of BMI and WC for multiple site-specific cancers (of which none have studied less frequently occurring cancer types).(125,126)

Finally, evidence about other BMI-derived exposures assessed during longer periods such as the duration and cumulative overweight/obesity exposure, and the existence of critical age periods to develop overweight/obesity in relation to cancer risk is currently limited. Most of the studies of the BMI-cancer field have mainly focussed on single BMI measurements assessed at study baseline, which are measures of current BMI status. Whether overweight and obesity over the life course are more relevant risk factors for cancer is still in

dispute.(24,93,127) Capturing longitudinal BMI-derived exposures might better reflect the potential underlying biological mechanisms between long-term exposure to adiposity and the increase in the risk of cancer development. At an epidemiological level, this could translate into stronger associations between adiposity and obesity-related cancer risk and into adiposity being linked to the risk of a larger number of cancer types than currently recognized. Duration and cumulative adulthood overweight/obesity exposure have been positively associated with the risk of colorectal, breast postmenopausal, endometrium, kidney, pancreas, and multiple myeloma cancers in cohort studies from the US and Europe.(128–132) Studies investigating the age of onset of overweight and obesity in relation to cancer risk are currently lacking. While previous studies have provided relevant insights into the BMI trajectories-cancer risk association, they have certain limitations. They lack information on both longitudinal exposures and BMI at baseline which is essential to answer whether longitudinal exposures can better capture the effect of adiposity on cancer risk than baseline BMI. Other limitations involve excluding individuals without BMI information (increasing the risk of selection bias), having limited sample sizes that precluded them from analyzing a wider range of cancer types or relying on self-reported and self-recalled weight and height measurements (increasing the risk of exposure misclassification).

1.3.2. Possible mechanisms

Three mechanisms by which higher general adiposity can increase cancer risk have been extensively reported in the literature: sex hormonal metabolism, insulin and insulin-like growth factors (IGF) signaling, and adipokine pathways.(114,133–136) These, in conjunction with the altered immune function and chronic inflammatory state related to obesity, could also be responsible for an association between obesity and the risk of other (than those currently recognized) types of cancers, including rare cancers.(137–139) However, it could also be plausible that other pathways not yet recognized in the literature explain the associations between BMI and the risk of different types of cancers. The proposed mechanisms for the adiposity-cancers associations have evolved, partly due to the impossibility to explain the associations reported in epidemiologic studies using only one mechanism or a set of mechanisms.(114,133) This highlights the important role of epidemiologic research in understanding the biological mechanisms between adiposity and cancer risk which could, in turn, help inform preventive strategies.(114)

WC could be a better indicator of cancer risk compared to BMI because WC (as a surrogate marker of central adiposity) is more strongly correlated with visceral adiposity.(114,117,140) Visceral adipose tissue is thought to play a role in the development of different diseases, including cancers, given that metabolically active visceral fat releases substantial amounts of growth factors, inflammatory markers, free fatty acids that contribute to insulin resistance, locally produced estrogen and adipokines which might contribute tumor development.(114,141,142)

Longer, earlier, and cumulative overweight/obesity exposure have been related to key mechanisms on the obesity-cancer pathway. For instance, they can increase the risk and severity of hypertension (HTN), insulin resistance, chronic inflammation, oxidative DNA damage, and alterations in endogenous hormone metabolism. (114,143,144)

KEY MESSAGES

Adiposity and Cancer risk association

- Body mass index (BMI) has been convincingly associated with the risk of at least 12 cancer types. Prior studies suggest **BMI may be linked** to **more cancer** types than currently recognized but more data from large cohort studies is needed to confirm these associations.
- **More evidence** on the BMI-cancer relationship **stratified by smoking status** is needed from a biological (different pathophysiologic alterations of BMI in smokers), epidemiological (residual confounding), and clinical (targeted recommendations to patients) perspective.
- Waist circumference (WC) may be a better discriminator of the risk associated with obesity for certain cancers; however, a **systematic comparison** of **BMI** and **WC** for multiple site-specific cancers **is lacking.**
- Evidence about the association between **BMI from a life course perspective** and cancer risk, using indicators such as duration and cumulative overweight/obesity exposure, and age of onset of overweight/obesity, is **currently limited.**

1.4. Cardiometabolic conditions as modifiers of the adiposity-cancer association

Adiposity has also been associated with a higher risk of cardiometabolic conditions such as HTN, T2DM, and CVDs.(145)

1.4.1. Hypertension

HTN is a condition characterized by high blood pressure (ie, high force exerted by circulating blood against the walls of the body's arteries).(146) HTN can be captured in epidemiologic studies using diagnostic codes (eg, ICD-10 codes recorded by primary care physicians in the context of this Thesis). To diagnose this condition, there must be at least two altered measurements of blood pressure (either a systolic blood pressure ≥140 mmHg and/or a diastolic blood pressure ≥90 mmHg) observed in two different days.(146)

The global prevalence of HTN has nearly doubled in the last three decades: going from 594 million in 1975 to 1.13 billion adults in 2015 (22% of the global population).(146) This condition mostly affects low- and middle-income countries (two-thirds of the people with HTN live in these countries).(2) Estimations of the prevalence of HTN in Spain vary widely, going from 19.2% (in 2015) according to the WHO to 43% (in 2009-2010) according to a study representative of the Spanish population (Di@bet.es).(147,148)

Several modifiable risk factors of HTN have been identified such as unhealthy diets (high consumption of salt and saturated/trans fats, and low intake of fruits, vegetables, potassium, and calcium), sedentary lifestyle, stress, smoking, high alcohol intake, and having overweight or obesity.(146,149–151) Non-modifiable risk factors include family history of HTN, aging, and race (eg, in the US, HTN is more frequent and develops earlier among Blacks compared to Whites).(146,149–151) HTN is an important public health problem because it affects nearly a quarter of the global adult population and increases the risk of heart attack, stroke, aneurysm, heart failure, angina, kidney diseases, vision loss, dementia, and certain cancers, among others.(146,152,153)

A study with data from 7 cohorts (577k adults) from Norway, Austria, and Sweden found that blood pressure was positively associated with the risk of overall cancer incidence in men.(154) In addition, a literature review (including 15 cohort and 3 case-control studies) and a systematic review and meta-analysis (of 48 studies) consistently reported an association between HTN

and risk of kidney cancer.(155,156) Several studies included in these reviews also found positive associations between HTN and risk of stomach, colorectal, pancreas, breast cancer (overall and post-menopausal), brain, and malignant melanoma (only men) cancers -which are also cancer types associated with higher adiposity levels- as well as lung cancer.(90,93)

1.4.2. Type 2 diabetes mellitus

T2DM is a chronic disease that occurs when the body cannot use the insulin produced by the pancreas effectively.(157) This condition can be captured in epidemiologic studies using diagnostic codes (eg, ICD-10 codes recorded by primary care physicians in the context of this Thesis). To diagnose a patient with this condition, at least two observations of i) fasting blood sugar test ≥26 mg/dl (7 mmol/l), ii) oral glucose tolerance test ≥200 mg/dl (11.1 mmol/l), iii) glycated hemoglobin test ≥6.5%, or iv) a single observation of random blood sugar test ≥200 mg/dl (11.1 mmol/l) accompanied by typical symptoms of the disease must be observed.(158) However, T2DM is a condition that is frequently undiagnosed, likely because symptoms are often absent or less marked than those of type 1 diabetes. In high-income countries, the proportion of undiagnosed diabetes is estimated to be as high as 30% to 50%.(159–162)

The worldwide prevalence of diabetes among adults nearly doubled over the last three decades, going from 4.7% in 1980 (108 million adults affected) to 8.5% in 2014 (422 million).(159,163) This prevalence has risen substantially in all countries (regardless of the income level), although it has risen faster in low- and middle-income countries compared to high-income countries.(159) (*Nota bene*: given that sophisticated laboratory tests for pancreas function are needed to distinguish type 1 from type 2 diabetes, global estimates of the prevalence consider diabetes as a whole; nevertheless, T2DM accounts for approximately 90% of diabetes cases.(159,164)) The estimations of the prevalence of diabetes in Spain vary widely. A study reported estimates from nine other studies ranging from 4.8% to 18.7%.(165) A more recent (2009-2010) study having as the main aim to estimate the prevalence of diabetes in Spain estimated that 13.8% of the Spanish population have diabetes (although it has been only diagnosed among 7.8% of them).(148)

There are several modifiable (eg, overweight or obesity, high central adiposity, unhealthy diet, physical inactivity, smoking, low levels of high-density lipoprotein cholesterol, or high levels of triglycerides) and non-modifiable (eg, race/ethnicity, family history of diabetes, gestational diabetes or older age) factors which can increase the risk of T2DM.(159,166) T2DM is a major

global health problem because it affects nearly 8.5% of the adult population worldwide and because it increases the risk of cardiovascular events (eg, heart attack or stroke), kidney failure, leg amputation, vision loss, nerve damage, skin conditions, poor healing, hearing impairment, sleep apnea, dementia, and certain cancer types.(159,166–177)

T2DM has been positively associated with the risk of all cancer types combined and specifically with the risk of cancers of the stomach; colorectal; liver; gallbladder and biliary tract; pancreas; breast; corpus uteri; kidney; bladder; thyroid; non-Hodgkin lymphoma; multiple myeloma; and leukemia in different meta-analyses.(167–177) T2DM has also been negatively related to the risk of prostate cancer.(168,178) Except for bladder cancer, all other positive (and negative) associations between T2DM and the risk of different cancer types are concordant with those related to higher adiposity levels.(90,93)

1.4.3. Cardiovascular diseases

CVDs are a group of disorders affecting the heart and blood vessels including coronary heart disease (which occurs when the arteries that supply blood to the heart muscle become hardened and narrowed due to a buildup of plaque on the inner walls of the arteries) and cerebrovascular disease (which happens when the blood supply to part of the brain is cut off or temporarily disrupted), among other conditions.(179–181) CVDs can be captured in epidemiologic studies using diagnostic codes (eg, ICD codes recorded in hospitals and primary care centers in the context of this Thesis). CVDs are diagnosed using several tests such as electrocardiogram, exercise stress tests, coronary angiography, or intracoronary ultrasound.(180,181)

The worldwide prevalence of CVD has nearly doubled in the last four decades, going from 271 million in 1990 to 523 million in 2019.(182) The number of CVD deaths has also steadily increased from 12.1 million in 1990 to 18.6 million in 2019, being the leading cause of death globally.(179,181) Low- and middle-income countries are especially affected by these diseases (more than three-quarters of CVD deaths occur in these countries) compared to high-income ones.(179) In Spain, CVD is the leading cause of hospitalization and death being responsible for approximately 5 million hospital admissions and 125 000 deaths yearly.(183)

Non-modifiable risk factors of CVD include age, gender, ethnicity, and family history of CVD. Modifiable risk factors comprise high blood pressure, high cholesterol, smoking, overweight and obesity, physical inactivity, unhealthy diet, and high alcohol intake.(179–181)

Emerging evidence has suggested that CVD and cancer risk are associated.(184) This association might be explained by shared risk factors or by CVD being an independent risk factor for cancer (which we discuss in the following section [1.4.4.]).

1.4.4. Possible mechanisms

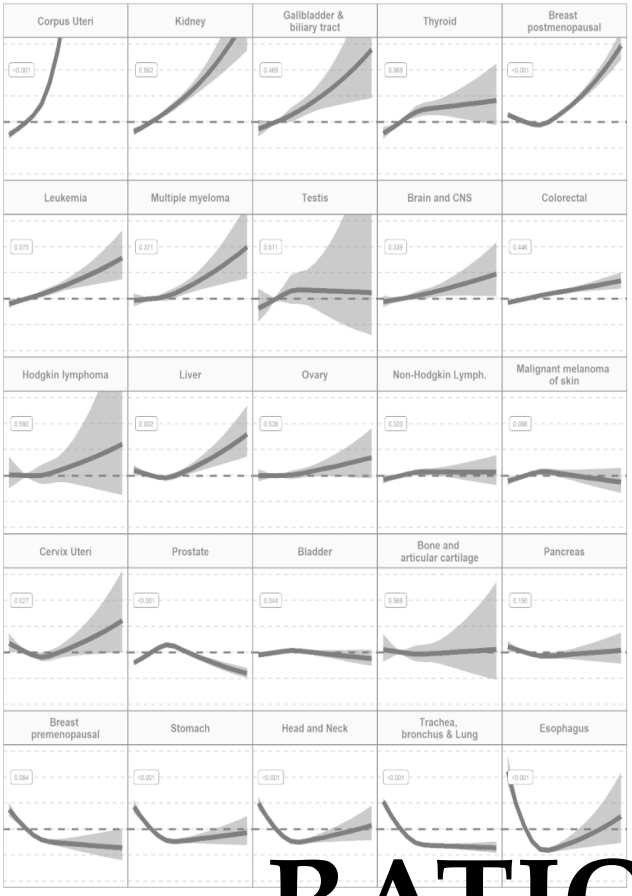
Adiposity, HTN, and T2DM are risk factors of both CVD and cancer and common pathways have been proposed to explain these relationships.(184) The association between adiposity and CVD and cancer risk might be mediated by several shared mechanisms including hormones (ie., sex hormones, insulin and IGF signaling, and adipokines), inflammation, and oxidative stress.(114,133–136,184) Emerging evidence also suggests CVD might be an independent risk factor for cancer (cardiac proteins excreted into the bloodstream after myocardial infarction might promote tumor growth).(185) While HTN is a well-established risk factor for CVD and the pathways behind this link have been widely investigated (eg, HTN induces oxidative stress on the arterial wall which can explain its atherogenic influence), the mechanisms by which HTN can promote cancer are less well established.(184,186) Some shared mechanisms might involve elevated levels of plasma vascular endothelial growth factor and oxidative stress.(184,187,188) The T2DM-CVD and T2DM-cancer associations may have hyperinsulinemia, hyperglycemia, IGF signaling, and inflammation as possible mediators.(184,189–193)

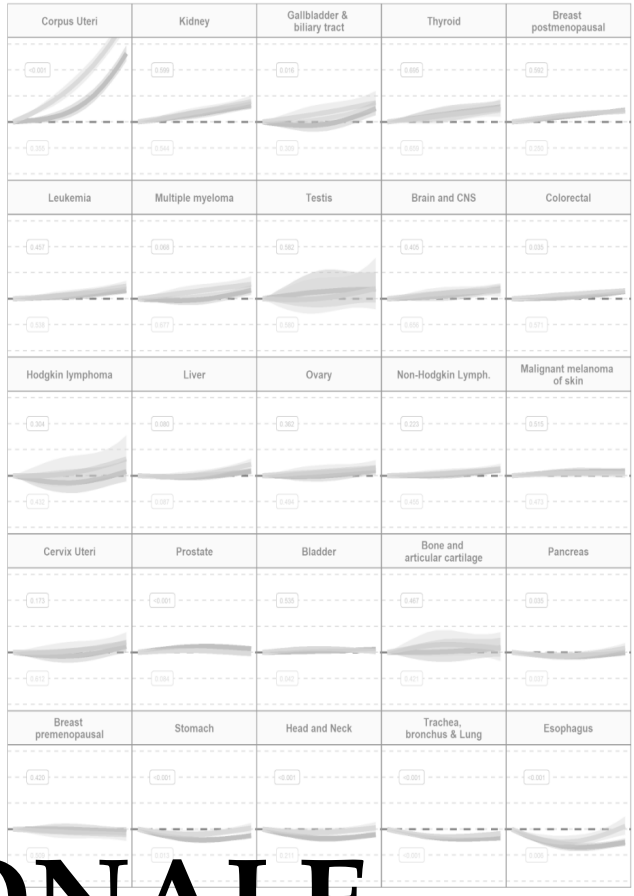
Because of the strong interrelation between adiposity, HTN, T2DM, CVD, and cancer, incident HTN, T2DM, CVD may modify cancer processes associated with obesity through the above-mentioned shared biological pathways. For example, by synergistically enhancing inflammatory, oxidative stress, or hormonal processes.(184,194) However, the extent to which these cardiometabolic conditions may modify the BMI-cancer association is unclear given that prior studies of the field have mostly focused on healthy or general populations.(104,105) In addition, prior studies have not investigated the combination of component risk factors (ie, adiposity and incident cardiometabolic conditions) and cancer risk.

KEY MESSAGES

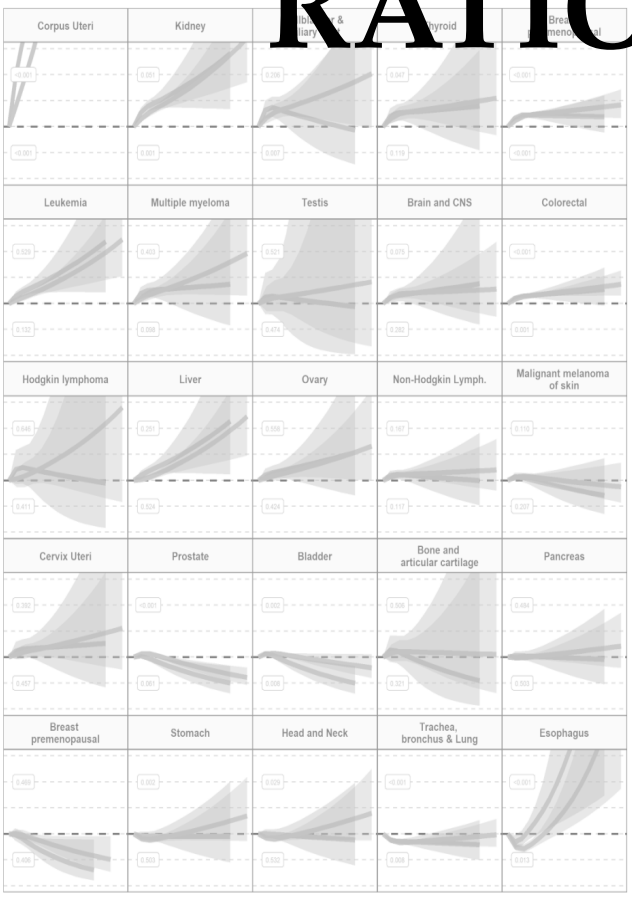
Cardiometabolic conditions as modifiers of the adiposity-cancer association

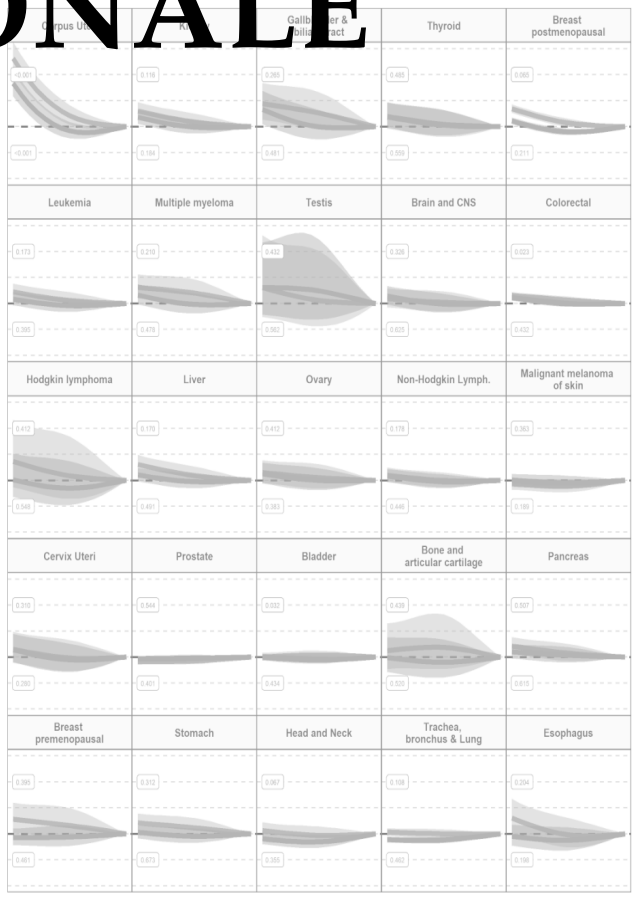
- A high **BMI** has been **related to** cancer risk and cardiometabolic conditions such as hypertension (**HTN**), type 2 diabetes (**T2DM**), and cardiovascular diseases (**CVD**).
- **HTN and T2DM** have been proposed as **risk factors for cancer**. **CVD and cancer share** common basic **biological pathways,** and emerging evidence also suggests CVD might be an independent risk factor for cancer.
- **Incident HTN, T2DM, and CVD may modify cancer processes** associated with obesity through a cascade of shared (eg, by synergistically enhancing inflammatory, oxidative stress, or hormonal processes) and non-shared biological pathways.
- However, the extent to which these cardiometabolic **conditions modify the BMI-cancer** association and the **combined effect** of component risk factors (ie, adiposity and incident cardiometabolic conditions) in relation to cancer risk **remains unclear.**


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

RATIONALE


Corpus Uteri
Kidney
Thyroid
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

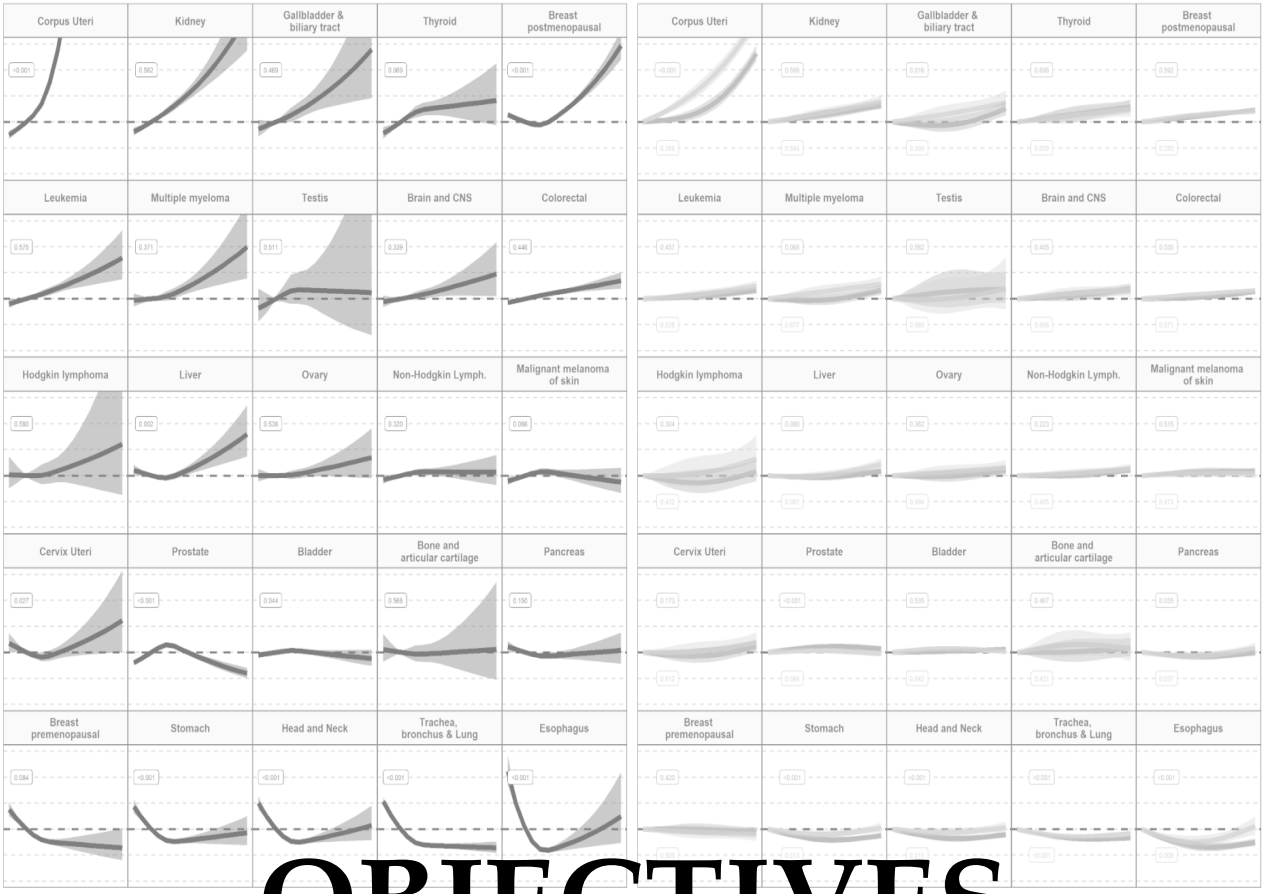
2. Rationale

Cancer is one of the leading causes of morbidity and mortality worldwide and in Spain.(3,6,7) A better understanding of the role of modifiable lifestyle factors such as obesity in cancer risk is essential for the implementation of preventive strategies of cancer at the individual and population levels.

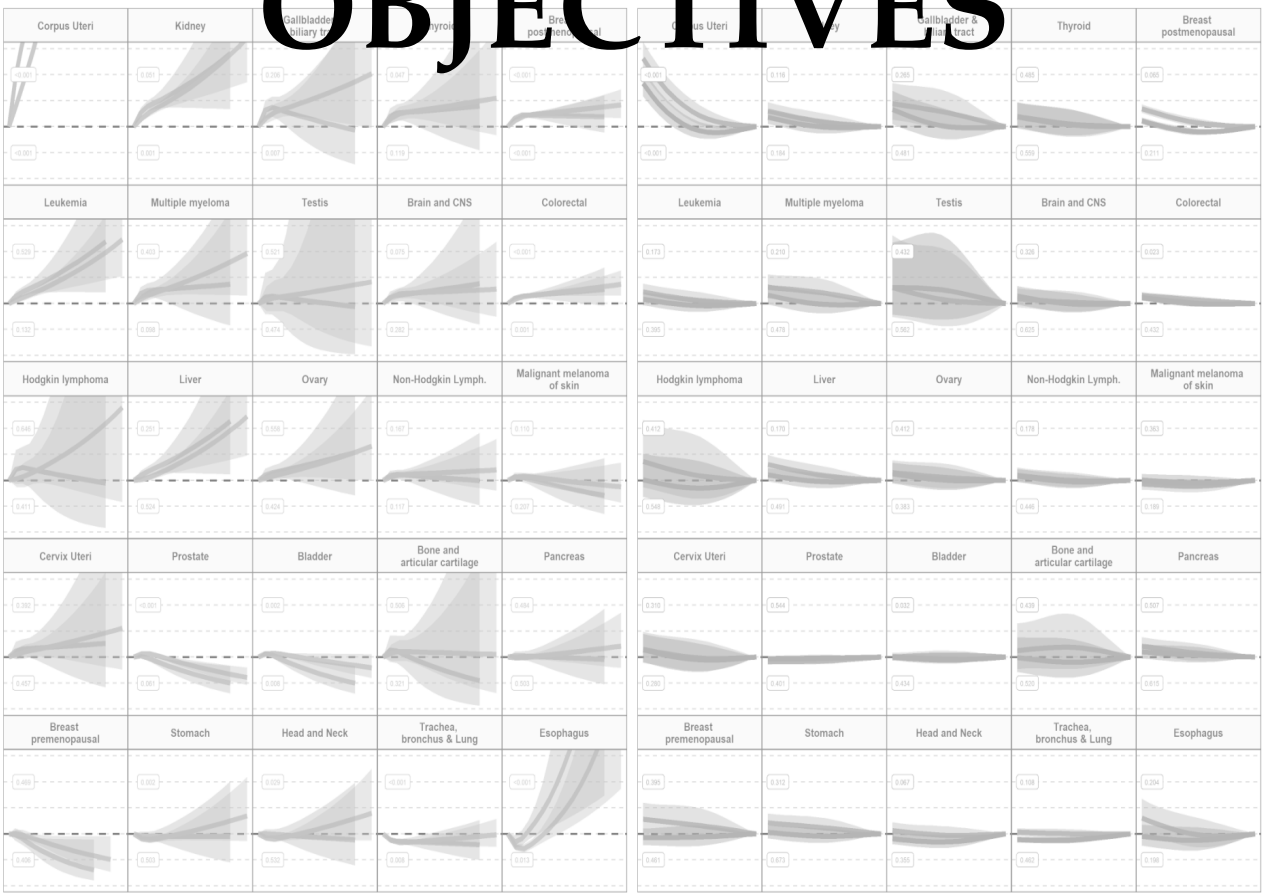
The use of databases of routinely collected EHRs has become more common in epidemiology and clinical research. EHRs can be defined as a longitudinal collection of electronic health information about individuals and populations.(195) Due to their size, amount of data availability, representativeness, long-term follow-up, and the sufficient statistical power they provide (eg, to detect uncommon outcomes such as rare cancer types), EHR databases offer a great opportunity to conduct cancer-related research.(196,197)

Cohort studies using EHRs could help fill in gaps in the adiposity-cancer association. Namely, whether BMI at baseline is related to the risk of more cancer types than currently recognized in the literature and if these associations differ by the smoking status of the participants. Another important question that still needs to be answered is whether BMI as a sole indicator of general adiposity fully captures the complex association between adiposity and cancer risk. In addition, evidence about the relationship between BMI from a life course perspective and cancer risk (using indicators such as duration and cumulative overweight/obesity exposure and age of onset of overweight/obesity) is currently limited. Finally, the extent to which cardiometabolic conditions such as HTN, T2DM, and CVD modify the BMI-cancer association and the combined effect of component risk factors (ie, adiposity and incident cardiometabolic conditions) in relation to cancer risk remains unclear.

However, to make good use of EHR databases for epidemiological research, some steps need to be priorly undertaken. Firstly, it is essential to understand if the available data sources are suitable (ie, what is the data collected and its scope) to achieve the aims of a specific study. This can be addressed by conducting exhaustive descriptive studies of the EHR databases of interest. Secondly, it is important to conduct validation processes to quantify the correctness of the data and increase the reliability for use in subsequent observational studies.(198) This issue can be dealt with through validation studies of the outcomes of interest (cancer diagnoses in the context of this Thesis).


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

OBJECTIVES


Corpus Uteri
Kidney
Thyroid
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

3. Objectives

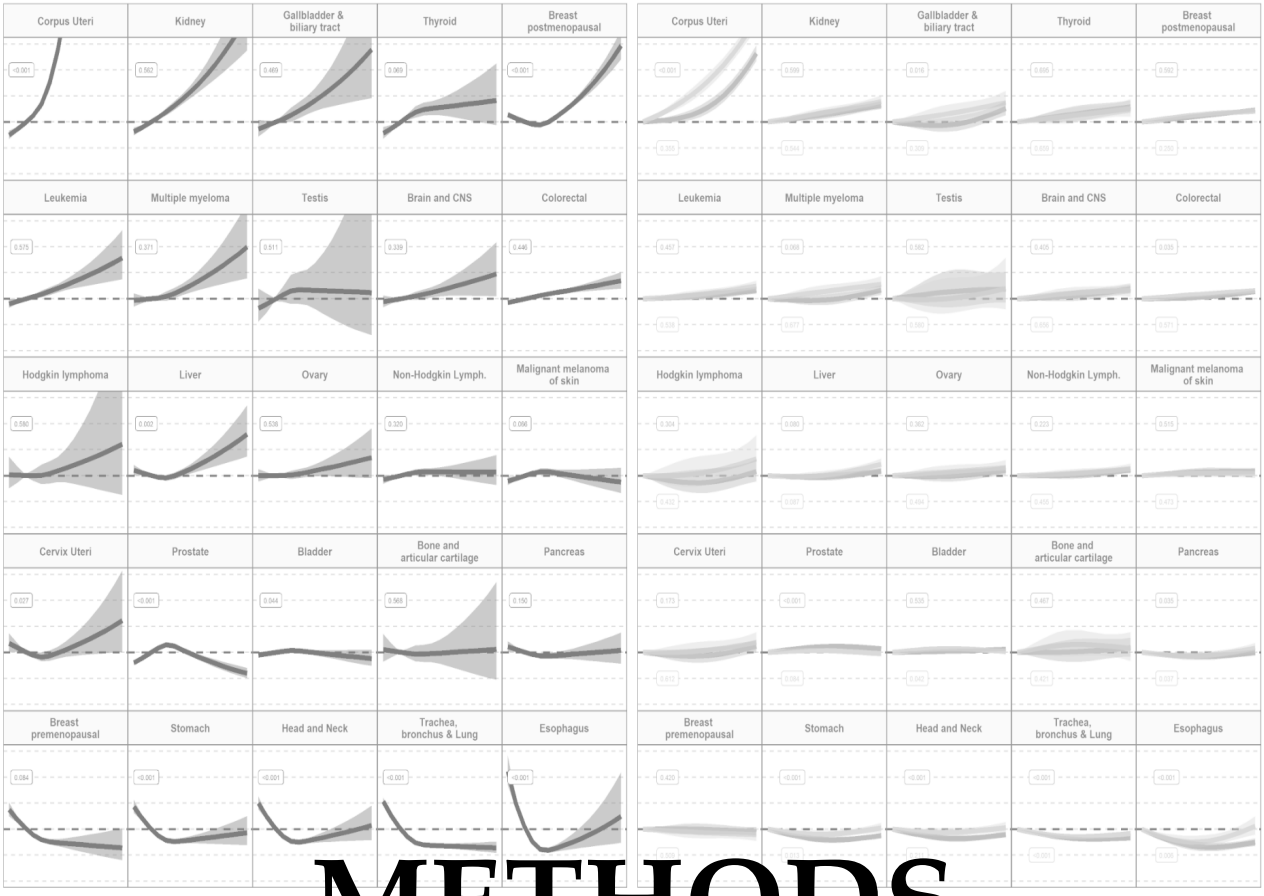
The main aim of this Thesis was to investigate the association between adiposity and cancer risk. As a prior step, we aimed to evaluate the suitability of a large EHR database from Catalonia, Spain for research and, more specifically, for cancer-related research. This Thesis had three specific objectives:

1. To provide an extensive characterization of the Information System for Research in Primary Care (SIDIAP) database (Study I in the Results section).

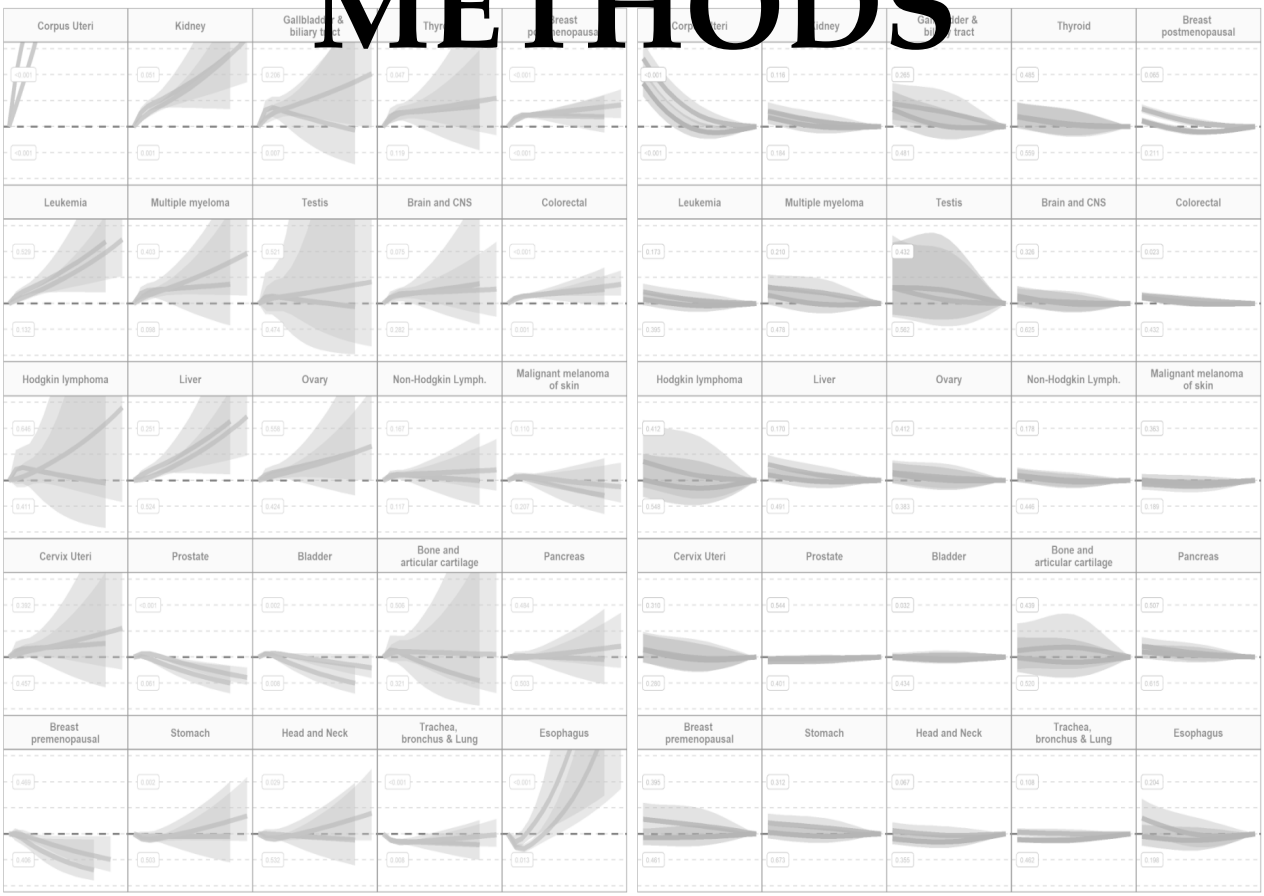
2. To validate twenty-five types of incident cancer cases in the SIDIAP using the population-based cancer registries of Girona and Tarragona as the gold standard and to assess the time difference in the date of diagnosis between the SIDIAP and these cancer registries (Study II in the Results section).

3. To investigate the association between adiposity and the risk of 26 types of cancer accounting for potential non-linearity, different adiposity indicators, and individual-level factors such as smoking status and incident cardiometabolic conditions (Study III in the Results section, Studies IV and V in the Appendix).

The specific objectives of each study can be consulted in Results section where we included the full manuscripts corresponding to Studies I-III (published articles) and in Appendices 1 and 2 where we included those corresponding to Studies IV and V (manusctipts submitted to scientific journals).


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

METHODS


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

4. Methods

In this section, we give a general overview of the study designs, setting, data sources, study populations, assessment of key variables, and statistical analyses used in this Thesis. A more detailed description of the methods and, specifically, of the statistical analyses corresponding to each study is included in Section 5 (in the Methods section of Studies I-III) and in the Appendices 1 and 2 (in the Methods section of Studies IV and V, presented as manuscripts submitted to scientific journals).

4.1. Study designs, setting, and data sources

Studies I and II were cross-sectional studies including data from 2006 until 2021 and from 2009 until 2015, respectively. Studies III, IV, and V were cohort studies with data from 2006 until 2018.

The five studies of this Thesis were population-based and included prospectively collected data from the SIDIAP (www.sidiap.org). SIDIAP includes routinely recorded information by health professionals from 287 primary care centers in Catalonia, a region in Northeastern Spain.(199) SIDIAP contains pseudo-anonymized records for 5.8 million people, including data on anthropometric measurements, disease diagnoses (using the ICD-10 coding system), demographic and lifestyle information, among others. Further, SIDIAP is linked to the Minimum Basic Dataset (CMBD in Spanish), a population-based registry that includes hospital discharge information in Spain, which was also used for four studies (Studies II to V).(200) A broad description of the SIDIAP database (including its possible linkages) is given in Study I.

For Study II, we also used data from the cancer registries of Girona (created in 1994) and Tarragona (in 1980) which cover 20% of the Catalan population.(201,202) They collect cancer diagnoses from public and private hospitals, anatomopathological and hematological laboratories, mortality registries, and other information sources.(203–205) Both cancer registries comply with the IARC quality requirements.(206)

4.2. Study population

Study I included all the individuals registered in the SIDIAP database (all-time population and population as of mid-2021).

Study II included all adults considered as incident cancer cases (with a first cancer diagnosis) from 2009 to 2015 among inhabitants of the provinces of Girona and Tarragona.

In Study III, we included all individuals registered in the SIDIAP database aged ≥18 years with a valid BMI (comprised between 15 and 60 kg/m^2) recorded between 2006 and 2017 and free of cancer at baseline. For the secondary objective of this study, we included an additional eligibility criterion, which was to have a valid WC assessment (WC values ≥ 40 and ≤ 160 cm) no more than 5 years previous to or 1 year later than the first BMI measurement recorded.

For Study IV, we included all the individuals registered in the SIDIAP database who were aged ≥40 years in 2009 who had at least one year of history in the SIDIAP database and were free of cancer at baseline.

Study IV included all individuals aged ≥40 years registered in the SIDIAP database in 2010 without prevalent HTN, T2DM, CVD, and/or cancer at baseline.

4.3. Variables

4.3.1. Exposures

Given the aims of Studies I and II we did not have any exposure in those studies. For Study I, we described the characteristics of the SIDIAP population based on different variables (eg, sociodemographics, lifestyle, or laboratory values), and for Study II, we compared the cancer diagnoses registered in the SIDIAP database and the cancer registries of Girona and Tarragona using the cancer definition reported in Section 4.3.2..

For Studies III and V, the main exposure of interest was BMI at study entry as a continuous variable (in kg/m^2). BMI was automatically calculated through a computer program ("Estació clínica d'atenció primària") after general practitioners (GPs) or nurses entered the weight (kg) and height (cm) of patients they directly assessed in a standardized manner.(207)

Additionally, for the Secondary objective of Study III, we also considered WC as an exposure; this indicator was routinely measured by GPs or nurses who follow a measurement protocol.(158) WC was measured at the umbilical level, midway between the anterior superior iliac spine and the inferior border of the rib while participants were standing.

For Study IV, we used BMI measurements from all the adults in the SIDIAP database to obtain imputed individual trajectories of BMI. We used the BMI trajectories to calculate the exposures

of interest. The window to capture longitudinal exposures was during early adulthood, between the ages of 18 and 40 years, and was separated from the time-to-event window. We generated six longitudinal exposures. The *duration of BMI ≥25* kg/m^2 (and of *≥30*, respectively) was the sum of years lived with a BMI ≥25 (≥30) kg/m^2. *Cumulative exposure to a BMI ≥25* kg/m^2 (and *≥30*) was calculated by summing the differences between the BMI measurements that were ≥25 (≥30) kg/m^2 and 24.9 (29.9) kg/m^2 for every year lived with a BMI ≥25 (≥30) kg/m^2. For all the other years, the value of the cumulative exposure was set to 0.(143,144,208) *Age of onset of a BMI ≥25* (and *≥30*) kg/m^2 was the age at which a person had a BMI measurement ≥25 (≥30) kg/m^2 for the first time in the trajectory and was only available for individuals who ever had a BMI ≥25 (≥30) kg/m^2.

Finally, for the Secondary objective of Study V, the exposure of interest was a composite variable of 16 categories combining binary BMI (< or ≥25kg/m^2) and cardiometabolic conditions which was coded as a time-varying variable with eight categories ("*healthy"*; *HTN*; *T2DM*; *CVD*; *HTN & T2DM*; *HTN & CVD*; *T2DM & CVD*; *HTN, T2DM, & CVD*). All study participants were in the "*healthy"* category at index date, and during follow-up, they could change states to one (*HTN*; *T2DM*; or *CVD*), two (*HTN & T2DM*; *HTN & CVD*; or *T2DM & CVD*), or three (*HTN, T2DM, & CVD*) cardiometabolic conditions.

4.3.2. Cancer definition

We created a definition of 25 incident cancer types in the SIDIAP adult (aged ≥18 years) population for Study II. Subsequently, we adapted this definition to satisfy the aims of Studies III, IV, and V where we used it to define the outcomes of these studies (changes mentioned in Section 4.3.3.).

We used ICD-10 codes and date of diagnosis to identify the following cancer types: head and neck (ICD-10 codes: C00-C14), esophagus (C15), stomach (C16), colorectal (C18–21), liver (C22), gallbladder and biliary tract (C23-24), pancreas (C25), larynx (C32), trachea, bronchus, and lung (C33-34), bone and articular cartilage (C40-C41), malignant melanoma of skin (C43), breast (C50), cervix uteri (C53), corpus uteri (C54-C55), ovary (C56), prostate (C61), testis (C62), kidney (C64), bladder (C67), brain, CNS, pituitary gland and pineal gland (C70-72, C75.1-C75.3), thyroid (C73), Hodgkin lymphoma (C81), non-Hodgkin lymphoma (C82-C86, C96), multiple myeloma (C90), and leukemia (C91-95).(19) We also considered diagnoses registered in the hospital discharge database (CMBD) which were coded in ICD-9.(209) We

mapped diagnostic codes to ICD–10 using available conversion codes eCIEMaps v3.1.9 which are available in the supplementary material of Study II.

For Study II, we also reproduced this definition using data from the cancer registries of Girona and Tarragona.

4.3.3. Outcomes

Given the aims of Studies I and II, we did not have any outcomes in those studies.

For Studies III and IV, we considered incident cancer diagnoses (registered in the SIDIAP database and the CMBD, the hospital discharge database) as the outcomes of interest. We used the same cancer types as specified in the cancer definition of Section 4.3.2.. Additionally, we categorized breast cancer into pre- and postmenopausal due to well-established evidence indicating different mechanisms of adiposity in the risk of this cancer according to menopausal status at cancer diagnosis.(133)

The outcome of interest for Study V was a binary variable indicating a diagnosis of any obesity-related cancer. We identified obesity-related cancers as diagnostic codes registered in the SIDIAP database and the CMBD (hospital discharge database). We defined obesity-related cancers as those listed in the IARC Viewpoint Working Group review report as having well-established evidence indicating that the presence of excess body fatness increases the risk of these cancers.(90) Therefore we included colorectum; liver; gallbladder and biliary tract; pancreas; postmenopausal breast; corpus uteri; ovary; kidney; brain and CNS; and multiple myeloma as obesity-related cancers using the cancer definition of Section 4.3.2.. We omitted esophagus and stomach cancers from this list given that with the available data we could not differentiate esophageal adenocarcinoma (obesity-related) from squamous cell carcinoma nor gastric cardia (obesity-related) from non-cardia cancers. Furthermore, the incidence of the non-obesity-related subtypes of these cancers is higher in Spain.(210) All the diagnostic codes used to define the outcomes can be consulted in the supplementary material of Study IV.

4.4. Statistical analyses

In Study I, we conducted different types of descriptive analyses. We described the origin and purpose of the SIDIAP database, the overall characteristics of the SIDIAP population, and the SIDIAP's population representativeness of the general population of Catalonia in terms of the

geographic, sex, and age distribution. We also described key data available in the SIDIAP database and the evolution in the recording of key variables and data domains in the SIDIAP population over the years.

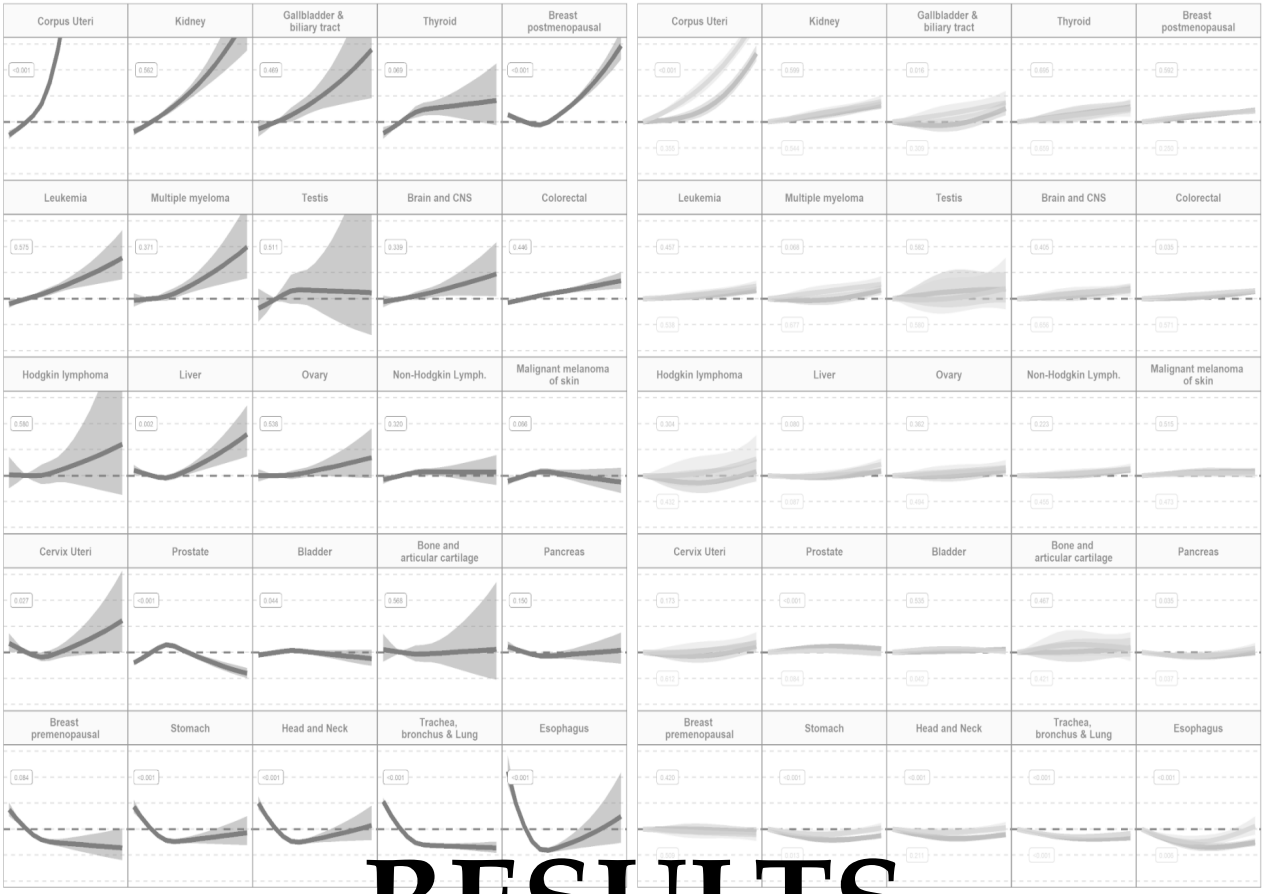
In Study II, we calculated the sensitivity, positive predictive values (PPV), and the time difference between the date of diagnosis entered into the SIDIAP and the cancer registries. We added hospital discharge cancer diagnoses to the SIDIAP to assess sensitivity changes.

For the primary objective of Study III, we fitted multivariable-adjusted Cox proportional hazard models to estimate cause-specific hazard ratios (HR) and 99% CIs for the relation between BMI and risk of each cancer type. We investigated potential non-linear associations between BMI and risk of each cancer by fitting models using restricted cubic splines for BMI. To assess residual confounding by smoking, we re-run the main models among never smokers. For the secondary objective of Study III, we compared risk estimates for general (BMI) and central (WC) adiposity in relation to the risk of 26 cancers by fitting multivariate-adjusted Cox proportional hazard models (one for each adiposity indicator). We estimated HRs and 99% CIs per 1 standard deviation (SD) increment of adiposity indicators (BMI and WC) to allow comparability between both estimates.(125)

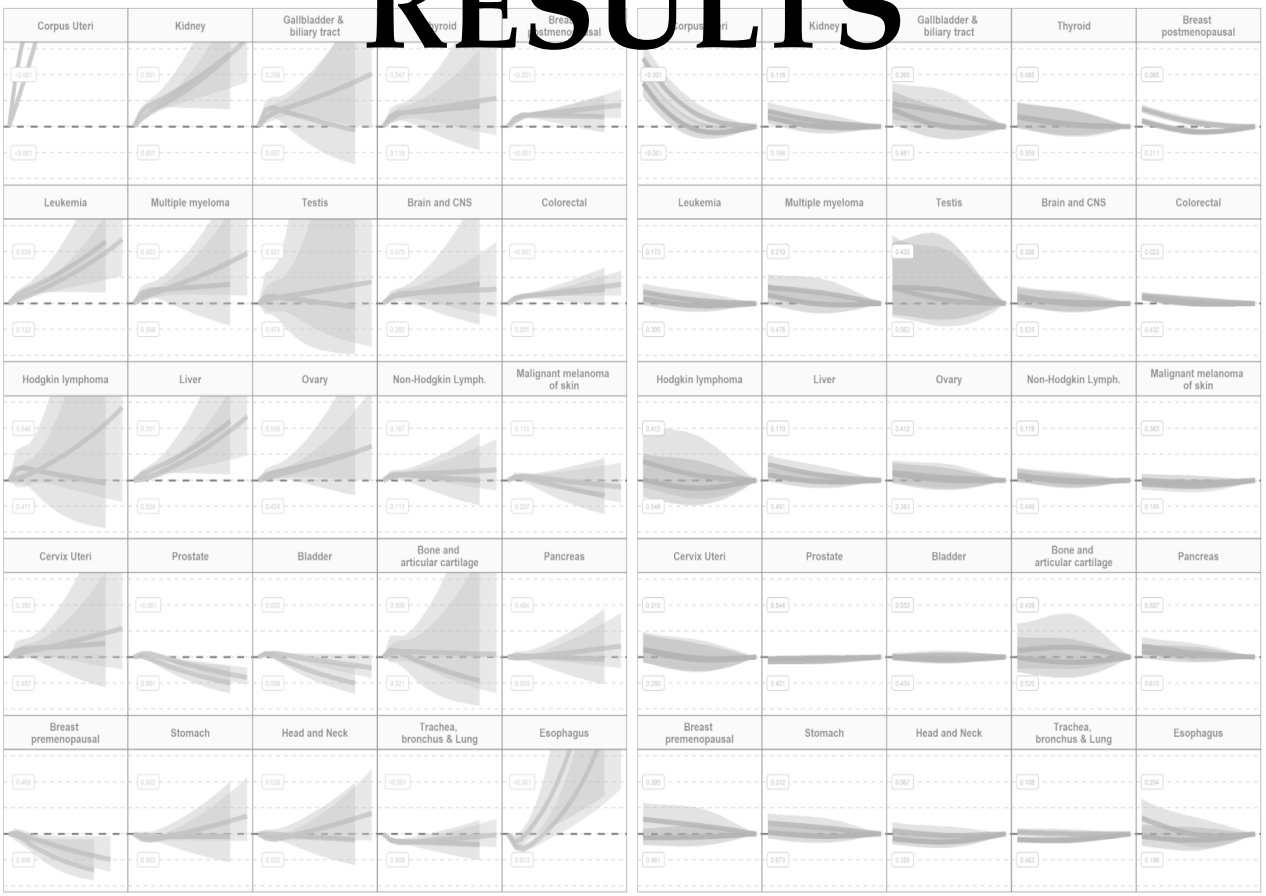
In Study IV, we investigated the association between each of the exposures of interest with the risk of the 26 cancer types by estimating multivariable-adjusted Cox proportional hazard models. We calculated the HRs and their respective 95% CIs for each cancer type per 1 SD increment of each exposure to allow comparability between the different HRs.(125) We checked whether the 95% CIs of the HR of each longitudinal exposure overlapped with that of BMI at index date to assess differences in the strength of the associations between the longitudinal exposures and BMI at index date. We assessed non-linearity in the relationship between each exposure and each cancer type using restricted cubic splines for each exposure.(211,212)

To investigate if incident HTN, T2DM, and CVD modify the association between BMI and the risk of obesity-related cancers (primary objective of Study V) we fitted multivariable-adjusted Cox proportional hazard models including BMI, cardiometabolic conditions as a time-varying variable, and an interaction of those with BMI. We estimated HRs and their 95% CIs per 5 kg/m^2 increment of BMI. For the secondary aim of Study V, we assessed the relative excess risk due to interaction (RERI) of obesity-related cancers between overweight/obesity (BMI≥25

kg/m^2) and incident cardiometabolic conditions by fitting a multivariable-adjusted Cox proportional hazard model using the above-mentioned composite variable with 16 categories.(213) The RERI was calculated as $RERI_{RR} = RR_{11} - RR_{10} - RR_{01} + 1$, where $_{11}$ denotes being exposed to both factors (eg, overweight/obesity and *HTN*), $_{10}$ to one factor (eg, overweight/obesity), and $_{01}$ to the other one (eg, *HTN*). A RERI of 0 was considered a lack of additive interaction and 95%CIs were calculated as proposed by Hosmer and Lemeshow.(214)


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

RESULTS


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

5. Results

In this section, we only show the results of three studies (Studies I-III) which are manuscripts published in scientific journals. For the sake of completeness, in Appendices 1 and 2, we included the manuscripts corresponding to Studies IV and V (which also address the third objective of this Thesis). These studies are a continuation of Study III and are submitted to scientific journals.

5.1. Study I

Recalde M, Rodríguez C, Burn E, Far M, Manuel-García D, Carrere-Molina J, Benítez M, Moleras A, Pistillo A, Bolíbar B, Aragón M, Duarte-Salles T.

Data Resource Profile: the Information System for Research in Primary Care (SIDIAP)

International Journal of Epidemiology, dyac068: 1-13.

International Journal of Epidemiology, 2022, 1–13
https://doi.org/10.1093/ije/dyac068
Data Resource Profile

Data Resource Profile

Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)

Martina Recalde [1,2†], Clara Rodríguez,[1†] Edward Burn [1,3], Marc Far,[1] Darío García,[1] Jordi Carrere-Molina,[1] Mencia Benítez,[1] Anna Moleras,[1] Andrea Pistillo,[1] Bonaventura Bolíbar,[1,2] María Aragón,[1*‡] and Talita Duarte-Salles [1‡]

[1]Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain, [2]Universitat Autònoma de Barcelona, Bellaterra, Spain and [3]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, UK

*Corresponding author. Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAPJGol), Gran Via Corts Catalanes, 587 àtic, 08007 Barcelona, Spain. E-mail: maragon@idiapjgol.info

[†]Joint first authors.

[‡]Joint senior authorship.

Received 4 November 2021; Editorial decision 1 March 2022; Accepted 29 March 2022

Data resource basics

Primary care in Catalonia, Spain

Spain has a universal taxpayer-funded health system that is decentralized to its 17 autonomous communities (the country's first level of political and administrative division).[1] Primary care is free of charge and is the main entry point for accessing public non-emergency health-related services that are delivered through primary care. Persons can be referred to secondary care if necessary. Primary care centres are composed of general practitioners (GPs), paediatricians, dentists, nurses, social workers, auxiliary nurses and administrative staff. Additionally, as part of primary care, there are a set of support services such as sexual and reproductive health or home care at the end of life.

In Catalonia, an autonomous community in north-eastern Spain, the *Institut Català de la Salut* (ICS, Catalan Health Institute) is the largest healthcare provider.[2] As seen in Figure 1, ICS covers most of the Basic Health Areas (territorial units by which primary healthcare services are organized in Catalonia) across Catalonia. It manages 328 primary care centres covering 5.8 million people, 75% of the population living in Catalonia (the remaining 25% is distributed among other providers whose services are hired by the Department of Health) (Figure 1 and Table 1).

The Information System for Research in Primary Care

The Information System for Research in Primary Care (SIDIAP; www.sidiap.org) database includes routinely collected data by >30 000 professionals from the ICS. During the 1990s, the ICS created a computerized programme [*estació clínica d'atenció primària* (e-CAP)] for the recording of information during primary care visits in a structured format that has been in use since 2005. In 2010, the ICS and the Institute for Primary Health Care Research Jordi Gol i Gurina (IDIAPJGol) created SIDIAP, which included the data collected through the e-CAP programme since 2006. SIDIAP was designed to provide a valid and reliable database of selected information from the patients' electronic health records (EHRs) for research.[3]

Table 1 presents the main characteristics of the SIDIAP population. The database has information on 8 036 948

Key Features

- The Information System for Research in Primary Care (SIDIAP) is a database of population-wide primary care electronic health records that was created to provide a useful tool for healthcare research.
- SIDIAP includes data from 328 primary care centres managed by the Catalan Health Institute in Catalonia, Spain. The database contains pseudo-anonymized records for >8 million people since 2006, with 5.8 million people active in June 2021 (75% of the Catalan population). SIDIAP is representative of the general population living in Catalonia in terms of age, sex and geographic distribution.
- SIDIAP is updated on a 6-monthly basis and the median follow-up time of the population is currently 15.5 years.
- SIDIAP includes high-quality data on demographics, all-cause mortality, disease diagnoses, prescription and dispensation of drugs, laboratory tests, socio-economic indicators, vaccinations, lifestyle information, parent–child linkage and clinical parameters, among others. SIDIAP can be linked on a project-by-project basis to other data sources such as hospital discharges, mental health centres or specific disease registries.
- Researchers from public institutions can request data access if they comply with certain requirements. Further information is available online (www.sidiap.org).

people, of whom 5 801 280 (72.2%) were still active as of 30 June 2021, 1 545 850 (19.2%) had been transferred out of the database (i.e. individuals who had moved out of the catchment area of SIDIAP) and 689 818 (8.6%) had died. Individuals are automatically incorporated into SIDIAP if they are registered in the public health system and have been assigned to a primary care centre of the ICS. The only requirement to do the self-registration in the public health system is to live in Catalonia (based on a census certificate). The registration process is free of charge and can be done online (without having to go to a primary care centre) or in person at a primary care centre. For births that take place in public healthcare facilities, the facility registers the newborn in the public health system. Individuals can subsequently leave the database when they move out of the catchment area (based on the census certificate) of SIDIAP or die. The median follow-up time of the population is 15.2 [interquartile range (IQR): 6.2–15.5] years (Table 1).

The current SIDIAP population (as of 30 June 2021) has a balanced sex distribution (50.7% are female) and a median age of 44 years (IQR: 25–60). The sex and age distribution of the SIDIAP population is similar to that of the general population in Catalonia (Figure 2). The large majority of the SIDIAP population is of Spanish nationality (83.9%), lives in urban areas (88.4%) and resides in the Barcelona region (75.3%) (Table 1). Interestingly, whereas the majority of the SIDIAP population resides in the Barcelona region, as seen in Figure 1, SIDIAP has a population coverage of ≤24% for several Basic Health Areas of Barcelona City.

Data collected

SIDIAP is a dynamic database containing pseudo-anonymized data recorded in primary care centres (e.g. disease diagnoses, lifestyle information, clinical parameters, etc.) on a daily basis. It also contains external information related to the primary care visit such as pharmacy dispensations and results of laboratory tests, among others (Table 2). Although SIDIAP systematically collects data since 2006, information prior to this date is also available due to professionals recording data retrospectively and to the data transferred from paper to the EHRs in certain centres during the computerization process. The database is updated every 6 months and is structured in data domains, each containing the person's pseudo-anonymized identifier, which allows linkage between them. Although the number of available data domains grows over time, a description of those most widely used is provided in Table 2.

SIDIAP includes socio-demographic characteristics of the population such as the date of birth (only month and year can be provided to avoid re-identification), sex, nationality, type of residential area (rural or urban), dates of entry and exit (if applicable) and the status at the moment of the data extraction (active, transferred out of SIDIAP or dead). Socio-economic status is captured through individual and ecological indicators. The individual income level (<18 000€, between 18 000€ and 100 000€, >100 000€ per year) and type of occupation (active, retired) are obtained through the pharmaceutical co-payment information.[4] Social class based on occupation is also available for those individuals who have taken sick leave at least once since 2014. The *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales* (MEDEA) deprivation index measures socio-economic status at the census tract level of both the residence and the primary care centre.[5] In addition, the *Índice de Privación* of 2011

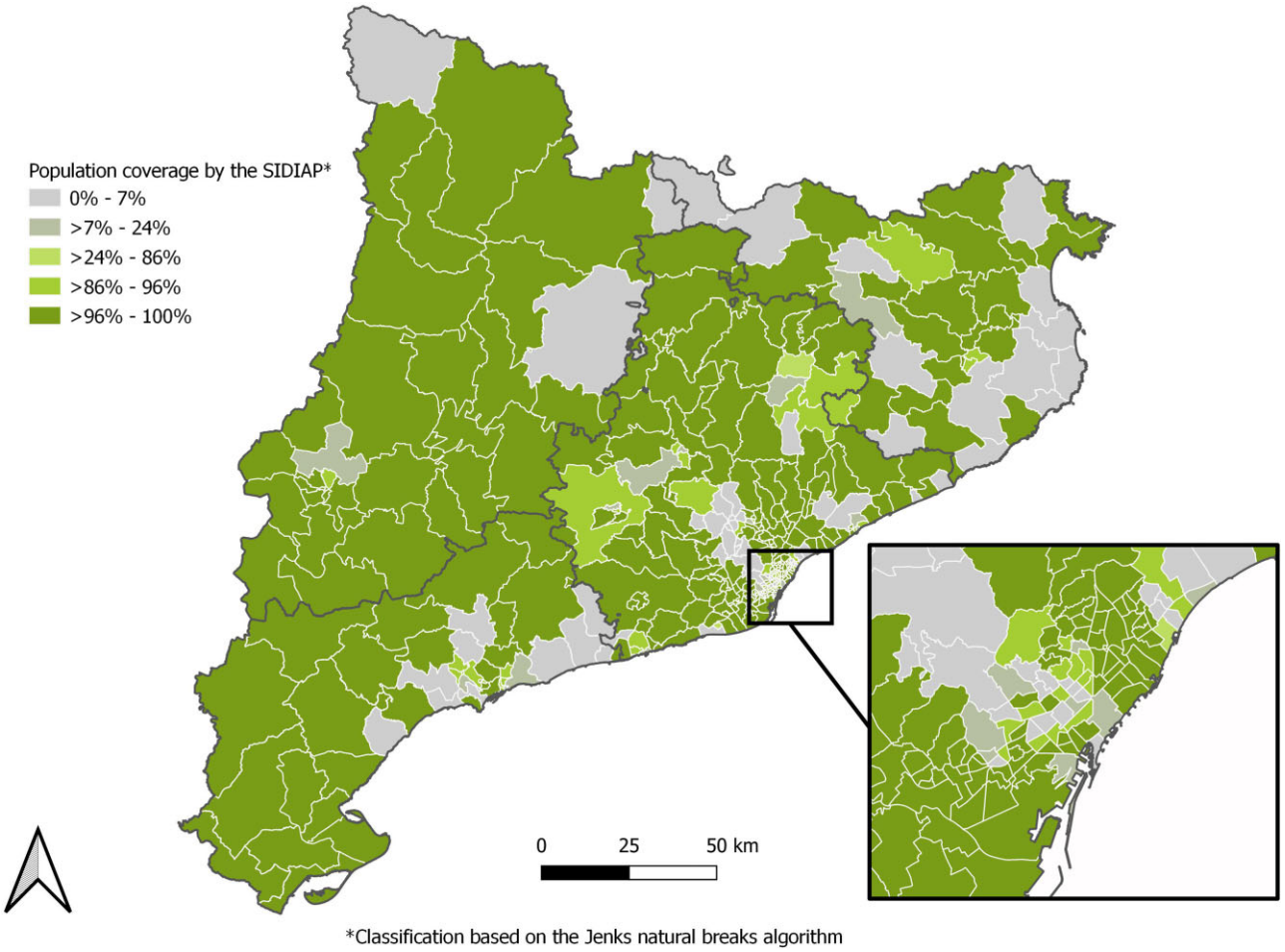


Figure 1 Population coverage by the Information System for Research in Primary Care database by Basic Health Area on 30 June 2021

Basic Health Areas are the territorial units by which primary healthcare services are organized in Catalonia. This delimitation is determined by the population's accessibility to health services, the efficiency of the organization of health resources and other factors (geographical, demographical, social and epidemiological).

(IP2011) is available at the residential census tract level and the *Índice de socioeconómico compuesto* (ISC) is calculated at the primary care centre coverage area level.[6,7]

Health conditions are captured via diagnoses registered by healthcare professionals using the International Classification of Diseases (ICD) codification system (dates of beginning and end of diagnosis given by a GP can be obtained). Currently, the Tenth Revision, Clinical Modification version of the ICD-10 is being used.

The database also contains comprehensive information regarding prescriptions and dispensations of medications. This includes the drugs (dosage and drug units per day) prescribed by ICS healthcare professionals (mostly GPs although specialists can also initiate a prescription for chronic medications that are continued by GPs in the mid-term and long term) that are financed by the Spanish National Health System and dispensed in community pharmacies (number of drug packages dispensed per month). For each drug, the corresponding code from the Anatomical Therapeutic Chemical (ATC) Classification System, defined daily dose recommended by the World Health Organization, the strength, the number of units per package and the administration route are available.

Data on therapeutic and requested procedures, physical examination results, routine measurements and laboratory tests are also captured. Therapeutic procedures include vaccinations (e.g. antigen and the number of administered doses) and health counselling information. Requested procedures comprise diagnostic imaging (e.g. echography, radiology, etc.), tests and scales (e.g. cognitive, pain, mental health, etc.) used in primary care, as well as other cardiovascular, digestive and respiratory diagnostic procedures (e.g. spirometry results, etc.). Physical examination results and routine measurements refer to blood pressure, weight, height, body mass index (BMI), measurements related to child growth and >500 other parameters (e.g. heart rate, cardiovascular risk calculator 'REGICOR', etc.). Laboratory tests include information such as cell count,

Table 1 Socio-demographic characteristics of the Information System for Research in Primary Care population, all-time and on 30 June 2021

	SIDIAP population	
	Total since 2006	Current (as of 30 June 2021)
Persons, *n*	8 036 948	5 801 280
Coverage of the population of Catalonia, %	–	74.9
Follow-up in years, median (IQR)	15.2 (6.2–15.5)	15.5 (12.7–15.5)
Sex, *n* (%)		
Female	4 029 112 (50.1)	2 940 521 (50.7)
Male	4 007 836 (49.9)	2 860 759 (49.3)
Age in years, median (IQR)	44 (27–63)	44 (25–60)
Age in years, *n* (%)		
<18	1 245 702 (15.5)	1 009 303 (17.4)
18–64	4 912 030 (61.1)	3 676 712 (63.4)
>64	1 879 216 (23.4)	1 115 265 (19.2)
Geographic region of nationality, *n* (%)		
Spain	6 476 556 (80.6)	4 867 045 (83.9)
Europe (other than Spain)	409 026 (5.1)	241 816 (4.2)
America	500 886 (6.2)	278 905 (4.8)
Asia	252 291 (3.1)	156 298 (2.7)
Africa	397 178 (4.9)	256 687 (4.4)
Oceania	1011 (0.0)	529 (0.0)
Type of residential area, *n* (%)		
Urban	6 652 818 (82.8)	5 125 779 (88.4)
Rural	433 639 (5.4)	332 827 (5.7)
Missing	950 491 (11.8)	342 674 (5.9)
Catalan regions, *n* (%)		
Barcelona	5 733 291 (71.3)	4 365 572 (75.3)
Lleida	518 810 (6.5)	390 365 (6.7)
Girona	667 816 (8.3)	512 985 (8.8)
Tarragona	706 192 (8.8)	527 997 (9.1)
Missing	410 839 (5.1)	4361 (0.1)
Status on 30 June 2021		
Active	5 801 280 (72.2)	5 801 280 (100)
Transferred out	1 545 850 (19.2)	–
Death	689 818 (8.6)	–

IQR, interquartile range.

serology and biochemistry, among others, that are collected in each laboratory and automatically integrated into the individual's EHR.

SIDIAP also contains lifestyle information. The most widely used indicators include smoking status (categorized into never, former or current smoker) and alcohol intake risk (categorized into no risk, low risk or high risk). The latter is calculated based on the reported amount and the frequency of consumption of alcoholic drinks (e.g. on a daily basis), the type of alcoholic drink and/or whether the consumption is made in risky situations (e.g. pregnancy). This information is converted into standard units of alcohol ingested on a weekly basis and converted into levels of alcohol consumption.

Data regarding primary care visits are available, including the date of the visit and the type of professional consulted as well as the cause and date of referral to specialists.

The database includes detailed pregnancy information such as dates of last period and of estimated delivery, along with the type of delivery, the circumstance of the end of the pregnancy (e.g. type of delivery, abortion, etc.), gestational age and trimestral obstetric ultrasounds, among others. SIDIAP contains information about paediatric (<15 years of age) health (e.g. nutrition, development, screening tests, etc.), collected under the framework of the *Programa de infancia amb salut* (Childhood and Health Program).[8] In addition, parent–child linkage is available

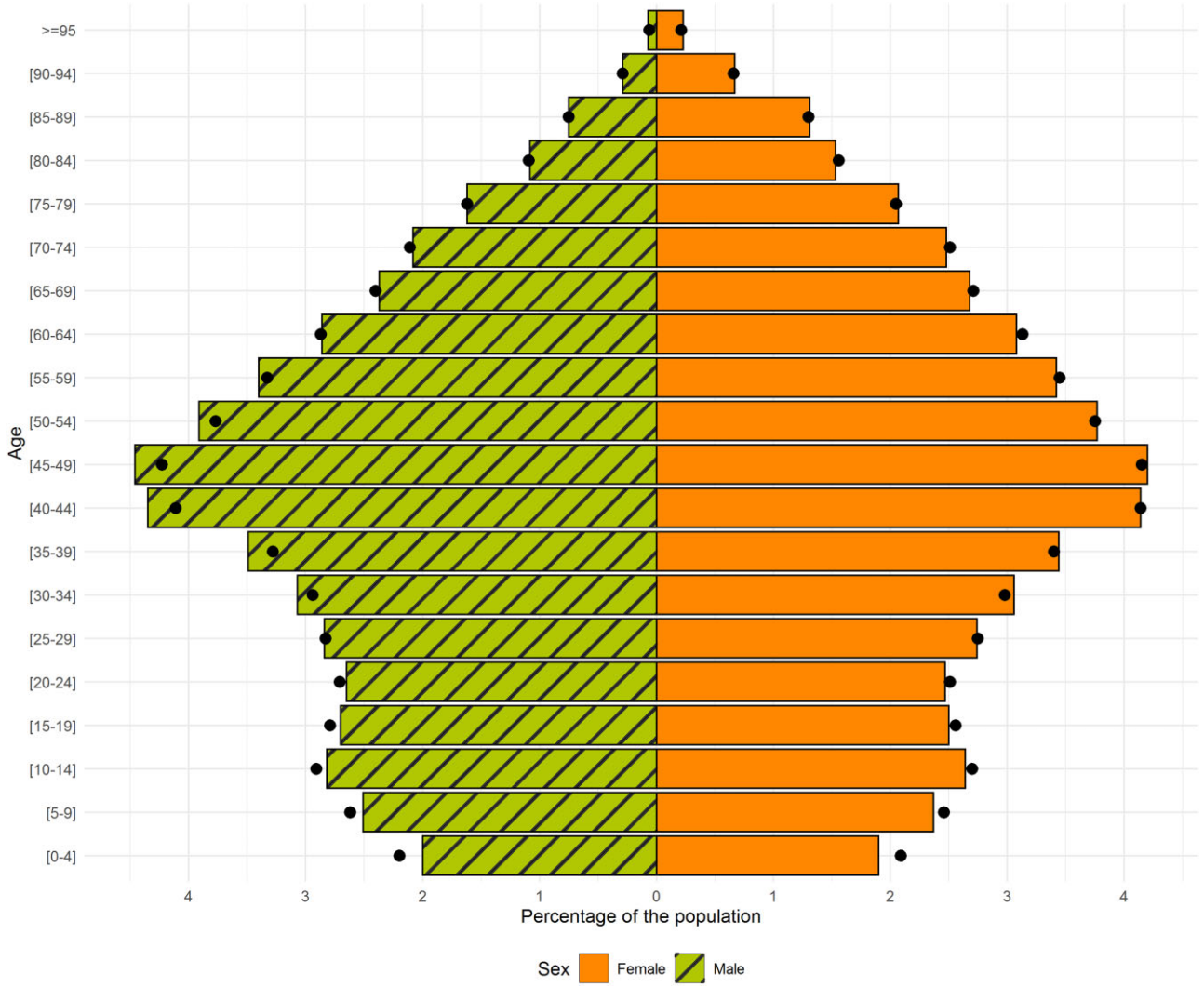


Figure 2 Age and sex distribution in the current Information System for Research in Primary Care population and in the general population of Catalonia

The Information System for Research in Primary Care data used for this graph was extracted on 30 June 2021. The data of the population of Catalonia were obtained from the Instituto Nacional de Estadística (National Institute of Statistics) website for the year 2020, 'Población por comunidades, edad (grupos quinquenales), Españoles/Extranjeros, Sexo y Año' tab, available from https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02002.px&L=0.

for children and adolescents born or entering the database after 2006.

SIDIAP continues to incorporate new information into the database when needed (e.g. to answer new research questions or to monitor more closely a specific condition or disease, etc.) and possible. For instance, during the coronavirus disease 2019 (COVID-19) pandemic, SIDIAP incorporated additional information needed to investigate this disease (e.g. polymerase chain reaction test results, administered vaccines, etc.) in a timely fashion.

Free text that has been previously anonymized is available when sufficient detail cannot be obtained from the structured data. Further information to complement the structured data or to validate diagnoses needed for research can also be obtained through surveys sent to health professionals administered by the ICS.

The growth in the recording of information in SIDIAP over time is shown in Figure 3a and b. For example, in 2019, ~80% of the SIDIAP population had at least one visit to primary care and >60% had one clinical diagnosis and/or a prescription/dispensation for a medication (Figure 3a). A decrease in the amount of recorded information can be observed in 2020 (likely due to the COVID-19 pandemic). By 2019, 75% of the population had at least one record available of blood pressure and >60% had a record of alcohol intake, BMI, glucose, total cholesterol and/or smoking status (Figure 3b).

Table 2 Description of key data available in the Information System for Research in Primary Care database

Type of data collected	Data domain in the SIDIAP database	Number of people with at least one entry for the data domain	Total number of entries for the data domain	Key information recorded in the data domain
Socio-demographics	Population	8 036 948	8 036 948	Date of birth, sex, date of entry to the database, date and cause of exit (death, transferral out of the database) if applicable, nationality (grouped into 11 categories)
	Socio-economic variables	8 036 948	8 036 948	Individual income level (<18 000€, between 18 000€ and 100 000€, >100 000€ per year), deprivation indices measured at the census tract level (MEDEA[a], IP2011[b]) and the basic health area level (ISC)[c], requested pharmaceutical contribution (exempted, or contribution of 10%, 40%, 50%, 60%), maximum monthly pharmaceutical contribution (without limit, or up to 8.23€, 18.52€ or 61.75€)
	Regional variables	8 036 948	18 036 948	Type of residential area (rural[d] or urban), province (Barcelona, Tarragona, Girona, Lleida) and basic health area ($n = 388$) where the person resides; productive unit (333 categories, e.g. EAP Vallirana, EAP Salt, EAP La Garriga, etc.), sanitary scope (10 categories, e.g. Barcelona, Girona, Terres del Ebre, etc.).
	Complexity	8 036 948	8 036 948	Clinical risk group (indicator based on the person's co-morbidities), state of fragility (categorizes persons into 'with chronic complexity', 'with advanced chronic disease' or none) and whether the person lives in a nursing home (yes, no)
Health conditions	Primary care diagnoses	7 531 524	171 215 580	ICD-10 code, dates of start and end (if there is any), SIDIAP grouper (e.g. 'Arthrosis' includes two ICD-10 codes with its descendants) and type of productive unit (hospital, primary care team, mental health centre, etc.) that registers the health problem
	Sick, maternity or paternity leaves	3 191 357	16 651 062	Coding of the health problem causing the sick leave (ICD-10) and dates of start and end of leave
Medications and vaccines	Prescriptions	7 083 845	320 380 795	ATC code, treatment group (e.g. anxiety, hormonal replacement therapy, hypertension), dates of start and end of prescription, posology, frequency of intake and DDD of the medication, speciality of the health professional (e.g. general medicine, gynaecology, paediatrician) and setting (hospital, primary care team, mental health centre, etc.) of prescription
	Dispensations	6 928 471	1 082 492 571	ATC code, dates of start and end of the dispensation, DDD, number of packages dispensed per month
	Adverse reactions	240 047	316 214	ATC code of the drug that produces the adverse reaction and date of occurrence
	Vaccinations	6 443 204	116 615 695	Code, date and dose of administration and grouper of the antigen of the vaccine

(Continued)

Table 2 Continued

Type of data collected	Data domain in the SIDIAP database	Number of people with at least one entry for the data domain	Total number of entries for the data domain	Key information recorded in the data domain
Laboratory tests	Analytical variables	5 703 343	1 129 564 883	Biomarker/measurement (e.g. glucose, cholesterol, bilirubin, PCR), date of test/measurement, result value and unit of the test, and interpretation of the value (in four categories: positive, negative, indeterminate, inconclusive or through free text if applicable)
	Serology	2 051 806	11 170 439	Serological test (e.g. hepatitis C antibodies, HIV) date and result (i.e. positive, negative, indeterminate)
Clinical practice and lifestyle information	Clinical and lifestyle variables	6 545 968	513 637 096	Clinical measurement (e.g. systolic/diastolic blood pressure, weight, height) and lifestyle variables (e.g. alcohol consumption, smoking status) date and value of the registry
	Request of complementary explorations/tests	4 314 375	21 741 943	Type (e.g. mammography, colonoscopy) and date of the exploration request
Visits	Visits	7 473 119	695 049 273	Visited service, SIDIAP grouper (e.g. dermatology, nursing, general practice), type of productive unit (hospital, primary care team, mental health centre, etc.) that does the visit, place (online, in person at the health centre or at the residence of the person) where the visit takes place
	Referrals	4 738 748	22 042 118	Health service to which the person is being referred to, date and cause (usually suspicion of an ICD-10 code) of referral
Sexual and reproductive health	Variables of sexual and reproductive health assistance	1 566 025	172 218 876	Variable (e.g. year of menopause, breastfeeding, or result of the last mammography), date and value (if none, the result is reported in another variable containing text in free format) of the registry
	Pregnancy	427 193	649 630	Date of last period and of estimated delivery, type of delivery (e.g. voluntary miscarriage, natural miscarriage, c-section), gestational age (in weeks), miscarriage risk (none, low, normal, moderate, high, very high), number of foetuses, trimestral obstetric ultrasounds [containing information about the mother (type of exploration, location of the placenta, type of amniotic fluid, etc.) and the foetus(es) (sex, cardiac activity, cephalic circumference, etc.)]
Paediatric health (<15 years of age)	Paediatric health	1 592 446	263 953 678	Variables related to birth (e.g. gestational age at delivery, sex, weight, height), screening (e.g. neonatal deafness, cystic fibrosis, congenital hypothyroidism), development (e.g. reaction to external stimuli, object manipulation, sphincter control), nutrition (e.g. gluten intolerance, breastfeeding length, beikost consumption), school (e.g. course, integration, performance), hygiene (e.g. correct fingernail, oral and genital hygiene), leisure and sports (e.g. extracurricular and weekend activities, screen time) and sleeping patterns (e.g. dyssomnia, sleeping schedule, snoring)

(Continued)

Table 2 Continued

Type of data collected	Data domain in the SIDIAP database	Number of people with at least one entry for the data domain	Total number of entries for the data domain	Key information recorded in the data domain
Other data sources available through external linkage	Diagnoses and medical procedures at hospital discharge[e]	3 223 390	59 748 073	ICD-10-CM/PCS code, SIDIAP grouper, position of the diagnosis/procedure at the time of admission (used to prioritize the diagnoses/procedures that caused the admission), dates and circumstance (urgent, scheduled) of admission, and dates and circumstance (eight categories, including home, voluntary discharge or death) of discharge. From 2016 onwards, type of anaesthesia used, ICU admission and length of it, date of the first ICU admission
	Hospital medication for outpatient dispensing[e]	516 557	14 563 331	ATC code, date of dispensation, content per box (tablets, syringes, suppositories, etc.), code of the pharmaceutical product, reason for the dispensation, price of the dispensation

[a]Deprivation index (based on five indicators related to work, education, housing conditions) calculated at the census tract level and available for urban areas.
[b]Deprivation index available at the residential census tract level based on six indicators of employment and education for urban and rural areas.
[c]Deprivation index calculated at the basic health area level for the assignment of the budgets of the primary healthcare teams in Catalonia valid for urban and rural areas.
[d]Rural areas are defined as municipalities with a population density of <100 people per km^2 and/or a population of <30 000 inhabitants.
[e]Information recorded in all Catalan public hospitals registered in the minimum basic set of hospital discharge data (CMBD-AH) available through linkage to the Data Analysis Program for Health Research and Innovation (PADRIS) of the Department of Health.
ATC, Anatomical Therapeutic Chemical; DDD, defined daily dose; GP, general practitioner; ICD-10, International Classification of Diseases, 10th Revision (CM: Clinical Modification or PCS: Procedure Coding System); ICU, intensive care unit; IP2011, Deprivation Index 2011; ISC, Composed Socio-economic Index; HIV, human immunodeficiency virus; MEDEA, *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales*; PCR, polymerase chain reaction.

Linkage to other data sources

SIDIAP is a pseudo-anonymized database and does not contain individual personal data. Nevertheless, it can be linked to other data sources on a project-by-project basis through a Trusted Third Party (TTP) using the individuals' unique personal identifier.

The information recorded in all Catalan public hospitals is registered in the minimum basic set of hospital discharge data (CMBD-AH) and is linked to SIDIAP through the *Programa d'analítica de dades per a la recerca i la innovació en salut* (PADRIS, Data Analysis Program for Health Research and Innovation) of the Catalan Department of Health.[9,10] This linkage has been widely used for SIDIAP research and includes the date and cause of hospitalization and discharge, as well as the codes registered during the stay (in ICD-10-CM and ICD-10-PCS, respectively).[11–19] Data from psychiatric hospitals, outpatient centres of mental health, dispensed medication in hospital settings and emergency rooms can also be obtained through the same linkage process.[10] In addition, SIDIAP has been previously linked to disease registries of cancer, arthroplasties, dementia, kidney transplants and dialysis, among others.[13–17,20,21] Finally, linkage to urban environment indicators (air pollution, noise, green spaces and built environment) at the census tract level[22] and to external cohorts at the individual level have also been conducted. An example of the latter is the population-based prospective peripheral arterial disease study (ARTPER) cohort that includes 3786 individuals aged >49 years recruited in 28 primary care centres of Catalonia through random sampling. The participants were given an appointment for an interview, a blood sample extraction and a visit at which anthropometric indicators were measured (including the ankle arm index examination). The collected data were used to estimate the prevalence and associated risk factors of peripheral arterial disease in the general population.[16]

Data quality

Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of a specific information under different codes). The measurement units of variables measuring one

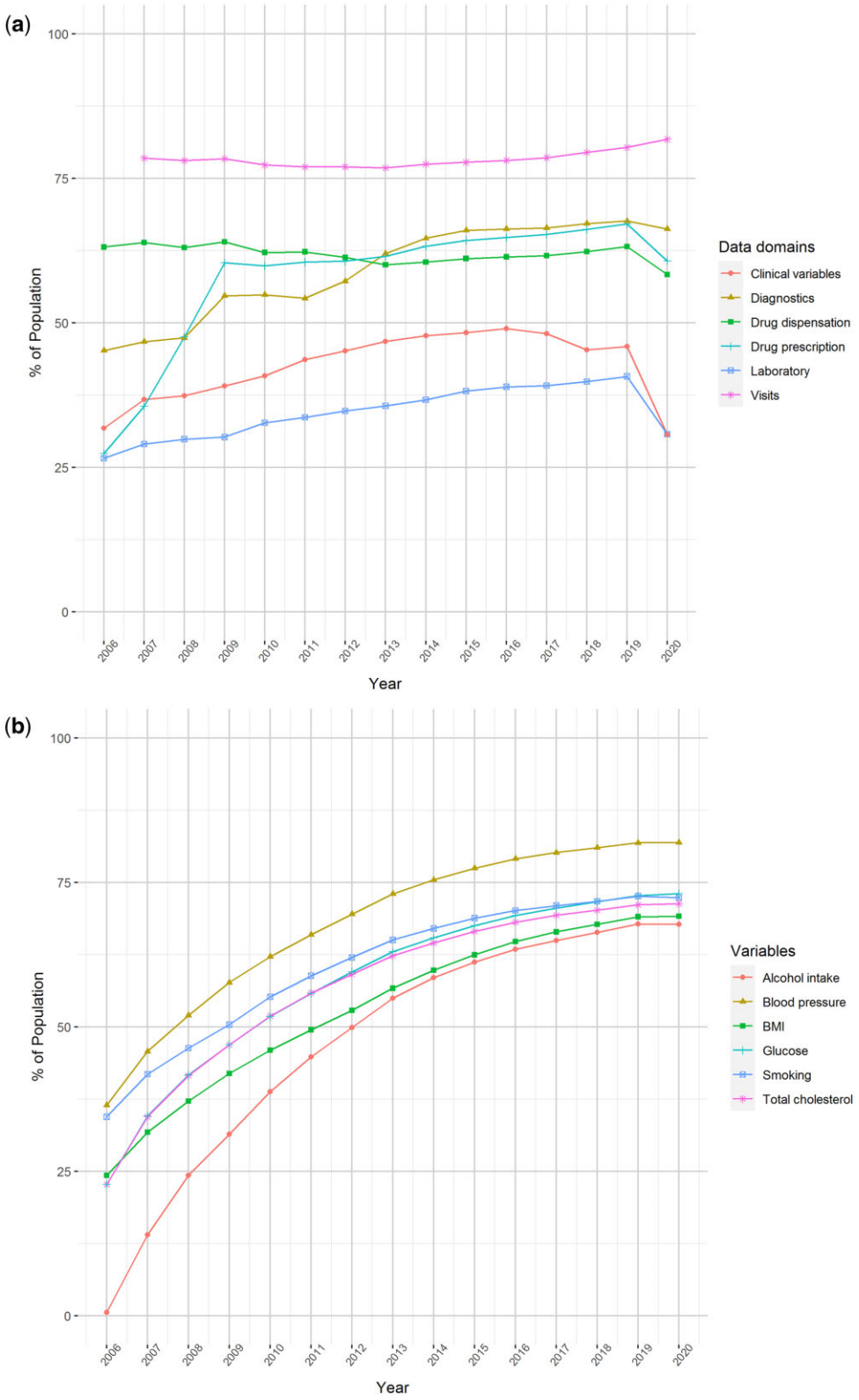


Figure 3 Recording of key variables and data domains in the Information System for Research in Primary Care population by calendar year

(a) Proportion of the Information System for Research in Primary Care population with yearly registries of key data domains, by calendar year. (b) Proportion of the population with at least one recording of key variables by calendar year. The data domain 'visits' of Figure 3a include all kinds of visits (in person at the primary care centre or at home and telematic). The clinical variables domain shown in Figure 3b contains information about data collected in primary care visits (BMI, blood pressure, smoking status, alcohol intake, etc.). BMI, body mass index.

characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g. there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinality and reliability), plausibility (range checks and unusual values) and consistency are inspected through visualization tools. In addition, before having access to the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences and prevalences of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery.

External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text or by sending questionnaires to health professionals. The quality of a wide number of data captured in SIDIAP (e.g. cancer, Alzheimer's disease, dementia, cardiovascular risk factors and musculoskeletal disorders) has been demonstrated in validation studies.[13–16,20,21,23,24]

Data resource use

SIDIAP data have been extensively used by national and international institutions to generate real-world evidence. A non-exhaustive list of 223 peer-reviewed published articles and of 306 projects (of which 37 are still ongoing) using the SIDIAP database is available on the SIDIAP website (www.sidiap.org, 'Projects' and 'Dissemination' tabs). These publications cover a wide range of research topics such as cardiovascular diseases, diabetes, musculoskeletal disorders, respiratory problems, cancer, mental health, multimorbidity, COVID-19, vaccinations; and research areas including pharmacoepidemiology, evaluation of safety and comparative effectiveness research, characterization of a disease, drug utilization, temporal trends of disease, health economics and evaluation of healthcare services, among others.[11,18,19,25–30]

Strengths and weaknesses

Strengths

SIDIAP has several strengths. First, the database is representative of the population of Catalonia in terms of age, sex and geographic distribution (Figures 1 and 2). This favours the generalizability of the findings of the studies conducted using SIDIAP to the general population living in Catalonia but also to other comparable regions. Second, due to SIDIAP's large size, this database can be used to answer research questions that would not be feasible in smaller-sized data sets. Third, the diverse type of data encompassed by this database is also an asset. Not only does SIDIAP include data typically recorded in EHRs (e.g. clinical diagnoses) but also contains socio-demographic information (e.g. socio-economic status or nationality) and lifestyle information (e.g. smoking status or alcohol intake). The parent–child linkage is also a major strength as it allows one to study the impact of parental health and early life exposures on health outcomes during childhood. Furthermore, SIDIAP contains data from external sources such as biomarkers' information originating from laboratories or prescription and dispensation of drugs, which makes the assessment of drug exposure quite complete. Data from different settings (e.g. disease and hospitalization registries) can also be obtained through diverse linkages, enriching the data available for studies. Finally, SIDIAP is being mapped to different common data models used in European projects. At present, it has already been mapped to the international Observational Medical Outcomes Partnership–Common Data Model (OMOP-CDM), which facilitates and promotes multi-database studies, helps with data management and data analyses, and ensures confidentiality throughout the studies using a federated analysis approach.[31]

Weaknesses

The SIDIAP database also has weaknesses. Although the database is representative of the population living in Catalonia and regions with similar socio-demographics, it is not necessarily so of other regions of Spain or other countries. In addition, data missingness is a common issue of EHRs (e.g. BMI is not recorded for every participant in the database, as seen in Figure 3b) and a recent measurement of a variable of interest might not be available at the index date for a particular study (e.g. the last BMI measurement available might have been recorded years before the index date). However, methodological approaches such as multiple imputations can be implemented to reduce collider bias in research studies.[32] Under-reporting of certain variables is also a limitation that can lead to the

underestimation of the frequency of a certain exposure or condition (e.g. less severe behavioural or mental disorders might be more likely to go undiagnosed in clinical practice). Furthermore, individual validation of a complete list of events of interest, as conducted ad hoc in cohort or case–control studies, is not possible for large EHR databases and may lead to misclassification. However, algorithms to capture diseases or conditions can be tested in validation studies and allow the quantification of the data quality. Also, relevant information for research might be recorded in unstructured format (i.e. free text) by health professionals. Although advanced techniques to process these data are not yet available in SIDIAP, previously anonymized free text can be manually explored by researchers. Another limitation refers to clinical practice standards and coding that can change over time, giving rise to observed changes in the incidence of a certain condition that might be unrelated to its epidemiology. Finally, due to the primary care nature of this database, studies conducted with SIDIAP could lack the granularity to answer certain research questions. For instance, specialist prescribing, drugs administered in the hospital setting, drugs purchased over the counter and actual drug intake are not available in the database.

Data resource access

Any researcher is able to request SIDIAP data to conduct a study. A five-step procedure takes place before data access is granted: (i) the researcher(s) must send an application (standardized form available at www.sidiap.org and study protocol) to the SIDIAP team; (ii) the application is approved by SIDIAP's Scientific Committee which evaluates the scientific quality and feasibility of the proposal; (iii) the study protocol is approved by the Clinical Research Ethics Committee of IDIAPJGol; (iv) the principal investigator of the study must sign a Good Practice form and, in some cases, an agreement between parties is needed; and (v) a meeting between the research team and the SIDIAP team is arranged to discuss the procedures and set the data extraction. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudesen/application-procedure) or by contacting Anna Moleras (sidiap@idiapjgol.org). Data access is limited to researchers from public organizations and collaboration with private institutions is possible when a study is required by a regulatory agency or for non-commercial studies within a European project financed by the European Commission.

In accordance with current European and national law, the data used in this study are only available for the researchers participating in this study. Thus, we are not allowed to distribute or make publicly available the data to other parties.

Ethics approval

The use of the data included in the Information System for Research in Primary Care (SIDIAP) is authorized by the Catalan Health Institute (ICS) and Data Analysis Program for Health Research and Innovation (PADRIS) who ensure the pseudo-anonymization of the information. When linkage with other public data sources is required, ICS or PADRIS act as a Trusted Third Party (TTP) to execute the linkage and provide the new data set already pseudo-anonymized; otherwise, informed consent of patients is needed to access their personal data, using the same TTP. SIDIAP does not provide information subject to re-identification and aggregations or deletions are applied in order to protect pseudo-anonymization. The data are managed in a secure server following all the present legal requirements of the General Data Protection Regulation (European Union) 2016/679 and of the Council of 27 April 2016 and Organic Law 3/2018 of 5 December on the protection of personal data and guarantee of Digital Rights.

This study was exempted from the approval of the Clinical Research Ethics Committee of the IDIAPJGol given that the data were directly analysed in the SIDIAP platform and only aggregated results were reported.

Data availability

See Data Resource access above.

Author contributions

All authors were involved in the study design. M.R. wrote the first draft of the manuscript. C.R. performed the data analyses. M.R., C.R., M.A. and T.D.S. prepared the tables and figures. All authors interpreted the results, contributed to drafting the article and approved the final version of the manuscript.

Funding

The SIDIAP database creation, development and maintenance are currently funded by the IDIAPJGol that seeks to recoup part of its expenditures through research fees obtained from research projects. SIDIAP has obtained funding for studies from the National Institute of Health Carlos III, the Catalan Department of Health, the Innovative Medicines Initiative, the European Medicines Agency, the European Commission, as well as several universities, hospitals and national and international research organizations.

In the context of this study, M.R. was funded by Wereld Kanker Onderzoek Fonds, as part of the World Cancer Research Fund International grant programme (grant number: 2017/1630). T.D.S. acknowledges receiving financial support from the Instituto de Salud Carlos III (ISCIII; Miguel Servet 2021: CP21/00023). The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript.

Acknowledgements

This study was carried out as part of the Doctoral Program in Biomedical Research Methodology and Public Health at the Autonomous University of Barcelona. We would like to thank all healthcare professionals of Catalonia who daily register information in the populations' EHRs. We also want to thank all the research teams who have worked with SIDIAP and have helped to improve and maintain the database, and the *Institut Català de la Salut* (ICS) and the *Programa d'analítica de dades per a la recerca i la innovació en salut* (PADRIS) for providing access to the different data sources accessible through SIDIAP.

Conflict of interest

None declared.

References

1. Ministry of Health, Social Services, and Equality. National Health System of Spain. 2012. https://www.sanidad.gob.es/gl/organizacion/sns/docs/sns2012/SNS012__Ingles.pdf (28 October 2021, date last accessed).
2. Generalitat de C, Població de referència del Servei Català de la Salut per a l'any 2020. 2020. https://catsalut.gencat.cat/web/.content/minisite/catsalut/proveidors_professionals/registres_catalegs/documents/informe-poblacio-referencia-2021.pdf (28 October 2021, date last accessed).
3. Bolíbar B, Fina Avilés F, Morros R *et al.*; Grupo SIDIAP. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Med Clin (Barc)* 2012;**138**:617–21.
4. CatSalut. Servei Català de la Salut. El model de copagament farmacèutic. 2020. https://catsalut.gencat.cat/ca/serveis-sanitaris/atencio-farmaceutica/financament-public-medicaments/model-copagament/ (28 October 2021, date last accessed).
5. Domínguez-Berjón MF, Borrell C, Cano-Serral G *et al.* Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). *Gac Sanit* 2008;**22**:179–87.
6. Duque I, Domínguez-Berjón MF, Cebrecos A *et al.*; en nombre del Grupo de Determinantes Sociales de la Salud, iniciativa contexto de la Sociedad Española de Epidemiología. Índice de privación en España por sección censal en 2011. *Gac Sanit* 2021;**35**:113–22.
7. Colls C, Mias M, García-Altés A. Un índice de privación para reformar el modelo de financiación de la atención primaria en Cataluña. *Gac Sanit* 2020;**34**:44–50.
8. Agència de Salut Pública de Catalunya (ASPCAT). Programa Infància amb Salut. 2018. https://salutpublica.gencat.cat/ca/ambits/promocio_salut/Infancia-i-adolescencia/Infancia/infancia-amb-salut/ (28 October 2021, date last accessed).
9. CatSalut. Servei Català de la Salut. Conjunt mínim bàsic de dades (CMBD). 2017. https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/ (28 October 2021, date last accessed).
10. Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS). Programa d'analítica de dades per a la recerca i la innovació en salut. 2020. https://aquas.gencat.cat/ca/ambits/analitica-dades/padris/index.html (28 October 2021, date last accessed).
11. Recalde M, Davila-Batista V, Díaz Y *et al.* Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain. *BMC Med* 2021;**19**:10.
12. Garcia-Gil M, Comas-Cufí M, Blanch J *et al.* Effectiveness of statins as primary prevention in people with different cardiovascular risk: a population-based cohort study. *Clin Pharmacol Ther* 2018;**104**:719–32.
13. Recalde M, Manzano-Salgado C, Díaz Y *et al.* Validation of cancer diagnoses in electronic health records: results from the information system for research in primary care (SIDIAP) in northeast Spain. *Clin Epidemiol* 2019;**11**:1015–24.
14. Ponjoan A, Garre-Olmo J, Blanch J *et al.* Epidemiology of dementia: prevalence and incidence estimates using validated electronic health records from primary care. *Clin Epidemiol* 2019;**11**:217–28.
15. Ramos R, Balló E, Marrugat J *et al.* Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol (Engl Ed)* 2012;**65**:29–37.
16. Pagès-Castellà A, Carbonell-Abella C, Avilés FF *et al.* Burden of osteoporotic fractures in primary health care in Catalonia (Spain): a population-based study. *BMC Musculoskelet Disord* 2012;**13**:79.
17. Robinson DE, Ali MS, Pallares N *et al.* Safety of oral bisphosphonates in moderate-to-severe chronic kidney disease: a binational cohort analysis. *J Bone Miner Res* 2021;**36**:820–32.
18. Ramos R, Comas-Cufí M, Martí-Lluch R *et al.* Statins for primary prevention of cardiovascular events and mortality in old and very old adults with and without type 2 diabetes: retrospective cohort study. *BMJ* 2018;**362**:k3359.
19. Burn E, Tebé C, Fernandez-Bertolin S *et al.* The natural history of symptomatic COVID-19 during the first wave in Catalonia. *Nat Commun* 2021;**12**:777.
20. Prieto-Alhambra D, Judge A, Javaid M *et al.* Predictors of poor outcome (high use of non-steroidal anti-inflammatory drugs) at year 1 following total knee/hip arthroplasty: the press-up cohort study. *Osteoarthritis and Cartilage* 2013;**21**:S155.
21. Ponjoan A, Garre-Olmo J, Blanch J *et al.* How well can electronic health records from primary care identify Alzheimer's disease cases? *Clin Epidemiol* 2019;**11**:509–18.
22. de Bont J, Díaz Y, de Castro M *et al.* Ambient air pollution and the development of overweight and obesity in children: a large longitudinal study. *Int J Obes* 2021;**45**:1124–32.
23. Fina-Aviles F, Medina-Peralta M, Mendez-Boo L *et al.* The descriptive epidemiology of rheumatoid arthritis in Catalonia: a retrospective study using routinely collected data. *Clin Rheumatol* 2016;**35**:751–57.
24. Muñoz-Ortego J, Vestergaard P, Rubio JB *et al.* Ankylosing spondylitis is associated with an increased risk of vertebral and nonvertebral clinical fractures: a population-based cohort study. *J Bone Miner Res* 2014;**29**:1770–76.
25. Mata-Cases M, Franch-Nadal J, Real J *et al.* Therapeutic inertia in patients treated with two or more antidiabetics in primary care: factors predicting intensification of treatment. *Diabetes Obes Metab* 2018;**20**:103–12.

26. Lane JCE, Butler KL, Poveda-Marina JL *et al.* Preschool obesity is associated with an increased risk of childhood fracture: a longitudinal cohort study of 466,997 children and up to 11 years of follow-up in Catalonia, Spain. *J Bone Miner Res* 2020;**35**:1022–30.
27. Monteagudo M, Nuñez A, Solntseva I *et al.* Treatment pathways before and after triple therapy in COPD: a population-based study in primary care in Spain. *Arch Bronconeumol (Engl Ed)* 2021;**57**:205–13.
28. Ortega Y, Aragonès E, Piñol JL, Basora J, Araujo A, Cabré JJ. Impact of depression and/or anxiety on the presentation of cardiovascular events in a cohort with metabolic syndrome. StreX project: five years of follow-up. *Prim Care Diabetes* 2018;**12**:163–71.
29. Troncoso-Mariño A, Roso-Llorach A, López-Jiménez T *et al.* Medication-related problems in older people with multimorbidity in catalonia: a real-world data study with 5 years' follow-up. *J Clin Med* 2021;**10**:https://doi.org/10.3390/jcm10040709.
30. Braeye T, Emborg H-D, Llorente-García A *et al.* Age-specific vaccination coverage estimates for influenza, human papillomavirus and measles containing vaccines from seven population-based healthcare databases from four EU countries—the ADVANCE project. *Vaccine* 2020;**38**:3243–54.
31. Observational Health Data Sciences and Informatics. The Book of OHDSI. 2019. https://ohdsi.github.io/TheBookOfOhdsi/TheBookOfOhdsi.pdf (28 October 2021, date last accessed).
32. Pedersen AB, Mikkelsen EM, Cronin-Fenton D *et al.* Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol* 2017;**9**:157–66.

5.2. Study II

Recalde M, Manzano-Salgado CB, Díaz Y, Puente D, Garcia-Gil MdM, Marcos-Gragera R, Ribes-Puig J, Galceran J, Posso M, Macià F, Duarte-Salles T.

Validation of cancer diagnoses in electronic health records: results from the Information System for Research in Primary Care (SIDIAP) in Northeast Spain

Clinical Epidemiology 2019:11 1015–1024

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain

This article was published in the following Dove Press journal:
Clinical Epidemiology

Martina Recalde,[1,2,*] Cyntia B Manzano-Salgado,[1,2,*] Yesika Díaz,[1] Diana Puente,[1,2] Maria Del Mar Garcia-Gil,[1] Rafael Marcos-Gragera,[3,4] Josefa Ribes-Puig,[5,6] Jaume Galceran,[7] Margarita Posso,[8] Francesc Macià,[8] Talita Duarte-Salles[1]

[1]Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGoL), Barcelona, Spain; [2]Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Cerdanyola del Vallès, Spain; [3]Unitat d'Epidemiologia i Registre de Càncer de Girona (UERCG), Pla Director d'Oncologia, Institut Català d'Oncologia, Institut d'Investigació Biomèdica de Girona (IdIBGi), Universitat De Girona, Girona, Spain; [4]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; [5]Catalan Cancer Plan, Department of Health of Catalonia, Barcelona, Spain; [6]Department of Clinical Sciences, University of Barcelona (UB), Barcelona, Spain; [7]Registre de Càncer de Tarragona, Fundació per a la Investigació i Prevenció del Càncer (FUNCA), IISPV, Reus, Spain; [8]Cancer Prevention Unit and Cancer Registry, Department of Epidemiology and Evaluation, Hospital del Mar, Barcelona, Spain

*These authors contributed equally to this work

Correspondence: Talita Duarte-Salles
Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGoL) Gran Via Corts Catalanes, 587 Àtic, Barcelona 08007, Spain
Tel +34 93 482 43 42
Email tduarte@idiapjgol.org

Background: Electronic health records are becoming an increasingly valuable resource for epidemiology but their data quality needs to be quantified. We aimed to validate twenty-five types of incident cancer cases in the Information System for Research in Primary Care (SIDIAP) in Catalonia with the population-based cancer registries of Girona and Tarragona as the gold-standard.

Methods: We calculated the sensitivity, positive predictive values (PPV), and the time-difference between the date of diagnosis entered into the SIDIAP and into the registries. We added hospital discharge cancer diagnoses to the SIDIAP to assess sensitivity changes.

Results: We identified 27,046 incident cancer diagnoses in the SIDIAP from 2009–2015 among the 949,841 residents of Girona and Tarragona. The cancer types with the highest sensitivity were breast (89%, 95% CI: 88–90%), colorectal (81%, 95% CI: 80–82%), and prostate (81%, 95% CI: 80–83%). Trachea, bronchus and lung cancers had the highest PPV (76%, 95% CI: 74%-78%) followed by stomach (72%, 95% CI: 68–75%) and pancreas (71%, 95% CI: 67–75%). Most cancer diagnoses were reported with less than three months of difference between the SIDIAP and the registries. More cases were registered first in the registries than in the SIDIAP. By adding cancer diagnoses based on hospital discharge data, sensitivity increased for all cancers, especially for gallbladder and biliary tract for which the sensitivity increased by 21%.

Conclusion: The SIDIAP includes 76% of the cancer diagnoses in the cancer registries but includes a considerable number of cases that are not in the registries. The SIDIAP reports most of the cancer diagnoses within a three-month period difference from the date of diagnosis in the cancer registries. Our results support the use of the SIDIAP cancer diagnoses for epidemiological research when cancer is the outcome of interest. We recommend adding hospital discharge data to the SIDIAP to increase data quality, particularly for less frequent cancer types.

Keywords: validation studies, cancer, electronic health records, primary health care, population-based cancer registries

Introduction

Cancer is one of the leading causes of morbidity and mortality worldwide.[1] In 2018, there were 18 million new cases and 9 million deaths.[2] In Spain, cancer is a significant burden for the National Health System: cancer is the second most frequent overall cause of death and results in more than 250,000 new invasive

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/CLEP.S225568

cancer cases every year.[3] Therefore, conducting research focused on understanding cancer epidemiology is important both at the national and international levels.

The use of databases of routinely collected electronic health records (EHRs) is becoming more common in epidemiology and clinical research. Due to their size, amount of data availability, representativeness, and long-term follow-up, EHR databases offer a great opportunity to conduct cancer research.[4] Another advantage of large health record databases is that they provide sufficient statistical power to detect uncommon outcomes such as rare cancer types.[5] However, validation processes are required to quantify the correctness of the data and to increase the reliability of large health record databases for use in subsequent observational studies.[6]

The information recorded in EHRs by primary health care professionals in Catalonia – a region in Northeast Spain with 7.5 million inhabitants (2017) – comprises the Information System for Research in Primary Care (SIDIAP) platform.[7] Since the SIDIAP aims to provide reliable information to support research in primary health care, validation studies are performed regularly.[8] A previous study assessed the validity of lung, colon and rectum, prostate, breast, and cervix uteri cancers in the SIDIAP during the period 2009–2012 with sensitivities ranging from 64% (cervix uteri) to 92% (breast).[9] However, this study compared SIDIAP cancer cases with those from the registry of a single hospital in Barcelona. Although the data collection for this hospital is rigorous for a specific area in Barcelona, this area is not representative of the general population of Catalonia. Furthermore, the hospital does not have data available for research use on hematological cancers. A study validating more cancer types and using population-based cancer registries as the gold-standard may increase the scope of the validity of cancer diagnosis in the SIDIAP as well as its use in new areas of research.

The aim of this study was to validate twenty-five types of incident cancer cases in the SIDIAP using the population-based cancer registries of Girona and Tarragona as the gold-standard and to assess the time-difference in the date of diagnosis between the SIDIAP and these cancer registries.

Methods

Data Sources

We performed a cross-sectional study in the SIDIAP during the years 2009–2015, using data from the two population-based cancer registries that exist in Catalonia, the Girona and Tarragona cancer registries, as the gold-standard. The SIDIAP includes information recorded in EHRs by health professionals during routine visits at 287 primary health care centers from the Institut Català de la Salut (ICS, Catalan Health Institute).[10,11] The SIDIAP has anonymized records for more than seven million people and is representative of the Catalan population in terms of age, sex, and geographic distribution.[11] It includes information on disease diagnoses (International Classification for Diseases, 10th revision [ICD-10]), drug prescriptions and dispensations in the primary care setting, and clinically relevant parameters (eg, weight, blood pressure, laboratory tests). It is also linked to a hospital discharge database for patients who attend ICS hospitals (30% of the SIDIAP population).[12,13] The cancer registries of Girona (created in 1994) and Tarragona (in 1980) cover 20% of the Catalan population.[14,15] They collect cancer diagnoses from public and private hospitals, anatomopathological and hematological laboratories, mortality registries, and other information sources.[16–18] Both cancer registries comply with the International Agency for Research on Cancer quality requirements.[19]

Study Population And Cancer Case Definition

In the SIDIAP, incident cancer cases were identified as the first cancer diagnosis from 2009 to 2015 among inhabitants of the provinces of Girona and Tarragona. We had the number of incident cancer cases from the cancer registries during 2005–2015 for Girona and during 2005–2013 for Tarragona available for reference. Cases registered during 2005–2008 were used to clean prevalent cases (Figure 1). The linkage between the SIDIAP and the cancer registries data was performed by a Trusted Third Party (the ICS in this study) using the unique personal identification number of patients. We obtained approval from the Clinical Research Ethics Committee of the IDIAPJGol (project code: P14/074) and the Research Ethics Committee of the Hospital Doctor Josep Trueta (project code: 2017.024).

We used ICD-10 codes and date of diagnosis to identify the following 25 cancer types in adults (aged ≥18 years): head and neck (ICD-10 codes: C00-C14), esophagus (C15), stomach (C16), colorectal (C18–21), liver (C22), gallbladder and biliary tract (C23-24), pancreas (C25), larynx (C32), trachea, bronchus, and lung (C33-34), bone and articular cartilage (C40-C41), malignant

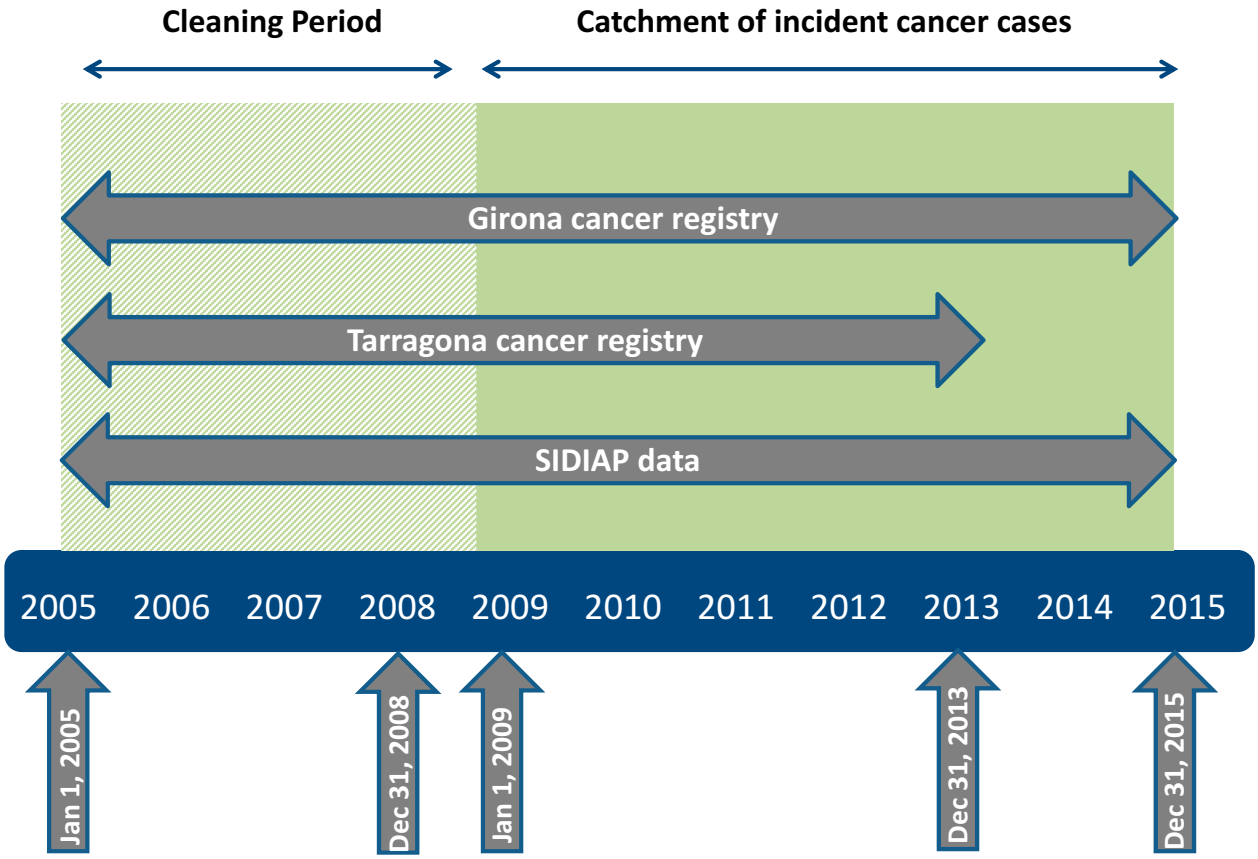


Figure 1 Time period covered by each data source with respect to the duration of the study.
Notes: Figure adapted from Margulis, A. et al. (2017). Validation of Cancer Cases Using Primary Care, Cancer Registry, and Hospitalization Data in the UK. Epidemiology, 29(2), 1.
Abbreviation: SIDIAP, Information System for Research in Primary Care.

melanoma of skin (C43), breast (C50), cervix uteri (C53), corpus uteri (C54-C55), ovary (C56), prostate (C61), testis (C62), kidney (C64), bladder (C67), brain, central nervous system, pituitary gland and pineal gland (C70-72, C75.1-C75.3), thyroid (C73), Hodgkin lymphoma (C81), non-Hodgkin lymphoma (C82-C86, C96), multiple myeloma (C90), and leukemia (C91-95).[20] We excluded other and unspecified malignant neoplasm of skin (C44). Other unspecified or very low-frequency cancers (n<100) were excluded. Diagnosis in hospital discharge data was registered using ICD-9 codes.[21] We mapped diagnosis codes to ICD–10 using available conversion codes eCIEMaps v3.1.9, which we have provided in Supplementary Table S1.

Other Variables

In the SIDIAP, we had information on the primary care center to which individuals were assigned in 2016 (Girona, Tarragona), date of diagnosis, sex (women, men), age (18–35, 36–50, 51–65, ≥66), and nationality (Spanish, non-Spanish). Socioeconomic status was assessed using the "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales" (MEDEA) deprivation index, which we categorized into quintiles for anonymization purposes. The 1st and the 5th quintiles represent the least and most deprived levels of the urban population in Catalonia, respectively.[22] We included a rural category since the MEDEA index was not available for people living in these areas.

Statistical Analysis

We performed a descriptive analysis of the overall number of cancer cases in SIDIAP and of the confirmed (ie, matched diagnoses between the SIDIAP and the cancer registries) vs non-confirmed cases (ie, in the SIDIAP but not in the cancer registries) by sex, age, nationality, MEDEA deprivation index, and year of diagnosis, in Girona and Tarragona, and we used a Chi-squared test to assess for significant differences.[23] We used the Catalonia Cancer Registries (CCRs, Girona and Tarragona

combined) data as the gold-standard to calculate the sensitivity and the positive predictive values (PPVs) for each cancer type (an illustration of our calculations is available in Figure S1. As secondary analyses, we stratified the sensitivity and PPV analyses by province (Girona and Tarragona) to assess if there were geographical differences and by sex, nationality, age, and the MEDEA deprivation index to assess if there were differences for specific population groups. We also checked if the sensitivities improved after including cancer diagnoses from the hospital discharge database.

For the confirmed cases, we calculated the time difference (months) between the date of diagnosis registered in the SIDIAP and the date registered in the CCRs.

We used R version 3.5.0 for all the statistical analyses and considered p-values <0.05 to be statistically significant.

Results

Sociodemographic Characteristics Of SIDIAP And Confirmed Cases

In the SIDIAP, we identified 496,356 inhabitants of Girona in 2016, of which 16,211 had a cancer diagnosis between 2009 and 2015, and 453,485 inhabitants of Tarragona, of which 10,835 had a cancer diagnosis between 2009 and 2013. There were more cancer cases registered in the SIDIAP among men (55%, 56% for Girona and Tarragona, respectively), people aged 66 years or older (45%, 49%), Spanish citizens (94%, 95%), and people living in rural areas (32%, 37%) (Table 1).

We confirmed 9,296 cancer cases in Girona and 7,182 in Tarragona. Compared to non-confirmed cases, confirmed cases had a higher proportion of men in Tarragona (58% vs 52%) as well as people aged 51 to 65 in both provinces (35% vs 31% in Girona; 34% vs 27% in Tarragona) but a lower proportion of socioeconomically deprived individuals in Girona (11% vs 14%) (Table 1).

Overall Validation

Out of the 21,559 cancer cases registered in the CCRs, 16,478 (76%) were in the SIDIAP. The cancer types with the highest sensitivities in Catalonia were breast (89%, 95% CI: 88–90%), colorectal (81%, 95% CI: 80–82%), and prostate (81%, 95% CI: 80–83%) (Table 2). Almost all cancer types had sensitivities above 60% in both provinces. The exceptions were head and neck (51%, 95% CI: 47–55%) and gallbladder and biliary tract (29%, 95% CI: 23–35%) (Table 2).

Out of the 27,046 SIDIAP cancer cases present in Catalonia, 16,478 (61%) were also in the CCRs. The trachea, bronchus and lung cancers had the highest PPV (76%, 95% CI: 74–78%) followed by stomach (72%, 95% CI: 68–75%) and pancreas (71%, 95% CI: 67–75%) cancers (Table 2). On the other hand, bone and articular cartilage (23%, 95% CI: 15–31%) and cervix uteri (28%, 95% CI: 24–33%) cancers had the lowest PPVs (Table 2).

Most cancer diagnoses were reported within less than three months of difference between the SIDIAP and the registries (Figure 2). More cases were reported first in the cancer registries than in the SIDIAP. Only kidney cancer had more than twenty-five percent of cases reported first in the SIDIAP compared to the CCRs.

Secondary Analyses

Overall, Girona had higher sensitivities than Tarragona, especially for cancers of the cervix uteri (68% vs 52%, for Girona and Tarragona, respectively), Hodgkin lymphoma (69% vs 56%) and head and neck (56% vs 45%) (Supplementary Table S2). The only cancer for which Tarragona had a higher sensitivity than Girona was for bone and articular cartilage (56% vs 75%). Regarding PPVs, Tarragona had higher estimates than Girona, except for six cancer types. We observed the biggest differences for bladder (33% vs 69% %, for Girona and Tarragona, respectively), colorectal (65% vs 77%) and larynx (52% vs 63%) cancers. The cancer types for which Girona had the biggest differences in PPVs with Tarragona were gallbladder and biliary tract (56% vs 44%) and Hodgkin lymphoma (71% vs 56%) (Supplementary Table S2).

Overall, sensitivity estimates differed by age groups, and PPVs estimates differed by age, nationality and socioeconomic status. Those older than 66 years showed lower sensitivities than those aged between 36 and 65 years for most cancer types (Supplementary Table S3). Overall, PPVs were lower in those aged between 18 and 35 years than in the rest of age groups, in non-Spanish than in the Spanish population and in the most deprived compared to the least deprived MEDEA quintiles (Supplementary Table S4). Besides the abovementioned situations, we did not observe any other change in the sensitivity and PPVs according to sex, age, nationality, and socioeconomic status, with exception of certain specific cancer types (Supplementary Tables S3 and S4).

Table 1 Descriptive Characteristics Of The SIDIAP Population With A Cancer Diagnosis In Girona (2009–2015) And Tarragona (2009–2013) By Confirmation Status From The Cancer Registries

	SIDIAP Cancer Cases, n (%)							
	Girona				Tarragona			
	SIDIAP Cases	Confirmed Cases	Non-Confirmed[a]	p-value[b]	SIDIAP Cases	Confirmed Cases	Non-Confirmed[a]	p-value[b]
Characteristics	**N=16,211**	**N=9296**	**N=6915**		**N=10,835**	**N=7182**	**N=3653**	
Sex								
Women	7300 (45.0%)	4207 (45.3%)	3093 (44.7%)	0.515	4732 (43.7%)	2994 (41.7%)	1738 (47.6%)	<0.001
Men	8911 (55.0%)	5089 (54.7%)	3822 (55.3%)		6103 (56.3%)	4188 (58.3%)	1915 (52.4%)	
Age (years)[c]								
18–35	624 (3.8%)	298 (3.2%)	326 (4.7%)	<0.001	375 (3.5%)	168 (2.3%)	207 (5.7%)	<0.001
36–50	2830 (17.5%)	1634 (17.6%)	1196 (17.3%)		1636 (15.1%)	1058 (14.7%)	578 (15.8%)	
51–65	5417 (33.4%)	3280 (35.3%)	2137 (30.9%)		3471 (32.0%)	2474 (34.5%)	997 (27.3%)	
≥ 66	7340 (45.3%)	4084 (43.9%)	3256 (47.1%)		5353 (49.4%)	3482 (48.5%)	1871 (51.2%)	
Nationality								
Spanish	15,182 (93.7%)	8731 (93.9%)	6451 (93.3%)	0.110	10,328 (95.3%)	6930 (96.5%)	3398 (93.0%)	<0.001
Non-Spanish	1029 (6.3%)	565 (6.1%)	464 (6.7%)		507 (4.7%)	252 (3.5%)	255 (7.0%)	
MEDEA deprivation index[d]								
Quintile 1	1784 (11.0%)	1171 (12.6%)	613 (8.9%)	<0.001	813 (7.5%)	528 (7.3%)	285 (7.8%)	0.022
Quintile 2	1461 (9.0%)	978 (10.5%)	483 (7.0%)		1062 (9.8%)	716 (10.0%)	346 (9.5%)	
Quintile 3	2180 (13.5%)	1152 (12.4%)	1028 (14.9%)		1201 (11.1%)	784 (10.9%)	417 (11.4%)	
Quintile 4	2540 (15.7%)	1209 (13.0%)	1331 (19.2%)		1664 (15.4%)	1097 (15.3%)	567 (15.5%)	
Quintile 5	2025 (12.5%)	1029 (11.1%)	996 (14.4%)		1366 (12.6%)	936 (13.0%)	430 (11.8%)	
Rural areas	5179 (31.9%)	3205 (34.5%)	1974 (28.5%)		4001 (36.9%)	2677 (37.3%)	1324 (36.2%)	
"Missing"	1042 (6.4%)	552 (5.9%)	490 (7.1%)		728 (6.7%)	444 (6.2%)	284 (7.8%)	
Year of diagnosis								
2009	2354 (14.5%)	1151 (12.4%)	1203 (17.4%)	<0.001	2143 (19.8%)	1150 (16.0%)	993 (27.2%)	<0.001
2010	2329 (14.4%)	1342 (14.4%)	987 (14.3%)		2154 (19.9%)	1433 (20.0%)	721 (19.7%)	
2011	2310 (14.3%)	1401 (15.1%)	909 (13.1%)		2046 (18.9%)	1443 (20.1%)	603 (16.5%)	
2012	2374 (14.6%)	1438 (15.5%)	936 (13.5%)		2212 (20.4%)	1553 (21.6%)	659 (18.1%)	
2013	2365 (14.6%)	1471 (15.8%)	894 (12.9%)		2280 (21.0%)	1603 (22.3%)	677 (18.5%)	
2014	2259 (13.9%)	1383 (14.9%)	876 (12.7%)		-	-	-	
2015	2220 (13.7%)	1110 (11.9%)	1110 (16.1%)		-	-	-	

Notes: [a]Non-confirmed cases either have different diagnoses in SIDIAP and the registry or were not available in the registry. [b]Comparison of confirmed vs non-confirmed cases using Chi-squared of independence test. [c]Age in 2009. [d]Quintile 1 of the MEDEA Index represents the least deprived and quintile 5 represents the most deprived. Rural was included as a category since the index cannot be calculated for people living in rural areas.

Abbreviations: SIDIAP, Information System for Research in Primary Care.

Table 2 Validity Of The ICD-10 Codes Used To Identify Incident Cancer Diagnoses Registered In The SIDIAP Database, Catalonia[a] (2009–2015)[b]

Cancer Type (ICD-10 CM)	Cancer Cases, n			Sensitivity, % (95% CI)	PPV, % (95% CI)
	CCRs	SIDIAP	Confirmed		
Head and neck (C00-C14)	650	819	332	51.1 (47.2–54.9)	40.5 (37.2–43.9)
Esophagus (C15)	211	255	157	74.4 (68.5–80.3)	61.6 (55.6–67.5)
Stomach (C16)	673	633	455	67.6 (64.1–71.1)	71.9 (68.4–75.4)
Colorectal (C18-C21)	3743	4329	3035	81.1 (79.8–82.3)	70.1 (68.7–71.5)
Liver (C22)	561	625	364	64.9 (60.9–68.8)	58.2 (54.4–62.1)
Gallbladder & biliary tract (C23-C24)	197	107	57	28.9 (22.6–35.3)	53.3 (43.8–62.7)
Pancreas (C25)	578	590	419	72.5 (68.8–76.1)	71.0 (67.4–74.7)
Larynx (C32)	337	403	226	67.1 (62.0–72.1)	56.1 (51.2–60.9)
Trachea, bronchus & lung (C33-C34)	2152	2155	1631	75.8 (74.0–77.6)	75.7 (73.9–77.5)
Bone and articular cartilage (C40-C41)	39	106	24	61.5 (46.3–76.8)	22.6 (14.7–30.6)
Malignant melanoma of skin (C43)	550	962	417	75.8 (72.2–79.4)	43.3 (40.2–46.5)
Breast (C50)	3325	4456	2958	89.0 (87.9–90.0)	66.4 (65.0–67.8)
Cervix uteri (C53)	198	416	118	59.6 (52.8–66.4)	28.4 (24.0–32.7)
Corpus uteri (C54-C55)	576	661	424	73.6 (70.0–77.2)	64.1 (60.5–67.8)
Ovary (C56)	263	398	190	72.2 (66.8–77.7)	47.7 (42.8–52.6)
Prostate (C61)	2820	3596	2286	81.1 (79.6–82.5)	63.6 (62.0–65.1)
Testis (C62)	139	175	102	73.4 (66.0–80.7)	58.3 (51.0–65.6)
Kidney (C64)	536	730	397	74.1 (70.4–77.8)	54.4 (50.8–58.0)
Bladder (C67)	1456	2370	1108	76.1 (73.9–78.3)	46.8 (44.7–48.8)
Brain and CNS (C70-C72, C75.1-C75.3)[c]	393	544	298	75.8 (71.6–80.1)	54.8 (50.6–59.0)
Thyroid (C73)	395	432	264	66.8 (62.2–71.5)	61.1 (56.5–65.7)
Hodgkin lymphoma (C81)	144	142	92	63.9 (56.0–71.7)	64.8 (56.9–72.6)
Non-Hodgkin lymphoma (C82-C86, C96)	709	909	472	66.6 (63.1–70.0)	51.9 (48.7–55.2)
Multiple myeloma (C90)	294	362	233	79.3 (74.6–83.9)	64.4 (59.4–69.3)
Leukemia (C91-C95)	620	871	419	67.6 (63.9–71.3)	48.1 (44.8–51.4)

Notes: [a]Provinces of Girona and Tarragona. [b]Data from the Tarragona Cancer Registry was only available for 2009–2013. [c]Include pituitary gland and pineal gland tumors.
Abbreviations: CI, Confidence Interval; CNS, Central Nervous System; CCRs, Catalonia Cancer Registries; ICD-10, International Classification for Diseases, 10th revision; PPV, positive predictive values; SIDIAP, Information System for Research in Primary Care.

When adding cancer diagnoses from hospital discharge to primary care data, we observed an increase in sensitivity for all cancer types. Gallbladder and biliary tract cancer had the most substantial change in sensitivity, changing from 29% to 50% (Supplementary Table S5). We also observed changes above 10% for larynx (67% to 83%), head and neck (51% to 66%) and liver (65% to 78%) cancers (Supplementary Table S5).

Discussion

This study validated cancer diagnoses recorded in primary care using the data of the two provincial population-based cancer registries that exist in Catalonia as the gold-standard. We found that 23 out of 25 cancer types had sensitivities above 60%. PPV estimates were generally lower than the sensitivities observed in most cancer types. The number of cancer cases in the SIDIAP that were not confirmed by the cancer registries was high for some specific cancer sites. More cases were first recorded in the cancer registries rather than in the SIDIAP, though for most cancer cases, the time difference between both data sources did not exceed three months. Including cancer diagnoses from hospital discharge data considerably improved the reliability of the data for specific cancer types.

We observed a high sensitivity for the majority of cancer types. Breast, colorectal and prostate cancers had the highest sensitivities, which are some of the most incident tumors and thoroughly screened cancers in systematic programs (breast and colorectal) and strongly sought by opportunistic screening (prostate) in Catalonia.[24,25] Furthermore, these cancers take part in the rapid diagnostic circuit program run in Catalonia, which could also contribute to an increase in the accuracy of diagnosis in primary care.[26] Previous studies conducted in the United Kingdom (UK) that compared primary care data with

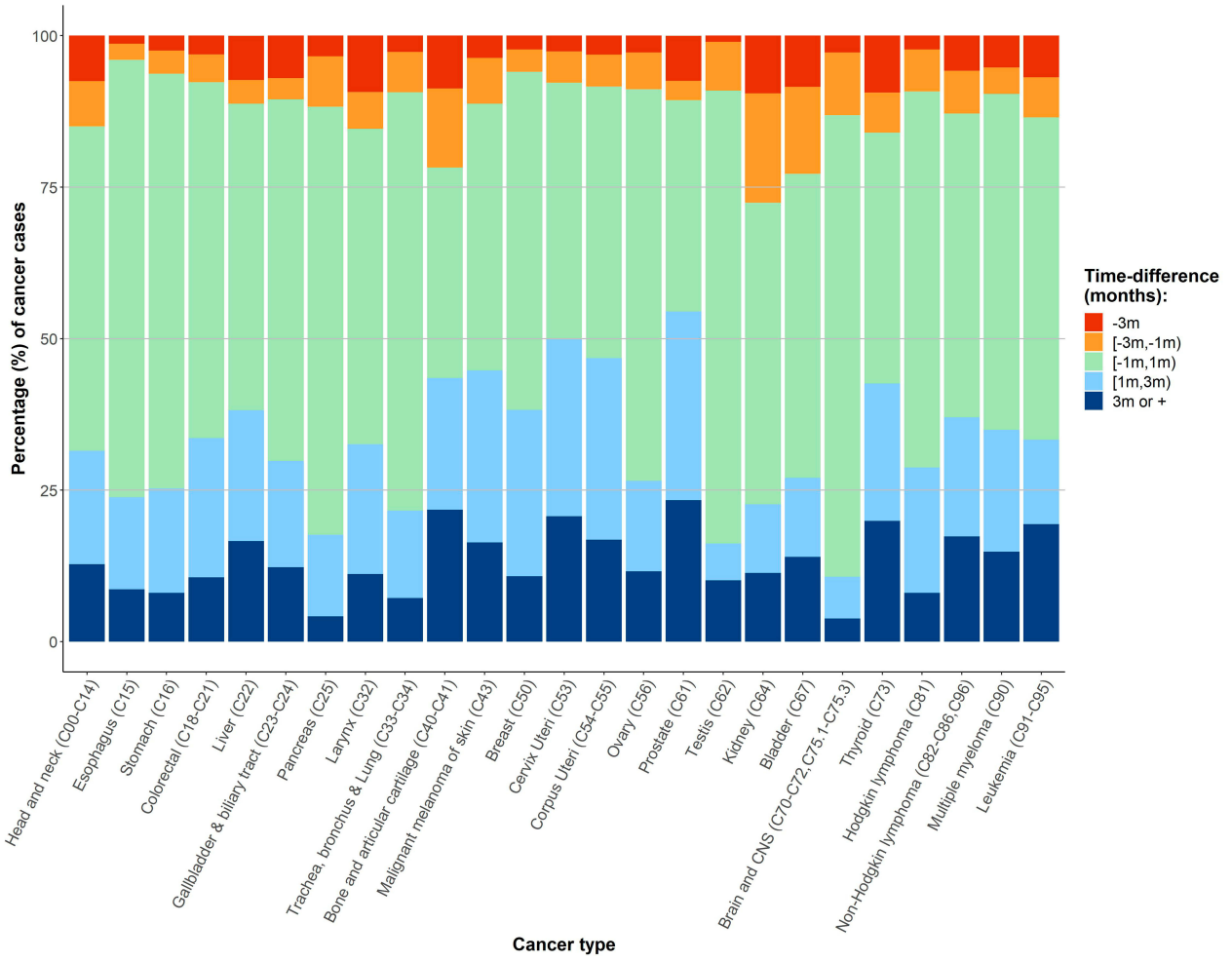


Figure 2 Time-difference (months) in the date of cancer diagnosis recorded in the SIDIAP and the population-based Catalonia Cancer Registries[a] (2009–2015)[b].
Notes: [a]Population-based cancer registries from the provinces of Girona and Tarragona. [b]Data from the Tarragona Cancer Registry was only available for 2009–2013. Negative values indicate SIDIAP diagnosis before the registries' diagnosis date. Brain and CNS include pituitary gland and pineal gland tumors.
Abbreviations: CNS, Central Nervous System; m, months; SIDIAP, Information System for Research in Primary Care.

hospital and cancer registry data also reported high sensitivities for breast, prostate, and colorectal cancers, highlighting that these cancers are usually managed by general practitioners.[27,28] In Catalonia, a previous study comparing SIDIAP cases with those registered in a hospital cancer registry in Barcelona, also reported high sensitivities for breast, colorectal and prostate cancers.[9] High sensitivities are important to enhance study inclusiveness and to be able to ascertain common exposures.[29] A high sensitivity paired with a high specificity (which is important for classifying outcomes) facilitates both the study of cancer as an outcome as well as the identification of the cases' common exposures. In our study, the lowest sensitivities were found for cancers that are less frequent and that are more commonly managed in hospitals, such as gallbladder and biliary tract or bone and articular cartilage.[9,24,30,31] We are not aware of any previous national or international studies validating the primary care diagnosis of these cancer types using external sources. Thus, our results indicate that using SIDIAP cancer diagnoses for research when cancer is the outcome of interest is reliable for most common cancer types in Catalonia but may be insufficient for less frequent types.

PPV estimates were generally lower than the sensitivities observed in most cancer types. The number of cancer cases in the SIDIAP that were not confirmed by the cancer registries was high for some specific cancer sites. A previous study validating only colorectal, lung, gastro-esophageal and urological cancer diagnosis in primary care in the UK reported higher PVV estimates than in our study, ranging from 92% to 98%.[28] This study hypothesized that some of the reasons behind non-confirmed cases might be

a disagreement in the type of cancer diagnosed in each data source, or the possibility of suspicious symptoms being registered as cancer diagnoses in primary care.[28] In agreement with this hypothesis, we found that approximately 10% of the non-confirmed cases by the cancer registries were due to disagreement in the type of cancer diagnosis between the data sources. The low PPV for cervix uteri cancer (included in the rapid diagnostic circuit in Catalonia) could be due to detected suspicious symptoms recorded as cancers in SIDIAP; however, we did not have the information needed to prove this hypothesis. Another factor that can influence PPVs is the prevalence of the cancer type which could partially explain the low PPVs of bone and articular cartilage (106 cases registered in the SIDIAP) and gallbladder and biliary tract (107 cases registered). High PPVs are important when we want to identify a cohort of people that only includes people with the condition of interest but do not need to be representative of all cases.[29] Therefore, the SIDIAP does not appear to be an appropriate database to create a cohort of cancer patients, except for certain cancer types (eg, trachea, bronchus and lung, stomach, pancreas or colorectal cancers). More research needs to be conducted to understand the reasons behind non-confirmed cancer cases in SIDIAP.

Most cancer diagnoses were reported within less than three months of difference between the SIDIAP and the registries, and generally, the cancer registries reported the cases earlier than the SIDIAP. Our results are in line with two previous studies in the UK which assessed the time difference between the date of cancer diagnoses registered in the cancer registries and primary care databases. One study reported a median time difference in the date of diagnosis of 11 days (range 6–30 days) between a UK primary care database and the Cancer Registry in England for colorectal, lung, gastro-esophageal and urological cancers.[28] The other study, also using information from the same UK primary care database and cancer registry but combining 11 cancer types, reported that 63% of cancer diagnoses were recorded with one month of difference between the data sources and 24% within one to three months of difference. However, the authors did not specify which source registered the diagnosis first.[32] Although the time difference between the data sources was not substantial in our study, investigators should be aware of it when addressing time-related research questions in the SIDIAP, such as those in the cancer survival field.

In our study, the inclusion of hospital discharge data to SIDIAP cancer diagnoses improved the sensitivity estimates for most cancer sites, with substantial improvements observed particularly for less frequent cancer types. The use of multiple data sources is highly recommended when using EHRs for epidemiological research since the advantages of each database can overcome the limitations of the others.[4,33] Specifically, the need to link primary care databases to those from hospitals and cancer registries to correctly identify certain cancer types has been proposed in the UK.[27] Therefore, considering both SIDIAP and hospital discharge databases can improve the reliability in the results of future research. This may be especially important for larynx, head and neck and liver cancers. For gallbladder and biliary tract cancer, despite the sizeable improvement in sensitivity after adding hospital discharge to SIDIAP cancer diagnoses, the final sensitivity estimate (50%) seems insufficient to perform future studies using this cancer type as an outcome. If data is available, future studies may consider restricting their analyses to confirmed cases only to avoid misclassifications and attain data robustness.

The main strengths of this study are first, the use of the SIDIAP database, which provides a large and representative sample of the Catalonian population and increases external validity.[11] Second, the use of two population-based cancer registries as the gold-standard allowed us to validate numerous cancer types. Third, we were able to calculate the sensitivity of the SIDIAP cancer diagnoses, a type of measure that is often not reported in cancer validation studies. However, our study has limitations. First, since the SIDIAP is a primary care database, certain cancer types are harder to be detected at this level; nevertheless, we assessed the inclusion of hospital discharge information to account for this limitation. Second, textual information in medical records could be of value to distinguish cancer suspicions from actual diagnoses in the SIDIAP, but this information was not available in this study. Third, for this study we were only able to add cancer diagnoses from hospital discharge from the ICS hospitals, therefore we cannot confirm whether including information from all Catalan hospitals would permit better identification of cases for the same cancer types we found. Finally, our population of reference was the population of individuals assigned to a primary care center in Girona and Tarragona provinces in 2016 and, thus, we could not account for changes in patient address during the whole study period.

Conclusion

The SIDIAP includes 76% of the cancer diagnoses present in the cancer registries of Catalonia but also includes a considerable number of cases that are not in the registries.

Overall, the SIDIAP reports cancer cases later than the registries but the time difference in the date of diagnosis between the databases is usually less than three months. Our results support the use of SIDIAP cancer diagnoses for national and international epidemiological research when cancer is used as an outcome, especially for the most frequent cancer types. The inclusion of cancer diagnoses from hospital discharge data is recommended to improve the reliability of certain cancer types such as head and neck, liver, larynx, and leukemia. However, our results do not support the use of SIDIAP data for all cancer sites when the purpose of the study is to identify a cohort of cancer patients. Further research is needed to understand the cancer cases recorded in the SIDIAP that were not confirmed by the cancer registries.

Abbreviations

CI, Confidence Interval; CCRs, Catalonia Cancer Registries; CNS, Central Nervous System; EHRs, Electronic Health Records; ICD-9,International Classification for Diseases, 9th revision; ICD-10, International Classification for Diseases, 10th revision; ICS, Institut Català de la Salut; MEDEA, Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales; PPV, Positive Predictive Value; SIDIAP, Information System for Research in Primary Care; UK, United Kingdom.

Ethics Approval

The Clinical Research Ethics Committee of the IDIAPJGol (project code: P14/074) and the Research Ethics Committee of the Hospital Doctor Josep Trueta (project code: 2017.024) approved this study.

Data Availability

In accordance with current European and national law, the data used in this study is only available for the researchers participating in this project. Thus, we are not allowed to distribute or make publicly available the data to other parties. However, researchers from public institutions can request data from the SIDIAP and other sources (eg, Cancer Registries) if they comply with certain requirements. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudes-en/application-proccedure) or by contacting Anna Moleras amoleras@idiapjgol.org).

Acknowledgements

We thank the Registre del conjunt mínim de bases de dades (CMBD), for providing data on Hospital Discharges. We also thank Leonardo Mendez-Boo and Eduardo Hermosilla for their support on the data linkage processes and Matthew Bennett for revising the English language composition.

Funding

Funding [for grant number: 2017/1630] was obtained from Wereld Kanker Onderzoek Fonds (WKOF), as part of the World Cancer Research Fund International grant program. TDS is funded by the Department of Health of the Generalitat de Catalunya, awarded on the 2016 call under the Strategic Plan for Research and Innovation in Health (PERIS) 2016–2020, modality incorporation of scientists and technologists, with reference SLT002/16/00308.

Disclosure

The authors report no conflicts of interest in this work.

References

1. International Agency for Research on Cancer. *World Cancer Report 2014*. Lyon; 2014.
2. Global Cancer Observatory. Cancer tomorrow. International Agency for Research on Cancer. https://gco.iarc.fr/tomorrow/home. Published 2018. Accessed March 10, 2019.
3. Galceran J, Ameijide A, Carulla M, et al. Cancer incidence in Spain, 2015. *Clin Transl Oncol*. 2017;19(7):799–825. doi:10.1007/s12094-016-1607-9
4. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93–101. doi:10.1002/bjs.9723
5. Haynes K, Forde KA, Schinnar R, Wong P, Strom BL, Lewis JD. Cancer incidence in the health improvement network. *Pharmacoepidemiol Drug Saf*. 2009;18:730–736. doi:10.1002/pds.v18:8
6. Nissen F, Quint JK, Morales DR, Douglas IJ. How to validate a diagnosis recorded in electronic health records. *Breathe*. 2019;15(1):64–68. doi:10.1183/20734735.0344-2018
7. Instituto Nacional de Estadística. Estadísticas Territoriales: cataluña. http://www.ine.es/FichasWeb/RegComunidades.do?fichas=49&busc_comu=&botonFichas=Ir+a+la+tabla+de+resultados. 2018. Accessed October 10, 2018.
8. Ramos R, Balló E, Marrugat J, et al. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Española Cardiol (English Ed*. 2012;65(1):29–37. doi:10.1016/j.recesp.2011.07.017
9. Garcia-Gil M, Elorza J-M, Banque M, et al. Linking of primary care records to census data to study the association between socioeconomic status and cancer incidence in Southern Europe: a nation-wide ecological study. *PLOS ONE*. 2014;9(10):e109706. doi:10.1371/journal.pone.0109706
10. Bolíbar B, Fina Avilés F, Morros R, et al. Base de datos SIDIAP: la historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Med Clin (Barc)*. 2012;138(14):617–621. doi:10.1016/j.medcli.2012.01.020
11. García-Gil MDM, Hermosilla E, Prieto-Alhambra D, et al. Construction and validation of a scoring system for selection of high quality data in a Spanish population primary care database (SIDIAP). *Inf Prim Care*. 2012;20(2):1.

12. Generalitat de Catalunya. Conjunt mínim bàsic de dades (CMBD). https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/index.html#googtrans(ca%7Ces. 2017. Accessed March 5, 2019.
13. Instituto Nacional de Estadística. Hospital morbidity Survey: general methodology. http://www.ine.es/en/daco/daco42/sanitarias/notaemh_en.htm. 2019. Accessed March 6, 2019.
14. Josep MB, Izquierdo A, Vilardell L, et al. Cancer Incidence in Girona (2008-2012). In: Bray F, Colombet M, Mery L, et al., editors. *Cancer Incidence in Five Continents*. Vol. XI (electronic version). Lyon: International Agency for Research on Cancer; 2017. Available from: http://ci5.iarc.fr. Accessed March 10, 2018.
15. Galceran J, Carulla M, Ameijide A, et al. Cancer incidence in Tarragona (2008–2012). In: Bray F, Colombet M, Mery L, et al., editors. *Cancer Incidence in Five Continents*. Vol. XI (electronic version). Lyon: International Agency for Research on Cancer; 2017. Available from: http://ci5.iarc.fr. Accessed March 10, 2018.
16. Unitat d'Epidemiologia i Registre de Càncer de Girona. *El Càncer a Girona 2010–12: Projeccions De La Incidència 2017*. Girona: Institut Català d'Oncologia; 2016.
17. Navarro C, Martos C, Ardanaz E, et al. Population-based cancer registries in Spain and their role in cancer control. *Ann Oncol*. 2010;21(Supplement3):iii3–iii13. doi:10.1093/annonc/mdq094
18. International Agency for Research on Cancer. Indices of data Quality: All sites except non-melanoma skin (C00-96 exc. C44). *Cancer Incidence in Five Continents Volume XI*. http://ci5.iarc.fr/CI5-XI/PDF/INDICES/21.pdf. 2017. Accessed March 10, 2018.
19. International Agency for Research on Cancer. Chapter 5: Data Comparability and Quality. Cancer Incidence in Five Continents Volume XI. http://ci5.iarc.fr/CI5-XI/Pages/Chapter5.aspx. 2017. Accessed March 10, 2018.
20. World Health Organization. *ICD-10 : International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*. 2nd ed. Geneva: World Health Organization; 2004.
21. World Health Organization. Ninth revision of the International Classification of Diseases. 1976.
22. Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). *Gac Sanit*. 2008;22(3):179–187. doi:10.1157/13123961
23. Subirana I, Vila J, Sanz H, Lucas G, Penafiel J, Gimenez D. Building bivariate tables: the comparegroups package for R. *J Stat Softw*. 2014;57(12):1–16. doi:10.18637/jss.v057.i12
24. Pla director d'Oncologia de Catalunya. *El Càncer a Catalunya*. Barcelona; 2016.
25. Instituto Catalán de Oncología. Prevención del cáncer. http://ico.gencat.cat/es/el-cancer/programes_de_deteccio_precoc/. Accessed April 17, 2019.
26. Generalitat de Catalunya. Circuit de diagnòstic ràpid. http://canalsalut.gencat.cat/ca/salut-a-z/c/cancer/recursos-per-a-professionals/diagnostic/circuit-de-diagnostic-rapid/index.html. 2018. Accessed April 16, 2019.
27. Margulis AV, Fortuny J, Kaye JA, et al. Validation of cancer cases using primary care, cancer registry, and hospitalization data in the United Kingdom. *Epidemiology*. 2018;29(2):308–313. doi:10.1097/EDE.0000000000000786
28. Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiol*. 2012;36(5):425–429. doi:10.1016/j.canep.2012.05.013
29. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65(3):343–349.e2. doi:10.1016/j.jclinepi.2011.09.002
30. American Cancer Society. Can gallbladder cancer be found early? https://www.cancer.org/cancer/gallbladder-cancer/detection-diagnosis-staging/detection.html. 2018. Accessed March 5, 2019.
31. American Cancer Society. Can bone cancer be found early? https://www.cancer.org/cancer/bone-cancer/detection-diagnosis-staging/detection.html. 2018. Accessed March 5, 2019.
32. Boggon R, Van Staa T, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer recording and mortality in the general practice research database and linked cancer registries. *Pharmacoepidemiol Drug Saf*. 2013;22:168–175. doi:10.1002/pds.v22.2
33. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol*. 2010;69(1):4–14. doi:10.1111/bcp.2010.69.issue-1

5.3. Study III

Recalde M, Davila-Batista V, Díaz Y, Leitzmann M, Romieu I, Freisling H, Duarte-Salles T.

Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain

BMC Medicine (2021) 19:10

Recalde et al. BMC Medicine (2021) 19:10
https://doi.org/10.1186/s12916-020-01877-3

BMC Medicine

RESEARCH ARTICLE Open Access

Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain

Martina Recalde[1,2], Veronica Davila-Batista[1,3,4], Yesika Díaz[1], Michael Leitzmann[5], Isabelle Romieu[6,7], Heinz Freisling[3†] and Talita Duarte-Salles[1*†]

Abstract

Background: A high body mass index (BMI) has been associated with increased risk of several cancers; however, whether BMI is related to a larger number of cancers than currently recognized is unclear. Moreover, whether waist circumference (WC) is more strongly associated with specific cancers than BMI is not well established. We aimed to investigate the associations between BMI and 26 cancers accounting for non-linearity and residual confounding by smoking status as well as to compare cancer risk estimates between BMI and WC.

Methods: Prospective cohort study with population-based electronic health records from Catalonia, Spain. We included 3,658,417 adults aged ≥ 18 years and free of cancer at baseline between 2006 and 2017. Our main outcome measures were cause-specific hazard ratios (HRs) with 99% confidence intervals (CIs) for incident cancer at 26 anatomical sites.

Results: After a median follow-up time of 8.3 years, 202,837 participants were diagnosed with cancer. A higher BMI was positively associated with risk of nine cancers (corpus uteri, kidney, gallbladder, thyroid, colorectal, breast post-menopausal, multiple myeloma, leukemia, non-Hodgkin lymphoma) and was positively associated with three additional cancers among never smokers (head and neck, brain and central nervous system, Hodgkin lymphoma). The respective HRs (per 5 kg/m^2 increment) ranged from 1.04 (99%CI 1.01 to 1.08) for non-Hodgkin lymphoma to 1.49 (1.45 to 1.53) for corpus uteri cancer. While BMI was negatively associated to five cancer types in the linear analyses of the overall population, accounting for non-linearity revealed that BMI was associated to prostate cancer in a U-shaped manner and to head and neck, esophagus, larynx, and trachea, bronchus and lung cancers in an L-shaped fashion, suggesting that low BMIs are an approximation of heavy smoking. Of the 291,305 participants with a WC measurement, 27,837 were diagnosed with cancer. The 99%CIs of the BMI and WC point estimates (per 1 standard deviation increment) overlapped for all cancers.

(Continued on next page)

* Correspondence: tduarte@idiapjgol.org
†Heinz Freisling and Talita Duarte-Salles share joint last authorship.
[1]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Gran Via Corts Catalanes, 587 àtic, 08007 Barcelona, Spain
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusions: In this large Southern European study, a higher BMI was associated with increased risk of twelve cancers, including four hematological and head and neck (only among never smokers) cancers. Furthermore, BMI and WC showed comparable estimates of cancer risk associated with adiposity.

Keywords: Body mass index, Waist circumference, Body size, Body fat distribution, Adiposity, Obesity, Cancer, Electronic health records

Background

The prevalence of obesity worldwide has nearly tripled over the past three decades, reaching 650 million adults in 2016 [1]. Body mass index (BMI), the most common indicator of general adiposity, has been convincingly associated with at least 12 cancer types [2]. Results from previous large cohort studies suggest that BMI is associated with a larger number of cancer types than currently recognized and that some of those associations may be non-linear [3, 4]. However, the main limitations of available studies include limited adjustment for potential confounding, reliance on self-reported weight and height, and lack of generalizability to different populations. Furthermore, although conducting analyses stratified by smoking status is critical to provide unbiased estimates of the impact of obesity on cancer risk [4, 5], many studies failed to present results stratified by smoking status, in part due to insufficient statistical power [3].

In addition, whether BMI as a sole indicator of general adiposity fully captures the complex association between adiposity and cancer risk is still in dispute. Central adiposity, typically assessed using waist circumference (WC), has been suggested to increase the risk of several cancer types and to better discriminate risk associated with obesity for colon and breast post-menopausal cancers [6–8]. However, only few studies have systematically compared the effect estimates of BMI and WC for multiple site-specific cancers, and none have studied less frequently occurring cancer types [9, 10].

The primary objective of the current study was to investigate associations between BMI and the risk of 26 types of cancer accounting for non-linearity and residual confounding by smoking status. Our secondary objective was to compare risk estimates for general (BMI) and central (WC) adiposity in relation to the risk of 26 cancer types.

Methods

Study design, setting, and data sources

We performed a cohort study with prospectively collected data from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org), from January 1, 2006, until December 31, 2018. SIDIAP includes routinely recorded information by health professionals from 287 primary care centers in Catalonia, a region in Northeastern Spain [11, 12]. SIDIAP contains anonymized records for approximately six million people (80% of the Catalan population) and is representative of the Catalan population in terms of age, sex, and geographic distribution [12]. It includes high-quality data on anthropometric measurements, disease diagnoses (International Classification for Diseases, 10th revision [ICD-10]), prescription and dispensation of drugs, laboratory tests, and demographic and lifestyle information. Further, SIDIAP is linked to the Minimum Basic Dataset (CMBD in Spanish), a population-based registry that includes hospital discharge information in Spain [13].

Participants

For the primary objective, we included all participants aged ≥ 18 years with a valid BMI (weight (kg)/height $(m)^2$ between 15 and 60 kg/m^2) recorded between January 1, 2006, and December 31, 2017, and subsequent eligible follow-up time (minimum of 1 year). The study's index date was the date of the first BMI assessment during this period. We followed participants from the study index date until first incident (primary) cancer diagnosis, death, transferal out of the SIDIAP, or until the end of the study period (December 31, 2018). We excluded individuals who were older than 100 years of age at index date, had a BMI assessment only available during pregnancy (from the 3rd month of pregnancy until 2 months after delivery), had any record of cancer before the study index date, or complied with any of the end-of-follow-up criteria described above before attaining 12 months of follow-up to avoid reverse causality (Fig. 1, dataset 1). For our secondary objective, we included an additional eligibility criterion, which was to have a valid WC assessment (WC values ≥ 40 cm and ≤ 160 cm) no more than 5 years previous to or 1 year later than the index date (first BMI measurement recorded) (Fig. 1, dataset 2). If a participant had more than one WC measurement available, we selected the closest one to the index date. Figure 1 shows the flow chart of inclusion and exclusion criteria for each study objective.

Assessment of anthropometric indicators and covariates

For our primary objective, the exposure of interest was BMI as a continuous variable (in kg/m^2). BMI was automatically calculated through a computer program

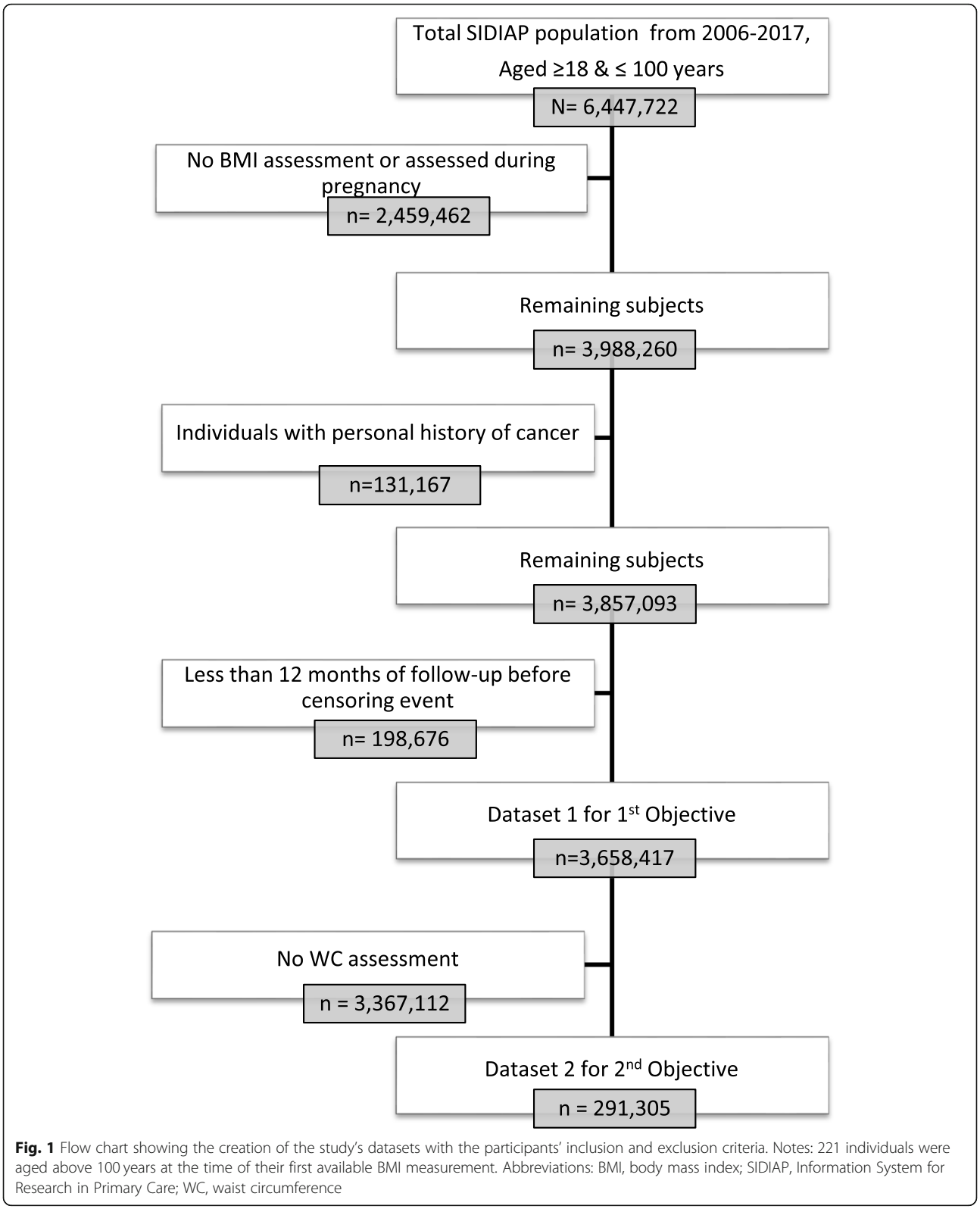


Fig. 1 Flow chart showing the creation of the study's datasets with the participants' inclusion and exclusion criteria. Notes: 221 individuals were aged above 100 years at the time of their first available BMI measurement. Abbreviations: BMI, body mass index; SIDIAP, Information System for Research in Primary Care; WC, waist circumference

("Estació clínica d'atenció primària") after general practitioners (GPs) or nurses entered the weight (kg) and height (cm) of patients they directly assessed in a standardized manner [14]. For participants without information from that computer program, we calculated the BMI using weight and height data available in their health records (if height was not available on the same date as the weight measurement, we calculated the individuals' mean height using all available measurements in their health records during adulthood (≥ 18 years) and we chose the closest real height value to the mean). For our secondary objective, we additionally considered WC as an exposure; this indicator was routinely measured by trained health professionals (GPs and nurses) who follow a measurement protocol [15]. WC was measured at the umbilical level, midway between the anterior superior iliac spine and the inferior border of the rib while participants were standing.

We also extracted information on sex (women, men), age (in years), and geographic region of nationality (Spain, European [non-Spanish], Africa, America, and Asia). We assessed socioeconomic status in urban areas using the "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales" (MEDE A) deprivation index, which is calculated at the census tract level and was categorized into quintiles by the SIDIAP for anonymization purposes [16]. The first and the fifth quintiles represent the least and most deprived groups of the population living in urban areas of Catalonia, respectively. We included a rural category since the MEDEA index was not available for participants living in those areas. We also extracted information on smoking status (never, former, or current smoker) and alcohol intake (none, low or high). If a participant had more than one record of smoking status and alcohol intake available, we selected the one closest to the index date within a 6-year period (5 years before and 1 year after the first BMI measurement). For type 2 diabetes, we considered any registry of a GP diagnosis (ICD-10 code E11) before the index date. For women, we included information on menopausal status and hormonal replacement therapy (HRT) use, the definitions of which can be consulted in Additional file 1: Appendix S1.

Ascertainment of cancer cases

We considered first incident cancer diagnoses as the outcomes of interest. We identified outcomes using ICD-10 codes in the SIDIAP database and ICD-9 codes in the CMBD from January 1, 2007, to December 31, 2018. We mapped ICD-9 diagnosis codes to ICD-10 using available conversion codes (eCIEMaps v3.1.9) which are provided in Additional file 1: Table S1. We used the following cancer types as outcomes: head and neck; esophagus; stomach; colorectal; liver; gallbladder and biliary tract; pancreas; larynx; trachea, bronchus, and lung; bone and articular cartilage; malignant melanoma of skin; breast (which we categorized into pre- and post-menopausal due to well-established evidence indicating different BMI relations) [17]; cervix uteri; corpus uteri; ovary; prostate; testis; kidney; bladder; brain and central nervous system (CNS); thyroid; Hodgkin lymphoma; non-Hodgkin lymphoma; multiple myeloma; and leukemia. All cancer diagnoses in the SIDIAP including the CMBD have been previously validated [18].

Statistical analysis

We described the number of excluded individuals in each step of the creation of the main dataset. We presented the overall baseline characteristics of the study participants and by the World Health Organization (WHO) BMI categories: underweight or normal weight (BMI $< 18.5\ kg/m^2$ and between ≥ 18.5 and $< 25\ kg/m^2$), overweight (BMI ≥ 25 and $< 30\ kg/m^2$), and obesity (BMI $\geq 30\ kg/m^2$).

We fitted Cox proportional hazard models with age as the time metric to estimate cause-specific hazard ratios (HR) and 99% confidence intervals (CI) for the relation between BMI and risk of each cancer type. We stratified all models by age (5-year categories) and sex to reduce the sensitivity to violations of the proportional hazards assumption. The first (basic) model included BMI only (model 1) and the second (multivariable-adjusted) model further adjusted for smoking status, alcohol intake, type 2 diabetes, socioeconomic status, and nationality (model 2). A directed acyclic graph was used to guide decisions on the control for confounding (Additional file 1: Fig. S1) [19]. We used a missing category for variables with missing data.

Firstly, we investigated potential non-linear associations between BMI and risk of each cancer. We considered non-linearity in BMI by fitting models using restricted cubic splines for BMI with 3 knots (placed at the 10th, 50th, and 90th percentiles) or 5 knots (placed at the 5th, 27.5th, 50th, 72.5th, and 95th percentiles). We evaluated linearity by comparing the Akaike information criterion of models with restricted splines to the model with BMI as a linear term in combination with a Wald test linearity hypothesis [20, 21]. To assess residual confounding by smoking, we re-run the multivariable-adjusted (adjusted for alcohol intake, type 2 diabetes, socioeconomic status, and nationality) models among never smokers for cancers for which we found evidence of non-linearity.

Secondly, we fitted model 2 with BMI as a linear term to estimate HRs of the relation between BMI (per $5\ kg/m^2$ increment) and risk of each cancer type. Again, we re-run the multivariable-adjusted models (adjusted for alcohol intake, type 2 diabetes, socioeconomic status,

and nationality) only among participants who reported having never smoked to explore residual confounding by smoking.

In the subsample of participants who had information on both BMI and WC (Fig. 1, dataset 2), we compared risk estimates for general (BMI) and central (WC) adiposity in relation to the risk of 26 cancers by fitting Cox proportional hazard models (one for each adiposity indicator) with age as the time metric. We estimated HRs and 99% CIs per 1 standard deviation (SD) increment of adiposity indicators (BMI and WC) to allow comparability between both estimates [9]. We considered estimates different if the 99% CIs of the point estimates of each adiposity indicator did not overlap. We adjusted the statistical models for the same variables as in model 2, and we used the same end of follow-up definition. We only analyzed cancer types for which we ascertained at least 100 cancer cases.

Model-checking and sensitivity analyses

For all models, we checked the proportional hazard assumption by using the Schoenfeld test of proportionality and by visual inspection of the scaled Schoenfeld residuals [22].

We assessed the robustness of our primary objective findings by performing six sensitivity analyses. First, we accounted for residual selection bias by additionally adjusting model 2 for the number of GP consultations in the year of the index date because participants who see their GP more often may have different health behaviors than those who see their GP less often. Second, we explored potential outcome misclassification by restricting the analyses to specific regions of Catalonia where we had access to population-based or hospital cancer registries. We considered as cancer cases only those who had the same diagnosis in the SIDIAP and a cancer registry. Third, we addressed potential reverse causality (i.e., undiagnosed cancer affecting BMI) by extending the minimum follow-up time (of 1 year in the main analyses) to 2 and 4 years. Fourth, we strengthened the validity of our results by performing multiple imputations (using the fully conditional specification approach, with 10 imputed data sets created) to deal with missing values of model 2 covariates [23, 24]. Fifth, we avoided confounding in the analyses of BMI and specific cancer types by re-running model 2, additionally adjusting for HRT use in post-menopausal women [women-only cancers] and excluding participants with a diagnosis of chronic hepatitis B/C [liver cancer risk factor] or a helicobacter pylori infection [stomach cancer risk factor]). Finally, to investigate to which extent the relationships between BMI and risk of each cancer type represents an effect of weight, height, or both weight and height, we re-ran the multivariable-adjusted models (model 2) with height and weight as the main exposures, mutually adjusted for each other.

To assess the robustness of our secondary findings, we performed two sensitivity analyses. We re-ran the analyses that compared BMI and WC in relation to cancer risk with mutual adjustment for both adiposity indicators using residuals of WC and BMI (e.g., we regressed WC on BMI, and we included the residuals from this analysis in the model using BMI as an indicator of general adiposity) to assess if this added valuable information to fully capture adiposity [9]. Finally, we added height as an adjustment variable to the analyses that compared BMI and WC in relation to cancer risk.

The a priori level of statistical significance was set at a 2-sided *P* value of 0.01 for all analyses. We used STATA version 15.1 (College Station, TX, USA) for data analysis and R version 3.5.0 for data visualization.

We obtained approval from the Clinical Research Ethics Committee of the IDIAPJGol (project code: P14/074) to perform this study.

Results

Of the 6,447,722 individuals aged between ≥ 18 and ≤ 100 years in the SIDIAP population, 2,459,462 were excluded due to the unavailability of a valid BMI, 131,167 due to personal history of cancer, and 198,676 due to less than 12 months of follow-up (Fig. 1). A total of 3,658,417 participants constituted the primary dataset of this study for whom follow-up ended at a median of 8.3 years (interquartile range [IQR] 5–11) after study entry. In total, 202,828 [5.6%] individuals were diagnosed with cancer over the study period (Table 1). Among all participants, 55% were women, the median age at inclusion was 46 years (IQR 32–61), and the median BMI was 26.3 kg/m^2 (IQR 23–30). When stratifying participants by WHO categories of BMI, the median follow-up and age increased with increasing categories of BMI. There were fewer participants from deprived areas and more current smokers in the underweight and normal weight category compared to those in the obesity category (Table 1). Compared to the overall SIDIAP adult population, the individuals included in this study were more likely to be women and older, as well as to have more comorbidities and complete information on lifestyle factors (the characteristics of the included and excluded individuals can be consulted in Additional file 1: Table S2).

Non-linear BMI associations and analyses restricted to never smokers

BMI was non-linearly associated with ten of twenty-six cancer types (p for non-linearity < 0.01) (Fig. 2). For cancers of the head and neck, esophagus, stomach, larynx, trachea, bronchus, and lung, low BMI values were

Table 1 Baseline characteristics of the study participants included in the analyses of the first objective (dataset 1) by body mass index categories and of the second objective (dataset 2)

	Dataset 1, *N* (%)				Dataset 2, *N* (%)
	Under and normal weight (BMI < 25)	Overweight (BMI ≥ 25 and < 30)	Obese (BMI ≥ 30)	BMI total	WC total
Characteristic	1,436,991 (39.3)	1,326,642 (36.3)	894,784 (24.4)	3,658,417 (100)	291,305 (100)
Follow-up (in years)[a,b]	7.7 (4.4–10.6)	8.5 (5.0–11.3)	9.1 (5.6–11.8)	8.3 (4.9–11.2)	9.9 (6.9–11.9)
Visits to health center[b,c]	5 (3–9)	7 (4–12)	8 (5–14)	6 (3–11)	10 (6–16)
BMI[b]	22.5 (20.8–23.8)	27.3 (26.1–28.5)	32.9 (31.2–35.6)	26.3 (23.2–29.9)	29.0 (25.9–32.5)
WC[b]	–	–	–	–	100 (91–108)
Age (in years)[b,d]	36 (27–50)	51 (37–65)	55 (42–66)	46 (32–61)	59 (46–71)
Sex					
Men	527,253 (36.7)	707,939 (53.4)	394,948 (44.1)	1,630,140 (44.6)	137,298 (47.1)
Women	909,738 (63.3)	618,703 (46.6)	499,836 (55.9)	2,028,277 (55.4)	154,007 (52.9)
MEDEA deprivation index[e]					
Quintile 1	226,165 (15.7)	180,162 (13.6)	98,899 (11.0)	505,226 (13.8)	32,542 (11.2)
Quintile 2	208,133 (14.5)	189,510 (14.3)	119,236 (13.3)	516,879 (14.2)	39,826 (13.7)
Quintile 3	198,978 (13.9)	192,240 (14.5)	131,763 (14.7)	522,981 (14.3)	40,019 (13.7)
Quintile 4	193,565 (13.5)	195,463 (14.7)	142,024 (15.9)	531,052 (14.5)	41,767 (14.3)
Quintile 5	190,155 (13.2)	185,886 (14.0)	144,632 (16.2)	520,673 (14.2)	34,249 (11.8)
Rural	263,435 (18.3)	251,378 (18.9)	169,075 (18.9)	683,888 (18.7)	73,535 (25.2)
Missing	156,560 (10.9)	132,003 (10.0)	89,155 (10.0)	377,718 (10.3)	29,367 (10.1)
Nationality (geographic region)					
Spain	1,216,424 (84.6)	1,169,166 (88.1)	804,483 (89.9)	3,190,073 (87.2)	271,950 (93,3)
Europe (non-Spanish)	68,689 (4.8)	36,592 (2.8)	22,475 (2.5)	127,756 (3.5)	5560 (1.9)
Africa	15,968 (1.1)	13,208 (1.0)	6379 (0.7)	34,655 (1.0)	1019 (0.4)
America	75,117 (5.2)	63,293 (4.8)	37,867 (4.2)	176,277 (4.8)	7310 (2.5)
Asia	61,693 (4.3)	44,383 (3.3)	23,580 (2.7)	129,656 (3.5)	5466 (1.9)
Smoking status					
Never	674,872 (46.9)	688,304 (51.9)	487,643 (54.5)	1,850,819 (50.6)	174,775 (60.0)
Former	113,105 (7.9)	154,969 (11.7)	106,333 (11.9)	374,407 (10.2)	33,958 (11.6)
Current	438,103 (30.5)	308,376 (23.2)	177,332 (19.8)	923,811 (25.3)	58,468 (20.1)
Missing	210,911 (14.7)	174,993 (13.2)	123,476 (13.8)	509,380 (13.9)	24,104 (8.3)
Alcohol intake					
None	541,451 (37.7)	464,399 (35.0)	315,211 (35.2)	1,321,061 (36.1)	107,230 (36.8)
Low	340,721 (23.7)	325,238 (24.5)	173,382 (19.4)	839,341 (22.9)	60,568 (20.8)
High	25,114 (1.7)	28,074 (2.1)	19,043 (2.1)	72,231 (2.0)	7077 (2.4)
Missing	529,705 (36.9)	508,931 (38.4)	387,148 (43.3)	1,425,784 (39.0)	116,430 (40.0)
Type 2 diabetes	34,847 (2.4)	109,302 (8.2)	123,313 (13.8)	267,426 (7.3)	50,269 (17.3)
Cause of end of follow-up					
End of study	1,170,596 (81.5)	1,037,513 (78.2)	683,190 (76.4)	2,891,299 (79.0)	207,329 (71.2)
Cancer[f]	47,609 (3.3)	87,344 (6.6)	67,875 (7.6)	202,828 (5.6)	27,837 (9.5)
Death	51,777 (3.6)	82,920 (6.2)	70,090 (7.8)	204,787 (5.6)	33,702 (11.6)
Transferred-out	167,009 (11.6)	118,865 (9.0)	73,629 (8.2)	359,503 (9.8)	22,437 (7.7)

BMI body mass index, *MEDEA* "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales", *WC* waist circumference
[a]Participants were followed from the study index date until cancer diagnosis, death, transferal out of the SIDIAP, or until the end of the study period (December 31, 2018)
[b]Median (interquartile range)
[c]Visits to general practitioners and nurses during the year of BMI assessment
[d]At baseline
[e]Quintile 1 represents the least deprived and quintile 5 represents the most deprived. Rural was included as a category since the index cannot be calculated for people living in rural areas
[f]Any, excl. non-melanoma skin cancer

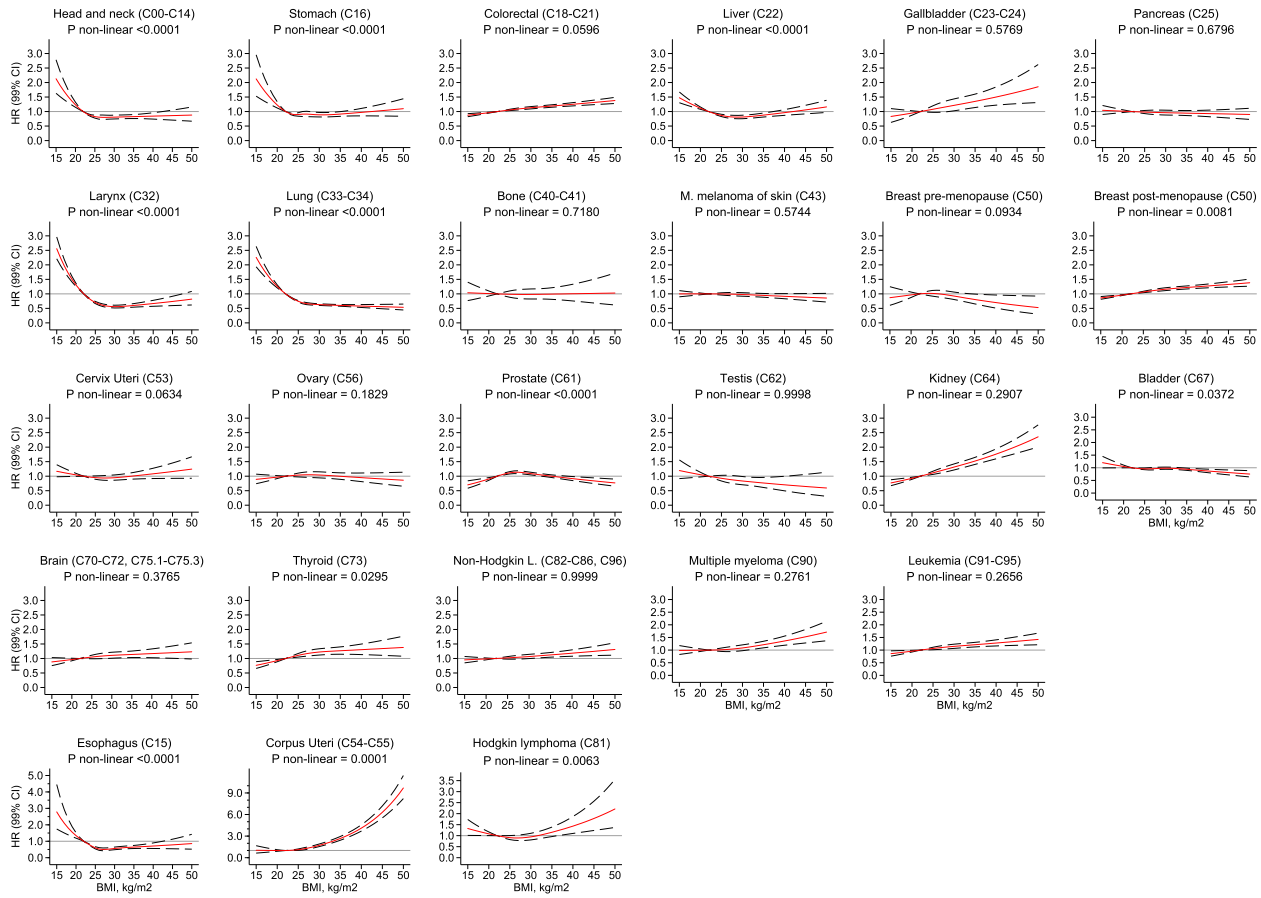


Fig. 2 Association between body mass index and the risk of 26 cancer types in the overall population, allowing for non-linear effects, with 99% CIs. Notes: (1) The reference BMI for these plots was 22 kg/m^2. Separate models were fitted for each cancer type and adjusted for smoking status, alcohol intake, nationality, the MEDEA deprivation index, type 2 diabetes, and had sex and age (5-year categories) in the strata statement. Each model had a restricted cubic spline for BMI with 3 knots placed at 21, 26, and 34 kg/m^2 except for head and neck; stomach; trachea, bronchus, and lung; corpus uteri; and prostate and bladder cancers that had 5 knots placed at 19, 23, 26, 29, and 37 kg/m^2. (2) Gallbladder includes biliary tract; lung includes trachea and bronchus; bone includes articular cartillage; brain includes the CNS, pituitary gland and pineal gland tumors. M. melanoma of skin stands for Malignant melanoma of skin; Non-Hodgkin L. stands for Non-Hodgkin lymphoma. (3) Models for ovary, cervix, and corpus uteri cancers were only computed in women, for breast pre-menopausal only in pre-menopausal women, for breast post-menopausal only in post-menopausal women, and for prostate and testis only computed in men. (4) All models have a scale up to a HR of 3 and are ordered by ascending ranking of ICD-10 codes, except for esophagus, corpus uteri, and Hodgkin lymphoma. Abbreviations: BMI, body mass index; CI, confidence interval; CNS, central nervous system; HR, hazard ratio; KG, kilograms; M, meters

associated with a higher risk of these cancers. The risk stabilized above values of 22 kg/m^2 (with HRs either at or below one). These non-linear relations disappeared when we restricted the analyses to never smokers (Fig. 3).

The curves for the associations between BMI and risk of cancers of the liver, breast post-menopausal, corpus uteri, prostate, and Hodgkin lymphoma were non-linear and were similarly shaped in the overall cohort and among never smokers (Figs. 2 and 3). Liver cancer showed an attenuated U-shaped curve, with a higher risk among participants with very low or very high BMI values. The risk of breast post-menopausal cancer seemed to increase linearly up to a BMI of 30 kg/m^2, at which point the increase in risk diminished. For prostate cancer, the risk curve displayed an attenuated inverse U-shape, with a lower risk of cancer among those with low, normal, and very high BMIs, but an increased risk for those in the overweight range. For corpus uteri cancer, the risk increased faster than linear at higher BMI values. Finally, the association between BMI and Hodgkin lymphoma was J-shaped, with a modest higher risk of this lymphoma in people with low BMIs and a more markedly higher risk for those with high BMIs.

Linear BMI associations and analyses restricted to never smokers

A BMI increment of 5 kg/m^2 (in multivariable analyses) was positively associated with risk of cancers of the corpus uteri (HR 1.49, 99%CI 1.45–1.53), kidney (1.16,

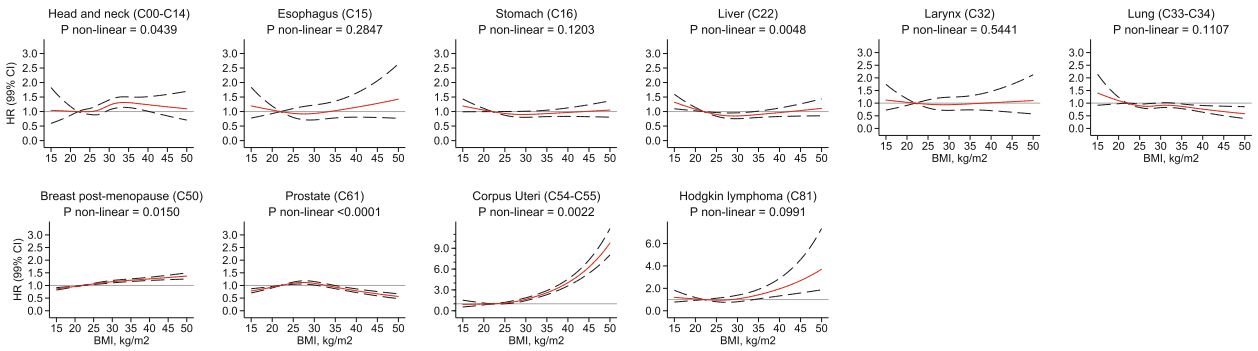


Fig. 3 Association between body mass index and the risk of 10 cancer types in never smokers, allowing for non-linear effects, with 99% CIs. Notes: (1) The reference BMI for these plots was 22 kg/m^2. Separate models were fitted for each cancer type and adjusted for alcohol intake, nationality, the MEDEA deprivation index, type 2 diabetes, and had sex and age (5-year categories) in the strata statement. Each model had a restricted cubic spline for BMI with 3 knots placed at 21, 26, and 34 kg/m^2 except for head and neck, bronchus and lung, and corpus uteri that had 5 knots placed at 19, 23, 26, 29, and 37 kg/m^2. (2) Lung includes trachea and bronchus tumors. (3) The association for corpus uteri cancer was only computed in women, for breast post-menopausal only in post-menopausal women, and for prostate cancer only in men. (4) All models have a scale up to a HR of 3, except for corpus uteri and Hodgkin lymphoma. Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; KG, kilograms; M, meters

1.12–1.20), gallbladder and biliary tract (1.10, 1.03–1.19), multiple myeloma (1.09, 1.04–1.15), thyroid (1.08, 1.03–1.13), leukemia (1.07, 1.04–1.11), colorectal (1.06, 1.04–1.08), breast post-menopausal (1.07, 1.05–1.08), and non-Hodgkin lymphoma (1.04, 1.01–1.08) (Fig. 4). Results from the basic model are presented in Additional file 1: Table S3. Results for corpus uteri and breast-postmenopausal cancers should be interpreted in combination with the splines of Fig. 3 due to the evidence of non-linearity. For the five cancer types (trachea, bronchus and lung, larynx, esophagus, head and neck, and prostate) for which we observed an inverse association between BMI and cancer risk, there was evidence of non-linearity as shown in Fig. 3. After restricting the analyses to never smokers, BMI remained inversely associated only with risk of prostate cancer (0.95, 0.92–0.98), but became positively associated with risk of Hodgkin lymphoma (1.16, 1.01–1.35), and cancers of the head and neck (1.09, 1.03–1.16), and brain and CNS (1.07, 1.00–1.10).

BMI and WC comparison in relation to cancer risk

Of the 291,305 participants who also had a WC assessment available, 27,837 were diagnosed with cancer from 2007 to 2018 (Table 1). Among eligible participants, the median follow-up time was 9.9 (IQR 7–12) years and the median age was 59 (IQR 46–71) years. The median WC was 100 (IQR 91–108) cm and the median BMI was 29 (IQR 26–33) kg/m^2. Compared to the overall BMI cohort, these participants were older and had a higher median BMI and a higher prevalence of type 2 diabetes (Table 1).

We ascertained more than 100 cases for all cancers of interest except cancers of the bone and articular cartilage (64 cases), Hodgkin lymphoma (63), testis (52), and breast pre-menopausal (44) (Fig. 5). For all cancer sites, the 99% CIs of the HRs for WC (per 1 SD increase) and BMI overlapped. We observed the largest differences between the WC and BMI effect estimates for cancers of the bladder (HR for BMI 0.97, 99%CI 0.91–1.03; WC 1.04, 0.98–1.10), larynx (HR for BMI 0.77, 99%CI 0.65–0.91; WC 0.91, 0.78–1.06), and trachea, bronchus, and lung (HR for BMI 0.85, 99%CI 0.79–0.91; WC 0.97, 0.90–1.03), although the 99%CIs overlapped. Nonetheless, these results should be interpreted with caution due to evidence of non-linearity in the association between WC and risk of bladder and trachea, bronchus, and lung cancers (Additional file 1: Table S4).

Sensitivity analyses

We assessed the robustness of our results by comparing the HRs of our main analyses to those from sensitivity analyses. We found that the HRs from our primary model (model 2) were similar to those from the sensitivity analyses. The CIs of the sensitivity analyses consistently included the main point estimate with only two exceptions (Additional file 1: Tables S5-S8). In the analysis in which we extended the minimum follow-up time from 1 to 4 years, the HRs from the main model for stomach and trachea, bronchus, and lung cancers (1-year follow-up) were not included in the CIs from the models with a 4-year minimum follow-up (stomach cancer with 1-year follow-up HR 0.99, 99%CI 0.99–1.00, vs. 4-year follow-up HR 1.01, 99%CI 1.00–1.01; trachea, bronchus, and lung cancer with 1-year follow-up HR 0.96, 99%CI 0.96–0.97, vs. 4-year follow-up HR 0.97, 99%CI 0.97–0.97; all HRs are per 1 kg/m^2 increment in BMI) (Additional file 1: Table S5). We also re-ran the multivariable-adjusted models (model 2) using height on one hand and weight on the other as the main exposures

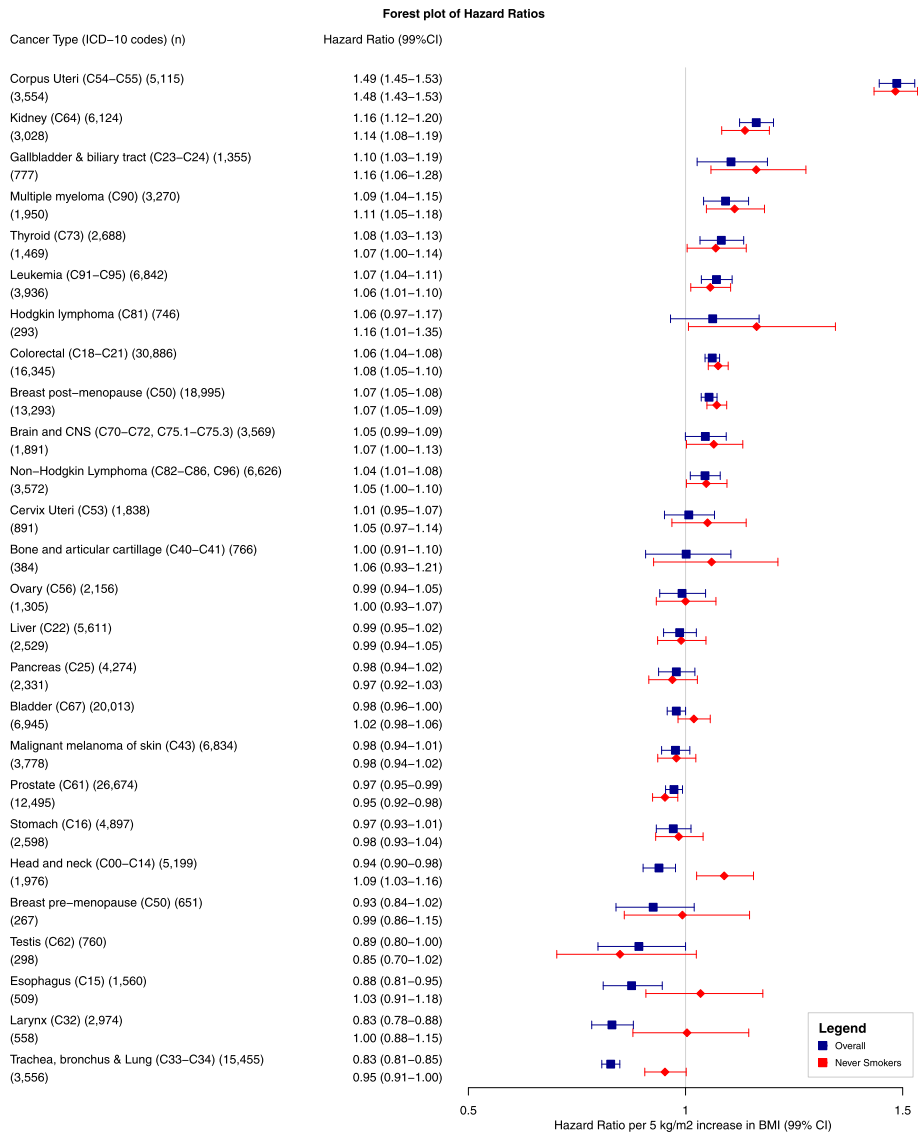


Fig. 4 Forest plot of hazard ratios of 26 cancer types related to a linear increment in body mass index of 5 kg/m^2 with 99% CIs, in the overall population, and in never smokers. Notes: (1) Separate models were fitted for each cancer type and adjusted for smoking status, alcohol intake, nationality, the MEDEA deprivation index, type 2 diabetes, and had sex and age (5-year categories) in the strata statement. (2) Cancer types are ordered by descending ranking. (3) Brain and CNS include pituitary gland and pineal gland tumors. (4) Models for ovary, cervix, and corpus uteri cancers were only computed in women, for breast pre-menopausal only in pre-menopausal women, for breast post-menopausal only in post-menopausal women, and for prostate and testis only computed in men. Abbreviations: BMI, body mass index; CI, confidence interval; CNS, central nervous system; KG, kilograms; M, meters; WC, waist circumference

(Additional file 1: Table S9). The nine cancer types that were positively associated with BMI were also all positively associated with weight, while six were so with height (colorectal, breast post-menopausal, kidney, thyroid, non-Hodgkin lymphoma, and leukemia). Corpus uteri cancer was negatively associated with height. The five cancer types for which we found a negative association with BMI were also negatively associated with weight while two of these were positively associated with height (trachea, bronchus, and lung and prostate cancers).

Furthermore, in the analysis comparing WC and BMI in relation to cancer risk, we assessed whether adding the residuals of the complementary adiposity indicator added valuable information to fully capture adiposity. This was not the case as the 99%CIs of the models comprising residuals always included the HRs from the main models (Additional file 1: Fig. S2). For example, for corpus uteri cancer, the model that only included BMI (HR 1.60, 99%CI 1.47–1.74) was similar to the one that included BMI and the residuals of WC (HR 1.61, 99%CI

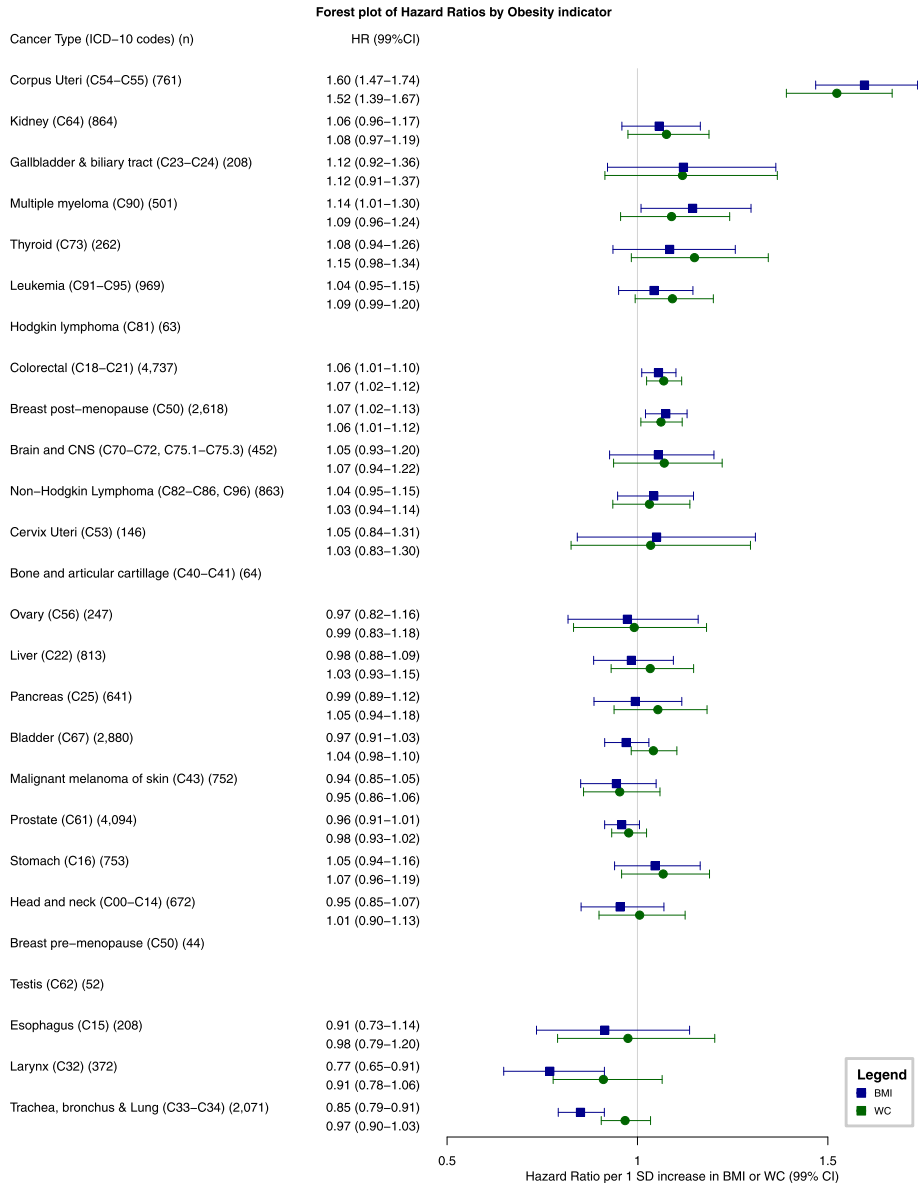


Fig. 5 Forest plot of hazard ratios of 22 specific cancer sites related to a 1 standard deviation increase in body mass index and a 1 standard deviation increase in waist circumference. Notes: (1) SD for BMI and WC were 5.3 and 13.9 overall, 5.8 and 14.5 for women, 6.5 and 16.1 for pre-menopausal women, 5.4 and 13.3 for post-menopausal women, and 4.7 and 12.9 for men. (2) Separate models were fitted for each cancer type and adjusted for smoking status, alcohol intake, nationality, the MEDEA deprivation index, type 2 diabetes, and had sex and age (5-year categories) in the strata statement. (3) HRs are ordered by the descending ranking of BMI estimates from Fig. 4. (4) Brain and CNS include pituitary gland and pineal gland tumors. (5) Models for ovary, cervix, and corpus uteri cancers were only computed in women, for breast post-menopausal only in post-menopausal women, and for prostate only computed in men. (6) We only calculated hazard ratios for cancer types for which we ascertained at least 100 cancer cases. Abbreviations: BMI, body mass index; CI, confidence interval; CNS, central nervous system; SD, standard deviation; WC, waist circumference

1.48–1.76); the same was observed for the model that only included WC (HR 1.52, 99%CI 1.39–1.67) and the one that included WC and the residuals of BMI (HR 1.53, 99%CI 1.39–1.68). The CIs of the sensitivity analysis that further adjusted for height also consistently included the main point estimate of the main analyses comparing WC and BMI in relation to cancer risk (Additional file 1: Table S10).

Discussion

Main findings

In this prospective study that included 3,658,417 participants and 202,837 cancer cases, we found that a higher BMI was associated with risk of 18 of 26 cancer types, although these relations differed in terms of direction, shape, and smoking status at baseline. BMI was positively associated with risk of cancers of the *corpus uteri*,

kidney, *gallbladder and biliary tract*, *thyroid*, *colorectum*, *breast post-menopausal*, *multiple myeloma*, *leukemia*, and *non-Hodgkin lymphoma* (in descending order of linear effect sizes). After restricting the analyses to never smokers to account for incomplete adjustment for smoking, BMI was also positively associated with *Hodgkin lymphoma* and cancers of the *head and neck*, and *brain and CNS*. BMI was associated in an inverse U-shaped manner with the risk of prostate cancer and in an L-shaped fashion with the risk of four cancers (*head and neck*, *esophagus*, *larynx*, and *trachea*, *bronchus, and lung*) in the overall cohort likely indicating residual confounding by smoking since the shape of these associations drastically changed among never smokers, except for prostate cancer.

In a subsample of 291,305 participants with a WC measurement and 27,837 cancer cases, we compared cancer risk estimates of WC and BMI. The 99% CIs of the WC and BMI effect estimates consistently overlapped, indicating that WC provides risk associations similar to BMI across a wide range of cancer types in our population.

Strengths and limitations of this study

This study has several strengths. Firstly, to our knowledge, this is the first study to systematically compare both BMI and WC indicators in relation to the risk of a wide variety of cancers, including less frequently occurring ones. Secondly, owing to the large scale of the SIDIAP database, we were able to investigate the association between BMI and numerous cancer types in a Southern European region, increasing the external validity of results previously reported in Northwestern European countries [3, 4]. Lastly, we previously demonstrated the high quality of cancer diagnoses in the SIDIAP data and we conducted sensitivity analyses in regions where we could include cancer cases confirmed by population-based cancer registries (Additional file 1: Table S6) [18].

This study also has limitations. Firstly, the inclusion of individuals with a BMI measurement (62% of the SIDIAP adult population) could result in selection bias. However, the study participants were not substantially different from the overall SIDIAP population (Additional file 1: Table S2). Secondly, although we cannot exclude the possibility of exposure misclassification, we were empirically reassured that this was not a serious bias. The distribution of BMI in the SIDIAP was similar to population-based survey data and representative studies of the Spanish population (Additional file 1: Table S11). Thirdly, outcome misclassification could have biased our results towards the null because modest positive predictive values have been reported in a validation study of SIDIAP cancer diagnoses [18]. Fourth, residual confounding is an inherent limitation of observational studies; an example in our study was residual confounding for smoking status at baseline. Fifth, we did not have data on factors in the possible causal path between obesity and cancer, such as specific reproductive variables (e.g., parity, breastfeeding history), physical activity, and diet. Neither did we have information on cancer subtype or stage at diagnosis, which could have helped sharpen the analyses for certain cancers (e.g., *prostate* cancer). Fifth, while the magnitude of this study's sample size has its advantages, some of the significant findings of this study could have been related to the large sample size. Another limitation was the missing covariate data which ranged from 10% (for the MEDEA deprivation index) to 39% (for alcohol intake risk). However, the results from our main analysis did not differ when we performed multiple imputations of these data (Additional file 1: Table S5). Finally, we had information for both BMI and WC for only 10% of the study participants. This limited our interpretation of the comparison of adiposity measures associated with cancer risk to individuals with both indicators and does not enable us to extrapolate the WC effect estimates to the general population.

Interpretation and comparison with previous studies

The observed positive associations between BMI and different cancer types are in line with previous studies. The increased risk of breast post-menopausal and corpus uteri cancers has been consistently reported in the literature [25, 26]. Furthermore, our non-linear analyses showed that the higher the BMI, the greater the magnitude of risk of corpus uteri cancer which concurs with previous studies [4, 27]. The positive association between BMI and cancers of the colorectum, kidney, thyroid, and gallbladder and biliary tract is well recognized in the literature; however, nuances by subtype (kidney) [2, 28], histology (thyroid) [29], and sex (colorectal and gallbladder and biliary tract) have been reported [25, 30, 31]. In our data, we observed a stronger effect of BMI for gallbladder and biliary tract cancer in women and colorectal cancer in men, which is in line with previous studies (Additional file 1: Table S12) [25, 31]. Further, our results showed a clear pattern in the association between BMI and hematological cancers. The association observed between BMI and higher risk of leukemia and multiple myeloma has been consistently reported in the literature [25, 32–34], but the association between BMI and the lymphomas is less well established. Although our results for non-Hodgkin lymphoma are supported by two meta-analyses [25, 35], other studies have only reported a link with the subtype of diffuse large B cell lymphoma [36]. For Hodgkin lymphoma, we observed a J-shaped association with BMI, which concurs with a large study from the United Kingdom (UK) [37]. The positive association observed between BMI and cancers

of the brain and CNS might have been driven by the inclusion of meningioma in this broad cancer group [2].

We also observed that the associations between BMI and respiratory tract cancers (head and neck, esophagus, larynx, and trachea, bronchus and lung) were L-shaped, suggesting that low BMIs are an approximation of heavy smoking. In the linear analyses restricted to never smokers, the associations between BMI and cancers of the larynx and esophagus became null, likely due to the opposite effects of BMI in adenoma and squamous cell carcinoma [25]. Also, among never smokers, BMI became positively associated with cancer of the head and neck and remained negatively associated with cancer of the trachea, bronchus, and lung, which concurs with other meta-analyses [25, 38–40]. For prostate cancer, we found an attenuated inverse U-shaped association which coincided with a large UK study [4]. The shape of this association could be explained by the dual effect of BMI on prostate cancer (inversely and positively associated with localized and advanced prostate cancer, respectively) [41]. Unfortunately, we did not have data on prostate cancer subtypes to test this hypothesis.

There were also differences between our results and those of previous studies. Despite the evidence supporting the inverse association between BMI and risk of breast pre-menopausal cancer [25], we observed a negative trend only with BMI values greater than 27 kg/m^2. In addition, some studies described a positive association between BMI and cancers of the liver and stomach [42, 43]. Our results suggest these associations are non-linear and similarly shaped to a large UK study (U- and L-shaped for liver and stomach cancers, respectively) [4]. We noted that the non-linear association for stomach resembled the one for respiratory tract cancers, suggesting residual confounding by smoking status for this cancer as well.

In a post hoc analysis, modeling height and weight in mutually adjusted models, we found that the nine and five cancer types that were positively and negatively, respectively, associated with BMI (in linear models) were also all associated with weight in the same directions. On the other hand, height was positively associated with 14 cancer types (and only negatively associated with corpus uteri cancer) (Additional file 1: Table S9). This suggests that the associations observed for BMI (our main analysis) were driven by excess body weight rather than height. Height is a complex exposure and likely reflects the fact that more stem cells are at risk of acquiring driver mutations during cell division over time. A second possible explanation is that a common factor (such as insulin-like growth factor (IGF) 1) directly affects cancer risk as well as increasing height [44].

Finally, our results indicate that BMI and WC have a comparable relationship with cancer risk. The effect estimates of BMI and WC were similar although we observed moderate differences for cancers of the bladder, larynx, and trachea, bronchus, and lung. Contrarily to BMI, WC was not negatively associated with the risk of cancers of the larynx and trachea, bronchus, and lung. We hypothesized that this could be explained by smoking since smokers tend to have a higher WC, more visceral adipose tissue, and leaner body mass [5].

Conclusion

In this large Southern European study, we found that a higher BMI was associated with higher risk of twelve cancer types. We provide novel evidence that higher BMI increases the risk of four hematological and head and neck (only among never smokers) cancers, and we confirmed associations reported in previous studies. Moreover, this study showed that BMI and WC result in comparable estimates of cancer risk associated with adiposity at a population level.

While the observational nature of this study prevents us from making policy and clinical recommendations, our findings reinforce the need for public health strategies focusing on the reduction of obesity for cancer prevention and indicate that assessing obesity-related cancer risk in primary care using BMI may be sufficient.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12916-020-01877-3.

Additional file 1: Appendix 1. definition of menopause and use of hormonal replacement therapy variables. **Appendix 2.** STROBE Statement-Checklist. **Table S1.** diagnostic codes used to define cancer cases. **Table S2.** characteristics of individuals with and without a BMI recorded. **Table S3.** BMI-cancer risk associations: results of the basic adjustment models. **Table S4.** P for non-linearity in WC-cancer risk associations. **Table S5.** A wide range of sensitivity analyses of BMI-cancer risk associations. **Table S6.** Sensitivity analyses of BMI-cancer risk associations using cancer registry data to confirm SIDIAP cases. **Table S7.** Sensitivity analyses of BMI-cancer risk associations excluding subgroups of participants. **Table S8.** Sensitivity analyses of BMI-cancer risk associations for women only cancers. **Table S9.** Sensitivity analysis including results of BMI/height/weight-cancer risk associations. **Table S10.** Sensitivity analysis of BMI/WC-cancer risk associations including additional adjustment for height. **Table S11.** Comparison of BMI information recorded in the SIDIAP and other studies' data. **Table S12.** BMI-cancer risk associations stratified by sex. Figure 1**.** Directed Acyclic Graph that guided our decisions in the control for confounding. Figure 2**.** Sensitivity analysis of BMI/WC-cancer risk associations, including mutual adjustment using residuals of BMI and WC.

Abbreviations

BMI: Body mass index; CI: Confidence interval; CMBD: Minimum Basic DataSet; CNS: Central nervous system; GP: General practitioner; HR: Hazard ratio; HRT: Hormonal replacement therapy; ICD-9: International Classification for Diseases, 9th revision; ICD-10: International Classification for Diseases, 10th revision; MEDEA (deprivation index): "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales"; SD: Standard deviation; SIDIAP: Information System for Research in Primary Care; UK: United Kingdom; WC: Waist circumference; WHO: World Health Organization

Acknowledgements
We would like to thank all primary care health professionals in Catalonia who routinely collected the information needed for this study in electronic health records of the population. We also thank the cancer registries of Girona, Tarragona, and the Hospital del Mar for providing data for one of the sensitivity analyses.

Disclaimer
Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Authors' contributions
All authors were involved in the study design. YD and MR performed the data management. MR conducted the statistical analyses and wrote the first draft of the manuscript. All authors interpreted the results, contributed to drafting the article, and approved the final version of the manuscript.

Funding
Funding [for grant number: 2017/1630] was obtained from Wereld Kanker Onderzoek Fonds (WKOF), as part of the World Cancer Research Fund International grant program. TDS is funded by the Department of Health of the Generalitat de Catalunya, awarded on the 2016 call under the Strategic Plan for Research and Innovation in Health (PERIS) 2016–2020, modality incorporation of scientists and technologists, with reference SLT002/16/00308. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
In accordance with current European and national law, the data used in this study is only available for the researchers participating in this project. Thus, we are not allowed to distribute or make publicly available the data to other parties. However, researchers from public institutions can request data from the SIDIAP and other sources (e.g., Cancer Registries) if they comply with certain requirements. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudes-en/application-proccedure) or by contacting Anna Moleras (amoleras@idiapjgol.org).

Ethics approval and consent to participate
The Clinical Research Ethics Committee of the IDIAPJGol (project code: P14/074) approved this study.

Consent for publication
Not applicable.

Competing interests
All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

Author details
[1]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Gran Via Corts Catalanes, 587 àtic, 08007 Barcelona, Spain. [2]Universitat Autònoma de Barcelona, Campus de Bellaterra, 08193 Bellaterra, Spain. [3]International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert Thomas, 69008 Lyon, France. [4]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain. [5]Department of Epidemiology and Preventive Medicine, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany. [6]Center for Research on Population Health, National Institute of Public Health, Mexico City, Mexico. [7]Hubert Department of Global Health, Emory University, Atlanta, GA, USA.

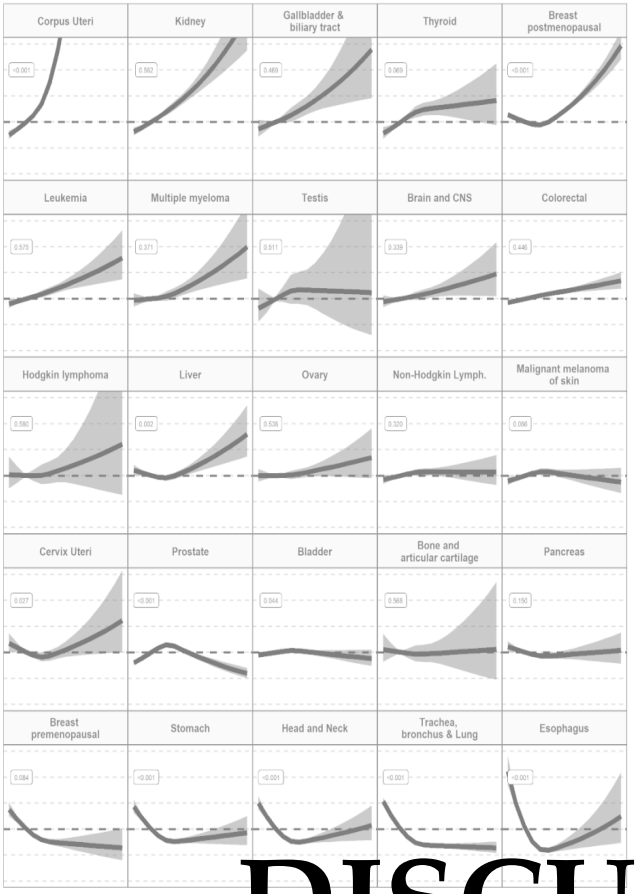
Received: 9 September 2020 Accepted: 1 December 2020
Published online: 14 January 2021

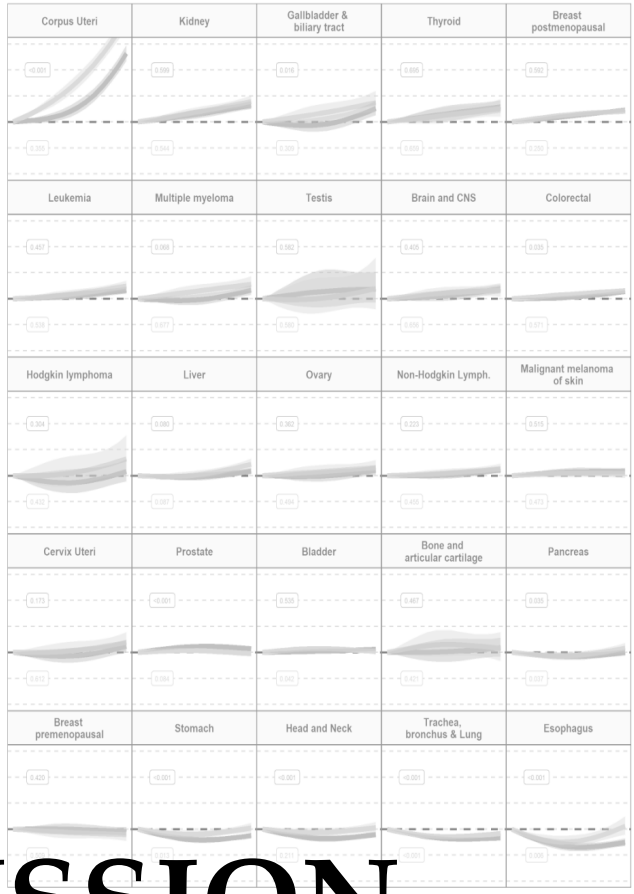
References

1. World Health Organization. Overweight and obesity. 2016 [cited 2018 Nov 5]. Available from: http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
2. Secretan BL, Ph D, Scoccianti C, Ph D, Loomis D, Ph D. Body Fatness and Cancer - Viewpoint of the IARC Working Group. Vol. 375, The New England Journal of Medicine. 2016.
3. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. Br Med J. 2007;335(7630):1134–9.
4. Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. Lancet. 2014;384(9945):755–65.
5. Song M, Giovannucci E. Estimating the influence of obesity on cancer risk: stratification by smoking is critical. J Clin Oncol. 2016;34(27):3237–9.
6. De Ridder J, Julián-Almárcegui C, Mullee A, Rinaldi S, Van Herck K, Vicente-Rodríguez G, et al. Comparison of anthropometric measurements of adiposity in relation to cancer risk: a systematic review of prospective studies. Cancer Causes Control. 2016;27:291–300.
7. White AJ, Nichols HB, Bradshaw PT, Sandler DP. Overall and central adiposity and breast cancer risk in the sister study. Cancer. 2015;121(20):3700–8.
8. Pischon T, Lahmann PH, Boeing H, Friedenreich C, Norat T, Tjønneland A, et al. Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). JNCI J Natl Cancer Inst. 2006;98(13):920–31.
9. Freisling H, Arnold M, Soerjomataram I, O'Doherty MG, Ordóñez-Mena JM, Bamia C, et al. Comparison of general obesity and measures of body fat distribution in older adults in relation to cancer risk: meta-analysis of individual participant data of seven prospective cohorts in Europe. Br J Cancer. 2017;116(11):1486–97.
10. Barberio AM, Alareeki A, Viner B, Pader J, Vena JE, Arora P, et al. Central body fatness is a stronger predictor of cancer risk than overall body size. Nat Commun. 2019;10(1):383.
11. García-Gil MDM, Hermosilla E, Prieto-Alhambra D, Fina F, Rosell M, Ramos R, et al. Construction and validation of a scoring system for selection of high quality data in a Spanish population primary care database (SIDIAP). Inf Prim Care. 2012;20(2):1–1.
12. Bolíbar B, Fina Avilés F, Morros R, Del Mar G-GM, Hermosilla E, Ramos R, et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. Med Clin (Barc). 2012;138(14):617–21.
13. Generalitat de Catalunya. Conjunt mínim bàsic de dades (CMBD). 2017 [cited 2019 Mar 5]. Available from: https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/index.html#googtrans(ca%7Ces).
14. Lecube A, Monereo S, Rubio MÁ, Martínez-de-Icaya P, Martí A, Salvador J, et al. Prevención, diagnóstico y tratamiento de la obesidad. Posicionamiento de la Sociedad Española para el Estudio de la Obesidad de 2016. Endocrinol Diabetes y Nutr. 2017;64:15–22.
15. Institut Català de la Salut. Guies de pràctica cliń ica: Abordatge de la diabetis mellitus tipus 2. 2015. [cited 2019 Mar 5]. Available from: http://ics.gencat.cat/web/.content/documents/assistencia/gpc/GuiaDiabetis2015.pdf.
16. Domínguez-Berjón MF, Borrell C, Cano-Serral G, Esnaola S, Nolasco A, Pasarín MI, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). Gac Sanit. 2008;22(3):179–87.
17. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004;4(8):579–91.
18. Recalde M, Manzano-Salgado C, Díaz Y, Puente D, Garcia-Gil M del M, Marcos-Gragera R, et al. Validation of cancer diagnoses in electronic health records: results from The Information System For Research In Primary Care (SIDIAP) in Northeast Spain. Clin Epidemiol 2019;11:1015–1024.
19. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. Epidemiology. 1999;10(1):37–48.
20. Harrell FEJ. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
21. Orsini N, Greenland S. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. Stata J 2011;11(1):1–29.

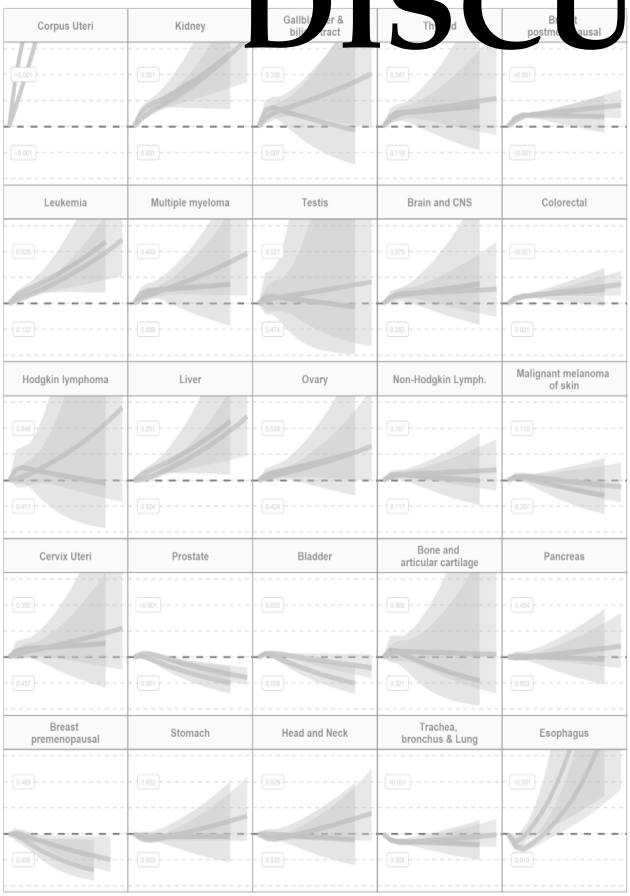
22. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515–26.
23. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66.
24. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol. 2010; 171(5):624–32.
25. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet. 2008;371(November):569–78.
26. Aune D, Navarro Rosenblatt DA, Chan DSM, Vingeliene S, Abar L, Vieira AR, et al. Anthropometric factors and endometrial cancer risk: a systematic review and dose–response meta-analysis of prospective studies. Ann Oncol. 2015;26(8):1635–48.
27. Crosbie EJ, Zwahlen M, Kitchener HC, Egger M, Renehan AG. Body mass index, hormone replacement therapy, and endometrial cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2010;19(12):3119 LP – 3130.
28. Callahan CL, Hofmann JN, Corley DA, Zhao WK, Shuch B, Chow W-H, et al. Obesity and renal cell carcinoma risk by histologic subtype: a nested case-control study and meta-analysis. Cancer Epidemiol. 2018;56:31–7.
29. Schmid D, Ricci C, Behrens G, Leitzmann MF. Adiposity and risk of thyroid cancer: a systematic review and meta-analysis. Obes Rev. 2015;16(12):1042–54.
30. Campbell PT, Newton CC, Kitahara CM, Patel AV, Hartge P, Koshiol J, et al. Body size indicators and risk of gallbladder cancer: pooled analysis of individual-level data from 19 prospective cohort studies. Cancer Epidemiol Biomarkers Prev. 2017;26(4):597 LP – 606.
31. Abar L, Vieira AR, Aune D, Sobiecki JG, Vingeliene S, Polemiti E, et al. Height and body fatness and colorectal cancer risk: an update of the WCRF-AICR systematic review of published prospective studies. Eur J Nutr. 2018;57(5):1701–20.
32. Abar L, Sobiecki JG, Cariolou M, Nanu N, Vieira AR, Stevens C, et al. Body size and obesity during adulthood, and risk of lympho-haematopoietic cancers: an update of the WCRF-AICR systematic review of published prospective studies. Ann Oncol Off J Eur Soc Med Oncol. 2019;30(4):528–41.
33. Wallin A, Larsson SC. Body mass index and risk of multiple myeloma: a meta-analysis of prospective studies. Eur J Cancer. 2011;47(11):1606–15.
34. Larsson SC, Wolk A. Overweight and obesity and incidence of leukemia: a meta-analysis of cohort studies. Int J Cancer. 2008;122(6):1418–21.
35. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: a meta-analysis of prospective studies. Eur J Cancer. 2011;47(16):2422–30.
36. Willett EV, Morton LM, Hartge P, Becker N, Bernstein L, Boffetta P, et al. Non-Hodgkin lymphoma and obesity: a pooled analysis from the InterLymph Consortium. Int J Cancer. 2008;122(9):2062–70.
37. Strongman H, Brown A, Smeeth L, Bhaskaran K. Body mass index and Hodgkin's lymphoma: UK population-based cohort study of 5.8 million individuals. Br J Cancer. 2019;120(7):768–70.
38. Gaudet MM, Kitahara CM, Newton CC, Bernstein L, Reynolds P, Weiderpass E, et al. Anthropometry and head and neck cancer:a pooled analysis of cohort data. Int J Epidemiol. 2015;44(2):673–81.
39. Yang Y, Dong J, Sun K, Zhao L, Zhao F, Wang L, et al. Obesity and incidence of lung cancer: a meta-analysis. Int J Cancer. 2013;132(5):1162–9.
40. Duan P, Hu C, Quan C, Yi X, Zhou W, Yuan M, et al. Body mass index and risk of lung cancer: systematic review and dose-response meta-analysis. Sci Rep. 2015;5:16938.
41. Discacciati A, Orsini N, Wolk A. Body mass index and incidence of localized and advanced prostate cancer—a dose–response meta-analysis of prospective studies. Ann Oncol. 2012;23(7):1665–71.
42. Chen Y, Liu L, Wang X, Wang J, Yan Z, Cheng J, et al. Body mass index and risk of gastric cancer: a meta-analysis of a population with more than ten million from 24 prospective studies. Cancer Epidemiol Prev Biomarkers. 2013;22(8):1395–408.
43. Chen Y, Wang X, Wang J, Yan Z, Luo J. Excess body weight and the risk of primary liver cancer: an updated meta-analysis of prospective studies. Eur J Cancer. 2012;48(14):2137–45.
44. Giovannucci E. A growing link—what is the role of height in cancer risk? Br J Cancer. 2019;120(6):575–6.

Publisher's Note


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

DISCUSSION


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

6. Discussion

In this Thesis we investigated the association between adiposity and cancer risk. Adiposity was captured mainly with the BMI indicator, but also with WC in a subset of the study population (both measured by trained healthcare personnel). As a first step, we evaluated the suitability of a large EHR database from Catalonia, Spain for research and, more specifically, for cancer-related research. In Sections 5 (Results) and 9 (Appendices, where we included the manuscripts corresponding to studies IV and V) we presented the results and preliminary results, respectively, as well as the detailed discussion of the findings of each specific objective of this Thesis. Therefore, in this section, we provide a general discussion of the Thesis, starting with a broad overview of the main findings, followed by its strengths and limitations, contributions to the current knowledge, implications for public health, and ending with future research possibilities.

6.1. Main findings

6.1.1. The SIDIAP database for epidemiological research

We assessed the suitability of SIDIAP for general health-related and cancer-related research. More specifically, in Study I, we explained what the SIDIAP database is and we described the overall characteristics of the SIDIAP population. SIDIAP is a database of population-based primary care EHRs containing pseudo-anonymized records for >8 million individuals since 2006 (median follow-up time: 15.5 years), with 5.8 million individuals active in June 2021 (75% of the Catalan population). The SIDIAP population is representative of the general population living in Catalonia in terms of age, sex, and geographic distribution. The large majority of the SIDIAP population is of Spanish nationality, lives in urban areas, and resides in the Barcelona region. The database includes high-quality data on demographics, all-cause mortality, disease diagnoses, prescription and dispensation of drugs, laboratory tests, socioeconomic indicators, lifestyle information, and clinical parameters, among others. SIDIAP can be linked on a project-by-project basis to other data sources such as hospital discharges, mental health centers, or specific disease registries. In Study II, we assessed the suitability of the SIDIAP database for cancer research by validating site-specific cancer diagnoses recorded in SIDIAP using the data of the two regional population-based cancer registries that exist in Catalonia as the gold-standard. The sensitivities of the SIDIAP cancer cases were above 60% for 23 out of 25 cancer types and the PPV estimates were generally

lower than the sensitivities observed in most cancer types. While more cases were recorded first in the cancer registries compared to SIDIAP, the time difference between both data sources generally did not exceed three months. Including cancer diagnoses from hospital discharge data considerably improved the reliability of SIDIAP data for specific cancer types.

6.1.2. Adiposity and cancer risk association

We investigated the association between adiposity and cancer risk as well as the role of cardiometabolic conditions in this link. In Study III, we investigated associations between baseline BMI and the risk of 26 types of cancer accounting for non-linearity and residual confounding by smoking status. That study included 3,658,417 participants and revealed that higher levels of BMI were associated with the risk of 18 out of 26 cancer types. However, these relations differed in direction, shape, and smoking status at baseline. BMI was positively associated with risk of cancers of 12 cancer types (corpus uteri, kidney, gallbladder and biliary tract, thyroid, colorectum, breast post-menopausal, multiple myeloma, leukemia, and non-Hodgkin lymphoma as well as Hodgkin lymphoma and cancers of the head and neck, and brain and CNS [the latter three cancers only among never smokers]). BMI was associated in an inverse U-shaped fashion with the risk of prostate cancer and in an L-shaped fashion with the risk of cancers of the head and neck, esophagus, larynx, and trachea, bronchus, and lung in the overall cohort. The latter findings likely indicated residual confounding by smoking amount since the shape of these associations drastically changed among never smokers, except for prostate cancer. In study III we also compared cancer risk estimates of WC and BMI in a subsample of 291,305 participants with a WC measurement. The 99% CIs of the WC and BMI effect estimates consistently overlapped each other, indicating that WC provides risk associations similar to those of BMI across a wide range of cancer types.

We also aimed to investigate longitudinal BMI-derived exposures (with data on baseline BMI for comparison) in relation to cancer risk. Study IV (which is available in Appendix 1 as a manuscript submitted to a scientific journal) was a population-based cohort study that included 2,645,885 individuals aged 40 years or older living in Catalonia, Spain, where we found that longitudinal BMI-derived exposures and BMI at baseline were positively associated with the risk of 12 cancers (corpus uteri, kidney, gallbladder and biliary tract, multiple myeloma, leukemia, breast postmenopausal, colorectal, liver, thyroid, brain and CNS, as well as head and neck and bladder [among never smokers]). At least one of these longitudinal exposures was additionally positively associated with the risk of six cancer types (ovary, non-Hodgkin

lymphoma, malignant melanoma of skin, prostate, pancreas, and stomach cancers). BMI at index date and BMI duration ≥25kg/m^2 were negatively associated with the risk of stomach and respiratory tract cancers, which likely indicates residual confounding by smoking since these associations disappeared when we restricted these analyses to individuals who never smoked.

Figure 5. Summary of the findings of this Thesis relating different adiposity exposures to the risk of specific cancer types.

Notes: Own elaboration with data from Studies III, IV, and V. Given that Studies IV and V are submitted to scientific journals, respectively, we included them in the Appendices of this Thesis. Cells filled in dark red denote positive linear associations (light red, positive trends) and dark green denote negative linear associations. Letters in the intersection between exposures and cancer types represent the shape of observed non-linear associations (Studies III & IV). Cancer types marked with "*" in Study IV are cancer types positively associated with BMI among never smokers. Outcomes of Study V were not specific cancer types, but a binary variable of obesity-related cancers as defined in the IARC viewpoint report (2016) on the association between excess body fatness and cancer risk.(90) Cells in gray were not analyzed for that study.
Abbreviations: BMI: Body mass index; CNS: central nervous system; CVD: cardiovascular disease; HTN: hypertension; T2DM: type 2 diabetes mellitus

Finally, we investigated whether incident HTN, T2DM, and CVD modify the association between BMI and the risk of obesity-related cancers as well as the joint associations of

overweight/obesity and one or more cardiometabolic conditions with the risk of obesity-related cancer. In Study V (available in Appendix 2 as a manuscript submitted to a scientific journal), which included 1,774,904 individuals aged 40 years or older from the SIDIAP population, we found that the positive association between a higher BMI and the risk of obesity-related cancers was similar among "healthy" (free of cardiometabolic conditions) individuals and those with incident diagnoses of HTN and/or CVD. In contrast, among individuals with incident T2DM, the association between BMI and obesity-related cancer risk was null. We also found that individual cardiometabolic conditions, and combinations thereof, were each independently and positively associated with the risk of obesity-related cancer. A striking finding was that the association of overweight/obesity and CVD (only or in combination with more conditions) with obesity-related cancer risk was greater than the sum of their separate associations. On the contrary, the observed joint association of overweight/obesity and T2DM (only or in combination with other conditions) with obesity-related cancer risk was lower than the sum of their separate associations.

6.2. Strengths and limitations

6.2.1. Data source

The five studies included in this Thesis were performed using data from the SIDIAP database. The first study of this Thesis was conducted in order to understand if the SIDIAP was suitable (ie, what was the data collected and its scope) to achieve the aims of the subsequent specific studies. Not only the database can be considered suitable for general health- and cancer-related research, but SIDIAP has several strengths to investigate the association between adiposity and cancer risk. Firstly, the SIDIAP population is representative of the general population living in Catalonia in terms of age, sex, and geographic distribution which favors the generalisability of the findings of this Thesis to the general population of Catalonia but also to other regions of Spain and other countries with similar sociodemographics. While most studies of the field have been conducted in Northern Europe, this Thesis contributes with evidence from Southern Europe. Secondly, the size of SIDIAP offers the possibility to address research questions that would not be possible to investigate in smaller-sized studies. Studies III to V would not have been feasible in traditional cohorts given that in SIDIAP we had sufficient statistical power to analyze numerous cancer types (including rare ones) and we accounted for the smoking status and the incident cardiometabolic conditions of the study participants. Finally, SIDIAP offers the possibility to be linked to external data sources such as population-based cancer registries.

This allowed us to assess the quality of the SIDIAP cancer diagnoses for 25 cancer types (Study II) and to conduct a sensitivity analysis in Study III in which we restricted the analyses of the baseline BMI-cancer association to regions where we could include cancer cases confirmed by population-based cancer registries.

6.2.2. Exposure assessment

BMI was the main exposure of this Thesis since it was used (directly or indirectly) for Studies III to V. Using BMI as an exposure has several advantages such as being easy to calculate and interpret given its standardized categorization as well as being highly correlated with body fat (assessed with reference methods).(26,37,43,44,49–51) In SIDIAP, 70% of the population has at least one BMI assessment available since 2006 (as seen in Study I). This anthropometric indicator is measured by trained health professionals (GPs and nurses) in routine visits to primary care centers.(207) The characteristics of the population with a BMI assessment are similar to those of the general population of the SIDIAP and the BMI distribution in SIDIAP is similar to population-based survey data and representative studies of the Spanish population, likely reducing the possibility of selection bias or exposure misclassification of the studies using this exposure (Studies III to V). Moreover, in sensitivity analyses of Studies IV and V, we observed that conducting the analyses among individuals with a BMI assessment (complete case approach) or using multiple imputations to also include those without an assessment in their EHRs yield similar results. However, the use of BMI has its limitations such as not being able to distinguish between body fatness and lean body mass and having a differential accuracy to capture body fatness by sex, age, and ethnicity.(37,52) Unfortunately, we did not have access to anthropometric indicators obtained from reference methods as this lack of granularity is one of the trade-offs of using large EHR databases for epidemiologic research.

For Study III, WC was also an exposure of interest. Some of the main strengths of this indicator are its simplicity, inexpensiveness, high correlation with reference methods, and standardization of cut-offs.(37,43,44) Like BMI, WC is routinely measured by trained health professionals who follow a strict measurement protocol.(158) Having information on this indicator for ~300k participants (even if this represents 10% of the study population of Study III) is an advantage of this Thesis. To our knowledge, Study III was the first study to systematically compare the adiposity-cancer association for WC and BMI. Unfortunately, the high possibility of selection bias among individuals with this anthropometric information precluded us from extrapolating the WC effect estimates to the general population and limited

our interpretation of the comparison of adiposity measures associated with cancer risk among individuals with information on both indicators. In addition, WC also has its disadvantages; it is an imperfect indicator of intra-abdominal adipose tissue, the exact point of measurement can vary between centers or measurers, cut-offs vary according to age and ethnicity, and measuring WC in individuals with very high BMIs is difficult.(37,43,44,56–59) Nevertheless, as explained above, not having access to more accurate indicators of fat distribution is an inherent limitation of using EHR data for epidemiological research.

Finally, for Study IV we considered longitudinal exposures derived from BMI measurements) which is a great strength of this Thesis. Capturing long-term exposures might better reflect (than a single point in time measure of BMI) the potential underlying biological mechanisms between long-term exposure to adiposity which can lead to chronic inflammation and altered hormone metabolism, and the increase in the risk of cancer development. Moreover, in Study IV we split the window of exposure (from 18 to 40 years of age) from that of the time-to-event analysis. This allowed us to reduce the likelihood of bias related to the overlap between the two windows (eg, that individuals with very high BMIs who theoretically are at a higher risk of cancer have a shorter exposure period due to having the event earlier than other participants). However, a limitation of the calculation of these exposures was that they were solely based on multiple imputations (given that the longest possible follow-up of an individual in SIDIAP is of 16 years).

6.2.3. Outcome assessment

Study II of this Thesis was conducted to assess the quality of the cancer diagnoses registered in SIDIAP. The results of Study II supported the use of SIDIAP cancer diagnoses for epidemiological research when cancer is the outcome of interest, which is the case of the studies of this Thesis. As cancer outcomes were obtained from primary care and hospital discharge records (and were not individually validated), we cannot discard the possibility of outcome misclassification. However, we do not expect that this misclassification would have been differential according to the exposure, therefore, this likely did not affect our results. On the other hand, while the current literature shows an important role of the subtype, subsite, or stage at diagnosis of specific cancers (eg, esophagus, stomach, or prostate) in the association between adiposity and cancer risk, this information was not available in SIDIAP.(90,215) As mentioned above, the large size of SIDIAP comes, in certain cases, at the expense of data granularity. Another limitation, was the multiple comparison of each paper. While the magnitude of this

study's sample size has its advantages, some of the significant findings of this study could have been related to the large sample size. Finally, in Study V, the outcome of interest was a binary indicator of obesity-related cancers due to the fact that we did not have enough statistical power to look at specific cancers as separate outcomes. The number of at-risk individuals (especially for those with more than one cardiometabolic condition) was modest. However, in a secondary analysis, we explored associations for certain specific cancer types and the associations were consistent (although with wider CIs) with those of obesity related-cancers combined.

6.2.4. Covariate assessment

A limitation of using EHR data for epidemiologic research is the lack of information on confounding factors. For instance, while in SIDIAP there is high-quality data on smoking status, other indicators on smoking amount (such as number/packages of cigarettes smoked per day) are not widely used or collected for most of the SIDIAP population. Given the strong interrelation between body fatness/weight and smoking, in Study III we observed residual confounding by smoking amount in the association between adiposity and the risk of respiratory tract cancers.(106) Nevertheless, thanks to the size of SIDIAP we managed to overcome this limitation by restricting the analyses to never smokers. Another important variable in the possible causal path between adiposity and cancer risk is individual SES.(37) While we did not have that information available in SIDIAP, we tried to partially block this confounding path by adjusting our analyses for the *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales* (MEDEA) deprivation index, an ecological indicator of deprivation. Finally, in SIDIAP there is limited data collected on the physical activity levels and diet. While we cannot discard that some of the observed associations in Studies III to V could be somewhat confounded by these factors, we were reassured that this probably was not the case as other traditional cohort studies (eg, European Prospective Investigation into Cancer and Nutrition or National Health and Nutrition Examination Survey studies) investigating the association between BMI and risk of specific cancer types with data on these variables found similar results to ours.(118,216,217)

6.3. Contributions to the current knowledge

6.3.1. EHR databases for epidemiological research

Studies I and II contributed to the current knowledge with information about the suitability of SIDIAP for research and, more specifically, for cancer-related research. To this date, no studies

have tried to characterize the SIDIAP database in depth despite this being essential to understand if the available data source is suitable to achieve the aims of subsequent specific studies. Some general guidelines of the characteristics of the SIDIAP database have been given in Garcia-Gil et al and Bolibar et al.(199,218) The former included some information about the setting and the amount of data collected for the whole SIDIAP population and explored the representativeness (by age, gender, and geographic distribution) of a subset of the population. The latter, published only in Spanish, gave more insights into the creation of the data resource and the type of data collected for the general population included in SIDIAP and briefly characterized a subset of the population. Therefore, we contributed to the current literature with an exhaustive characterization of the SIDIAP database and its included population. Study I, which is published in English, will allow a broader range of the research community to be aware of the existence of this data resource and will provide enough information to enable researchers to understand the scope of the data resource as well as how to access and make the best use of the data.

In Study II, we provided evidence about the reliability of SIDIAP cancer diagnoses for 25 cancer types which is useful for subsequent observational studies of the field. To this date, there has only been one study conducted in SIDIAP that assessed the validity of specific cancer types, which only included lung, colon and rectum, prostate, breast, and cervix uteri cancers.(219) While this study provided relevant information on the quality of some of the most frequent cancer types in Catalonia, data on more cancer types was lacking. In addition, this study compared SIDIAP cancer cases with those from the registry of a single hospital in Barcelona. Although the data collection for this hospital is rigorous for a specific area in Barcelona, this area is not representative of the general population of Catalonia. Furthermore, the hospital does not have data available for research use on hematological cancers. On the other hand, there are available studies (mostly from the UK) that assessed the quality of other cancer types using cancer registry data as the gold standard, which can be useful as a reference on the registry of cancer cases in primary care databases but does not replace the richness of information that provides a study performed in the database of interest.(220–222) Therefore, Study II contributed to the current knowledge with quantified data about the quality of cancer diagnoses registered in SIDIAP (assessed through sensitivities and PPVs, overall and by sociodemographics) as well as the characteristics of the confirmed and non-confirmed cases in SIDIAP (compared to the cancer registries). In addition, we also assessed the inclusion of data from the hospital setting to the SIDIAP database to check if there were sensitivity changes and

the time difference between the date of diagnosis registered in the SIDIAP and the cancer registries. This broadness of data will allow researchers interested in conducting cancer research with the SIDIAP database to make evidence-based decisions about how to make the best use of the data.

6.3.2. Adiposity and cancer risk association

In Study III we aimed to address whether BMI at baseline is related to the risk of more cancer types than currently acknowledged (in linear and non-linear models) and if the BMI-cancer associations differ by the smoking status of the participants. In addition, whether BMI as a sole indicator of general adiposity fully captures the complex association between adiposity and cancer risk is unclear. Our findings revealed that BMI was positively associated with the risk of 12 cancer types (corpus uteri, kidney, gallbladder and biliary tract, thyroid, colorectum, breast post-menopausal, multiple myeloma, leukemia, and non-Hodgkin lymphoma as well as Hodgkin lymphoma and cancers of the head and neck, and brain and CNS [the latter three cancers only among never smokers]). Some of these associations (for corpus uteri, kidney, gallbladder and biliary tract, thyroid, colorectum, breast post-menopausal, and brain and CNS) have been widely reported in the literature; however, we contributed to the current knowledge with consistent evidence about the relationship between BMI and the risk of four hematological cancers. The association observed between baseline BMI and higher risk of leukemia and multiple myeloma has been reported in the literature;(93,138,223,224) however, the association between BMI and lymphomas is less well established. Our results for non-Hodgkin lymphoma are supported by two meta-analyses,(93,137) but other studies have only reported a link with the subtype of diffuse large B cell lymphoma.(225) For Hodgkin lymphoma, we observed a J-shaped association with BMI, which was also observed in a large study from the UK.(226) While in the overall cohort we observed inverse associations between BMI and risk of respiratory tract cancers; these associations became null when we restricted the analyses to never smokers (except for head and neck that became positive). This analysis in combination with the exploration of non-linearity suggested that low BMIs are an approximation of heavy smoking, and not that adiposity confers a protective effect for these cancers. Finally, for prostate cancer, we provided evidence about the attenuated inverse U-shaped association between BMI and cancer risk which coincided with a large UK study.(105) The shape of this association could be explained by the dual effect of BMI on prostate cancer (inversely and positively associated with localized and advanced prostate cancer risk, respectively).(215)

Unfortunately, we did not have data on prostate cancer subtypes to test this hypothesis. Finally, our results for the comparison between BMI and WC indicated that these exposures have a comparable relationship with cancer risk. The effect estimates of BMI and WC were similar although we observed moderate differences for cancers of the bladder, larynx, and trachea, bronchus, and lung. Contrary to BMI, WC was not negatively associated with the risk of cancers of the larynx and trachea, bronchus, and lung. We hypothesized that this could be explained by smoking since smokers tend to have a higher WC, more visceral adipose tissue, and leaner body mass.(106)

In Study IV, we investigated the association between baseline and longitudinal BMI-derived exposures during early adulthood in relation to the risk of 26 cancers. Given that capturing longitudinal BMI-derived exposures might better reflect the potential underlying biological mechanisms between long-term exposure to adiposity and the increase in the risk of cancer development, we aimed to assess if there were stronger associations between adiposity (assessed longitudinally) and obesity-related cancer risk and if adiposity was linked to a larger number of cancer types than currently recognized. A single measurement of BMI at study baseline has been convincingly associated with the risk of 13 cancer types in previous studies of which 10 (colorectum, liver, gallbladder and biliary tract, post-menopausal breast, corpus uteri, ovary, kidney, brain and CNS, thyroid, and multiple myeloma) cancers were also positively associated with the longitudinal BMI-derived exposures that we investigated.(90) Thus at a population level, not only attained higher levels of BMI at one point in time are positively related to the risk of these cancers, but also longer exposures to overweight and obesity (with or without accounting for the degree of overweight and obesity) as well as developing overweight and obesity at younger ages through early adulthood might be associated with cancer risk. We also provide novel evidence that longitudinal BMI-derived exposures are associated with the risk of leukemia, non-Hodgkin lymphoma, malignant melanoma of skin, prostate, head and neck, and bladder cancers (the latter two were only and more pronouncedly, respectively, associated among never smokers). The IARC viewpoint on the association between excess body fatness and cancer risk considered the evidence as inadequate for leukemia and non-Hodgkin lymphoma.(90) However, as commented before, several studies have reported an association between BMI and higher risk of leukemia and non-Hodgkin lymphoma (or only of diffuse large B cell lymphoma).(93,137,224,225) Our findings support and extend these results by providing evidence that higher levels of adiposity through a life course perspective are consistently associated with the risk of hematological cancers,

including multiple myeloma, leukemia, and non-Hodgkin lymphoma. Furthermore, we showed that among individuals who never smoked, longer exposures to overweight and obesity (with or without accounting for the degree of overweight/obesity) were positively associated with the risk of head and neck and bladder cancers. These are relevant findings not only because they expand on the extent to which adiposity can affect cancer risk but also because, at an epidemiological level, they highlight the importance of accounting for residual confounding by smoking.(106,114) Moreover, only duration of BMI ≥25 kg/m^2 was positively associated with the risk of malignant melanoma of skin and prostate cancers. This is in line with what was observed in the non-linear analysis of the association between baseline BMI and risk of these cancers where an inverted U-shaped association was found, indicating a higher risk of these cancers only for BMIs in the overweight range.(105) Future research should focus on confirming these associations and on understanding the pathways by which only being overweight (and a longer duration of it) could have a harmful effect on the risk of these cancers. As for Study III, residual confounding by smoking was also likely an issue for stomach and respiratory tract cancers in the main analyses (the negative associations observed between certain exposures and cancers disappeared when we restricted the analyses to never smokers). An interesting finding in this regard was that the longitudinal BMI-derived exposures (except for duration of BMI ≥25) seemed less prone to residual confounding by smoking than a single BMI assessment.

Finally, in Study V we aimed to address the extent to which cardiometabolic conditions such as HTN, T2DM, and CVD modify the BMI-cancer association (given that prior studies of the field have mostly focused on healthy or general populations) as well as the joint effect of living with overweight/adiposity and cardiometabolic conditions on cancer risk. To our knowledge, there are no other studies that have investigated this research question, therefore the findings of this study represent novel scientific contributions. We found a positive association between BMI and obesity-related cancers among "healthy" individuals, which is in line with well-established evidence in the field.(90,104,105) This association can be explained by previously reported biological mechanisms.(114,133–136) It has also been suggested that other factors, such as cardiometabolic conditions, could be mediators in the association between body fatness and cancer risk.(114,133,184) However, since the "healthy" population did not include individuals with HTN, T2DM, and CVD by definition, our results support the existence of pathways between body fatness and cancer risk that are independent of these conditions. Further, our results revealed that the BMI-obesity-related cancer association remains present

among individuals with an incident diagnosis of HTN and/or CVD. This observation could be explained by an independent (from HTN and CVD) pathway between BMI and cancer risk, but also by a weaker (compared to, for example, T2DM) association between HTN or CVD and risk of obesity-related cancers (pathway being blocked by conditioning on HTN or CVD). The latter hypothesis was sustained by the more "modest" effect of HTN and CVD on cancer risk among individuals with a BMI<25 kg/m^2 (findings of the secondary objective), compared to that of T2DM. Moreover, the evidence linking HTN and CVD to cancer risk is yet to be well established. Studies investigating HTN as a risk factor for cancer have mostly focussed on kidney cancer, and although a link has also been suggested for stomach, colorectal, pancreas, postmenopausal breast, brain, and malignant melanoma cancer risk (which are also cancer types associated with higher adiposity levels) as well as lung cancer, the evidence supporting these associations is not yet well-established.(90,93,155,156) Similarly, the evidence suggesting that CVD might be an independent risk factor for cancer is still at very early stages.(185) While we did not observe an interaction between BMI and CVD (nor in combination with T2DM or HTN and T2DM together) on the multiplicative scale, we did find ones on the additive scale (ie, in terms of absolute risks). This suggests that a higher incidence of obesity-related cancers can be expected among population sub-groups affected by both overweight/obesity and these (combination of) cardiometabolic conditions. The association between BMI and risk of obesity-related cancers among people with T2DM was null. This could be explained by shared biological pathways underlying the carcinogenesis of obesity-related cancers between adiposity and T2DM as well as by a strong association between T2DM and cancer risk (direct association being blocked by conditioning on T2DM). In meta-analyses, T2DM has been positively associated with the risk of cancers of the stomach, colorectal, liver, gallbladder and biliary tract, pancreas, breast, corpus uteri, kidney, bladder, thyroid, non-Hodgkin lymphoma, multiple myeloma, and leukemia.(167–177) Except for bladder cancer, all other associations are concordant with those described for adiposity.(90,93) Our findings (both for the interactions on the multiplicative and the additive scale) in combination with what has been reported in the literature reinforce the existence of shared mechanisms between adiposity and T2DM in relation to cancer risk. In fact, the pathways that have been proposed to explain the T2DM-cancer associations (hyperinsulinemia, hyperglycemia, IGF signaling, and inflammation) have also been proposed as possible mediators for the BMI-cancer one.(184,189–193)

6.4. Implications of the findings for public health

The results of this Thesis (Studies III to V) reinforce the need to focus on reducing overweight and obesity from a public health perspective. Not only can reducing the prevalence of overweight and obesity help prevent cancer, but also other obesity-related conditions and diseases such as HTN, T2DM, CVD, and hyperlipidemia.(25,26) Furthemore our findings (Studies III and IV) indicate that adiposity is related to more cancer types than currently recognized. Therefore, the current estimations of the population attributable fraction and number of cases attributable to overweight/obesity (eg, 6.3% and 22,761, respectively, in the UK) might be underestimating the true impact of overweight and obesity.(24)

As pictured in the framework of the determinants of obesity described in Section 1.2.3., we can distinguish three types of approaches to address the high prevalence of overweight and obesity: policy interventions focussed on reversing the environmental drivers of obesity, interventions aiming at behavioral changes, and pharmaceutical and surgical interventions targeting physiological changes.(60)

Policy interventions focussed on reversing the environmental drivers of obesity could be the most beneficial to reverse the obesity epidemic given their sustainability, the fact that they encompass the whole population (not only highly motivated groups), and their possibility to become systemic.(60) These interventions include food policies to facilitate healthy choices (eg, subsidizing healthy foods, increasing the taxes of unhealthy foods, or implementation of food and nutrition labeling) and physical activity policies that can promote higher levels of physical activity and reduction of sedentary lifestyle (eg, urban planning policies, transport policies, or provision of spaces to engage in physical activity such as parks or sports centers).(68,227–231) However these types of policies are harder (than interventions targeting behavioral changes) to implement due to factors such as the powerful force of the food lobby or the public reluctance to change behaviors.(60,232,233)

The interventions aimed at motivating behavioral changes can help counteract some of the drivers of obesogenic environments and include social marketing (eg, disseminating evidence-based information about healthy lifestyle or the health consequences of overweight and obesity through newspapers, billboards, or social media) as well as health promotion and educational programs.(60) These programs can take place in schools (eg, teaching evidence-based information about nutrition and physical activity or motivating students to be more active

through physical education), families (which can provide tools to children to make healthier choices), workspaces (eg, facilitating or giving access to healthy foods, sports facilities, or nutritionists), or health care centers. Especially primary care could have an important role in reducing overweight and obesity by increasing the number of nutritionists in primary care centers, giving incentives to GPs and nurses to provide evidence-based information to adult patients about healthy lifestyles and pediatricians to parents, guardians, and/or children, and by carrying out behavioral weight loss programs.

If behavioral weight-loss interventions are not effective, pharmacological and surgical interventions are also a possibility. While focusing on these tertiary prevention interventions should not be the main public health priority, there can be subsidies or shorter waiting lists for patients in need of them (eg, the findings of Study V highlighted that individuals newly diagnosed with CVD who have overweight or obesity might be a subgroup of individuals who would benefit the most regarding weight-loss interventions). Most of the medications currently used work by suppressing the appetite while Orlistat works on preventing the digestion and absorption of some dietary fats and surgical procedures (normally conducted among individuals with BMIs >40kg/m^2 or >35kg/m^2 if they have other conditions) result in weight loss by restricting the size of the stomach or by bypassing a portion of the intestines.(234)

Finally, the findings of this Thesis (especially Studies I and II), reinforce the usefulness of EHRs for providing evidence for public health action. EHRs have several strengths such as their relatively low cost, large size, amount of data availability, representativeness, long-term follow-up, and providing sufficient statistical power to conduct research compared to setting up traditional cohort studies, but most importantly their data are already collected and available for use.(197,198) The latter along with their low cost provide important incentives for promoting the use of EHRs for research in the public sector.

6.5. Recommendations for future research

In this Thesis, we have tried to fill in gaps that we identified in the adiposity-cancer literature; however, other questions still need to be addressed or for which more evidence is needed.

6.5.1. Data sources

Throughout this Thesis we have emphasized the advantages of using EHR databases for research; however, there are some limitations to focussing only on one database. While SIDIAP

provides sufficient statistical power to study numerous cancer types as outcomes and evidence from the South of Europe, future studies could try to have several EHRs or other "real world" databases as data sources from different countries (especially from low- and middle-income countries to promote globally representative research), and to be conducted using a federated analysis approach. For example, the Observational Health Data Sciences and Informatics (OHDSI) is an established open-science international network of researchers and observational health databases with a central coordinating center housed at Columbia University that could help address those limitations.(235) There are currently 115 de-identified healthcare databases already mapped or in progress of mapping to the Observational Medical Outcomes Partnership-Common Data Model (which is maintained by the OHDSI network along with a wide range of tools developed by its members to facilitate analyses of mapped data) from 20 countries across 6 continents and including records from over 1.5 billion individuals.(235,236) The included databases provide health records from different settings, including hospitals, primary care, biobanks or claims. The studies conducted within OHDSI use a federated approach (data does not need to leave local servers) for data analysis which facilitates conducting large-scale studies while respecting the confidentiality of patients' records.

6.5.2. Exposure assessment

As we have mentioned above, BMI is an indirect measure of adiposity, and individuals with the same BMI can have different body compositions.(237) On the contrary, objective measurements of body composition obtained with reference methods such as dual-energy X-ray absorptiometry, computed tomography, and magnetic resonance imaging can provide more accurate measures of body composition and adiposity.(127) While currently, these technologies remain expensive and hard to implement for large-sized studies, their cost and difficulties to be implemented will probably be reduced in the future.(37,43,44) Therefore, future studies could use these indicators as the main exposures. Moreover, if information on these exposures is only available for at least a subset of the population, future studies could be used to assess the robustness of the results of analyses with BMI or other field methods as main exposures in sensitivity analyses.

In this Thesis, we aimed to reduce some of the uncertainties associated with longitudinal BMI-derived exposures in relation to cancer risk (Study IV). However, our analyses were only based on time-raster multiple imputations during early adulthood. Therefore, future studies using data with longer follow-up times and with more observed and spaced BMI assessments as well as

analyzing other vulnerability windows (eg, early life development, childhood, and adolescence or later stages of adulthood) could add valuable information in the understanding of adiposity from a life course perspective in relation to cancer risk. Considering longitudinal exposures is also important to assess changes in body composition. More research is needed to determine if weight loss (differentiating behavioral, pharmacological, and surgical interventions) can also help reduce cancer risk associated with obesity (as tobacco cessation reduces cancer risk).(238)

Furthermore, mendelian randomization studies could also be helpful to increase our understanding in the adiposity-cancer association. Mendelian randomization is an analytic approach which utilizes genetic variation as a randomized instrument of the exposure of interest to provide insights into causality.(239) A recently published systematic review revealed that there is only robust associations between adiposity (measured by anthropometric indices) and the risk of breast, kidney, and endometrial cancers.(239) Therefore further research is needed to confirm associations for other cancer types reported in observational studies, especially those for which evidence is less well established .

6.5.3. Outcome assessment

One of the limitations of our outcome assessment was the lack of data granularity on cancer outcomes. Thus, future studies could aim at a better characterization of cancer outcomes by providing further information on subtype, subsite, or stage at diagnosis of specific cancers, but also on molecular differences as new evidence is showing their importance in the association between lifestyle exposures and cancer risk.(127)

On the other hand, most of the research conducted in the adiposity-cancer field focuses on incident cancer diagnoses as the outcome of interest. More research is needed to understand the impact of adiposity (especially assessed with longitudinal exposures) on outcomes after cancer incidence including recurrence, prognosis, and comorbidities.(127)

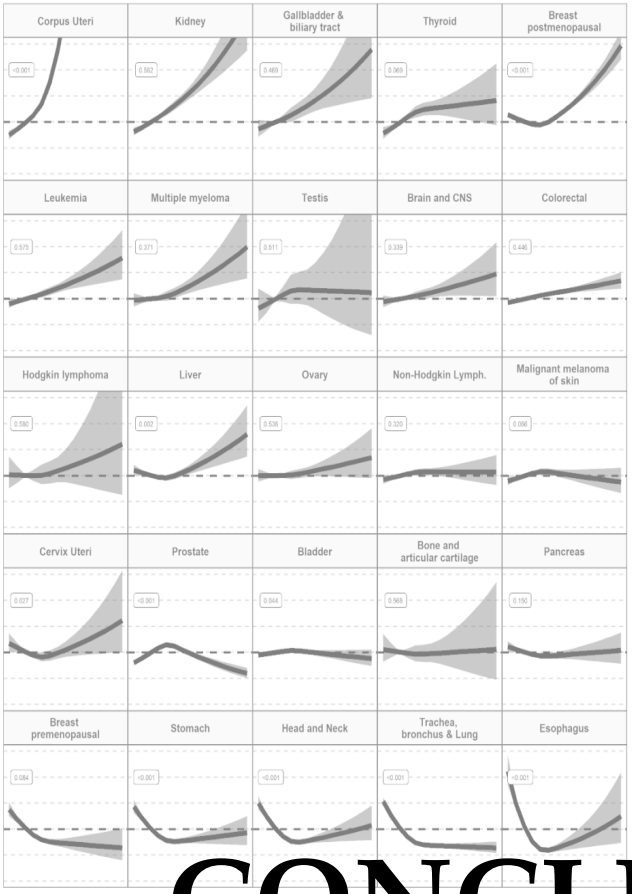
6.5.4. Covariate assessment

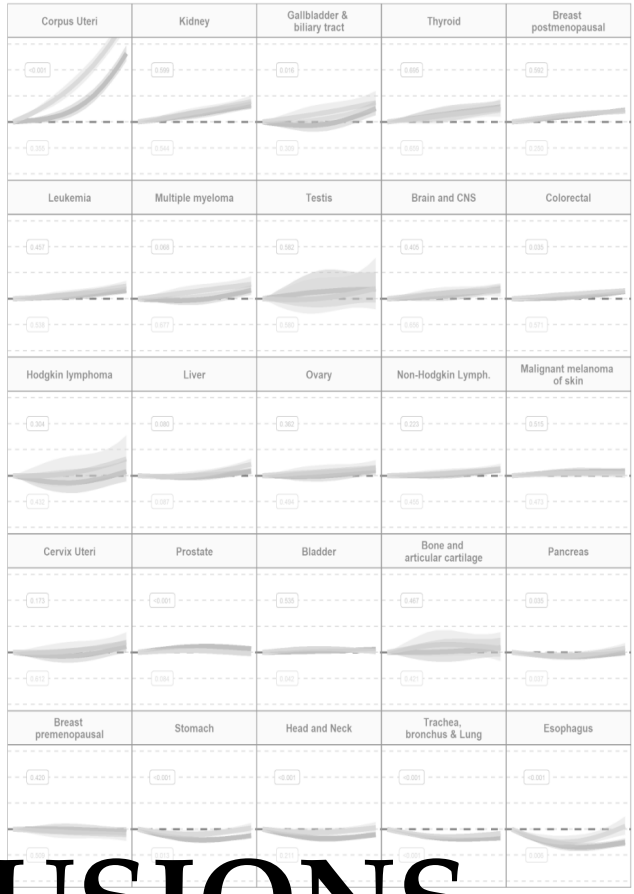
Because EHR databases contain data that is routinely collected in clinical practice for clinical purposes they lack certain data that would have been collected (or collected differently) in a traditional cohort study. As we pointed out, some examples of these data include smoking amount, individual SES, diet, and physical activity variables. However, the nature of EHR databases does not preclude relevant stakeholders of databases from implementing programs

to improve the recording of information of variables considered important for a wide range of studies. For example, the quality of the recording of certain conditions and key information could be enhanced by a program that would incentivize primary care professionals to register them in the EHRs through performance indicator fees.

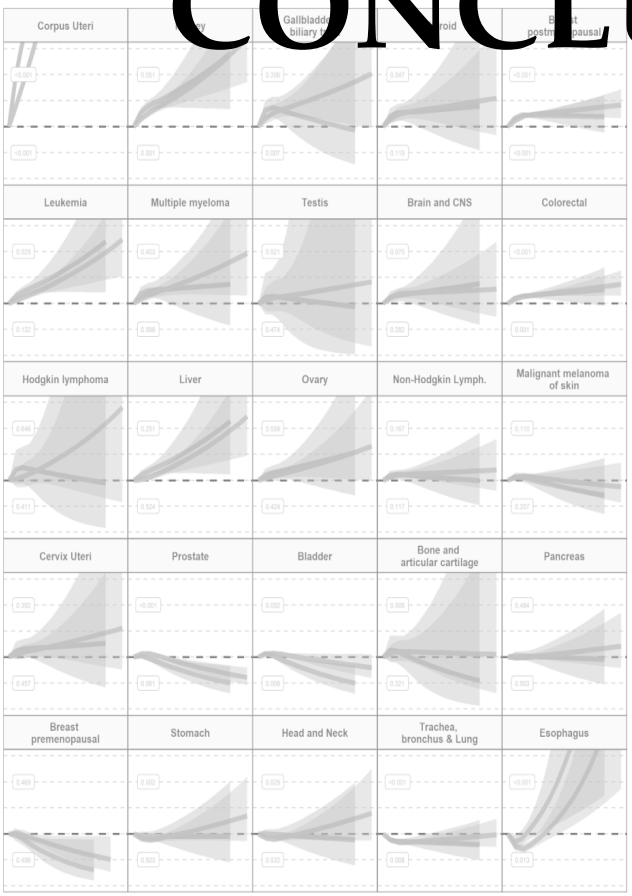
6.5.5. Biological mechanisms

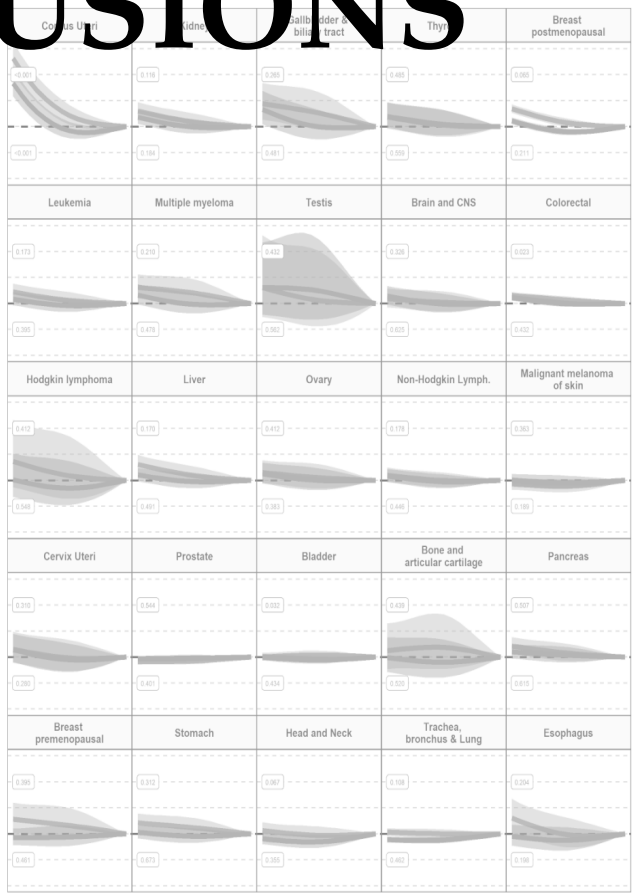
Finally, more evidence is needed to understand the biological mechanisms behind the adiposity-cancer association. As our results have shown, BMI is associated with more cancer types than currently recognized in the literature (four hematological and, only among never smokers, head and neck and bladder cancers); thus studies are needed to understand if these associations can be explained by the mechanisms currently postulated in the adiposity-cancer literature, or if other pathways could be behind this link. Moreover, the findings of Study V showed that the association between BMI and obesity-related cancers disappears once individuals are diagnosed with T2DM. Therefore research on the mechanisms between adiposity, T2DM, and cancer risk is also needed, especially considering the alarming worldwide prevalence of diabetes.(146,159,163,182)





CONCLUSIONS



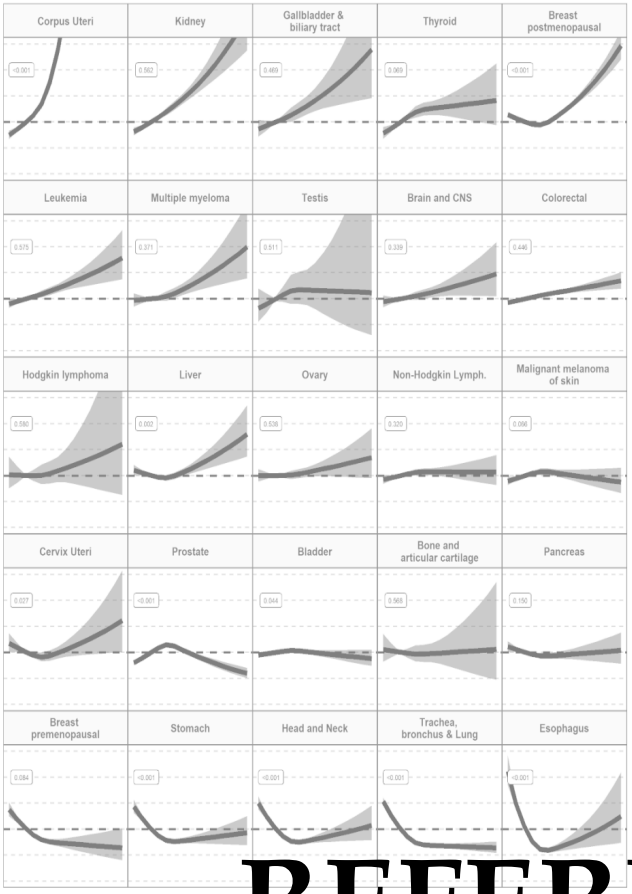


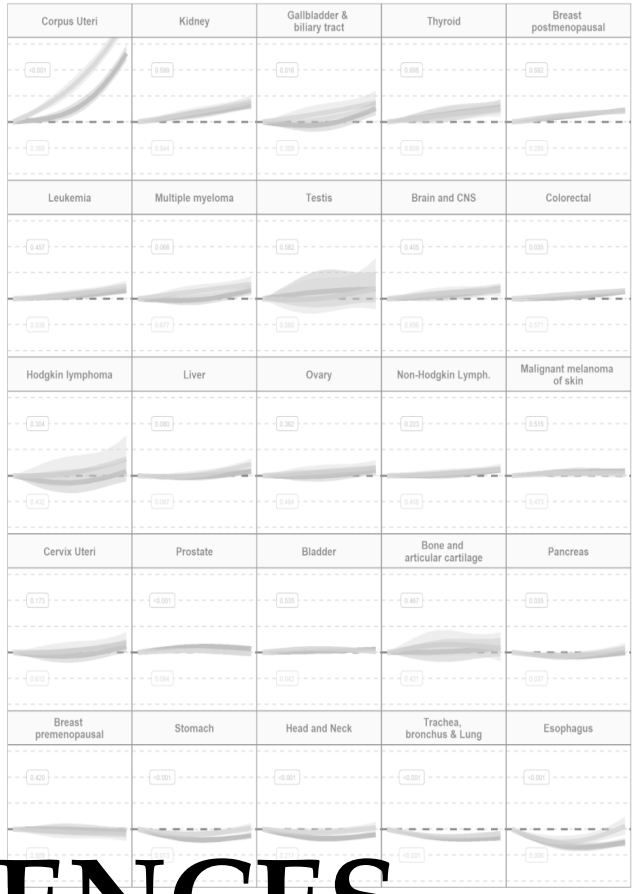
7. Conclusions

SIDIAP is a suitable database to conduct health- and cancer-related research. It contains population-based primary care electronic health records for >8 million individuals since 2006. The SIDIAP population is representative of the general population living in Catalonia in terms of age, sex, and geographic distribution. The database includes high-quality data on sociodemographics, all-cause mortality, disease diagnoses, lifestyle and clinical information, among others, and can be linked to external data sources on a project-by-project basis.

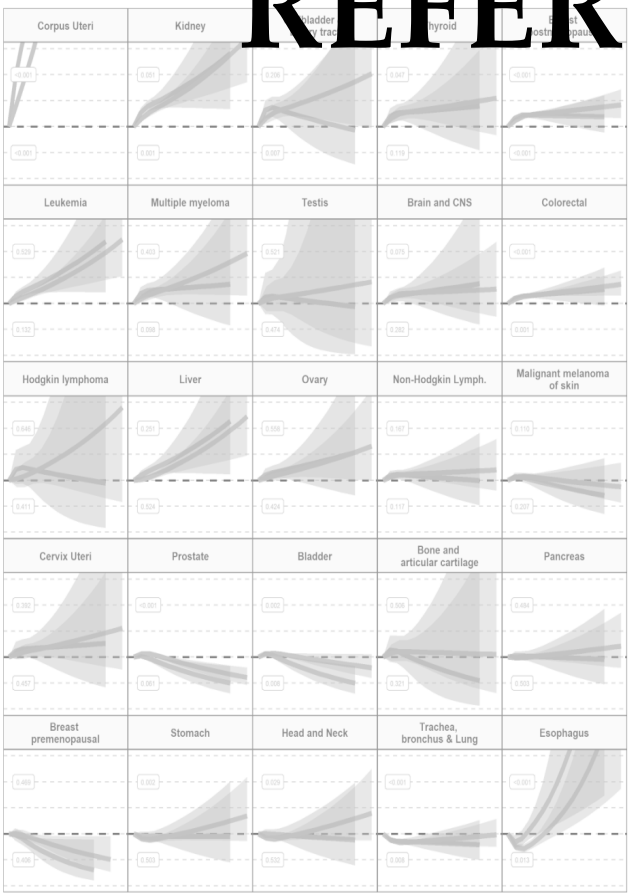
SIDIAP includes 76% of the cancer diagnoses in the population-based cancer registries of Catalonia but includes a considerable number of cases that are not in the registries. SIDIAP reports most of the cancer diagnoses three months difference from the date of diagnosis in the cancer registries. Our results support the use of the SIDIAP cancer diagnoses for epidemiological research when cancer is the outcome of interest. We recommend adding hospital discharge data to the SIDIAP to increase data quality, particularly for less frequent cancer types.

Adiposity is associated with an increased risk of several cancer types. We confirmed associations previously reported in studies focusing on baseline BMI and we provide novel evidence that higher and longer exposure to adiposity increases the risk of four hematological as well as head and neck and bladder (among never smokers) cancers. The BMI-cancer association is similar among individuals free of cardiometabolic conditions and those with incident hypertension and/or cardiovascular disease but it is attenuated among individuals with type 2 diabetes mellitus. BMI and waist circumference result in comparable estimates of cancer risk associated with adiposity at a population level.


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

REFERENCES


Corpus Uteri
Kidney
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

8. References

1. World Health Organization. Cancer [Internet]. 2018 [cited 2021 Aug 16]. Available from: https://www.who.int/health-topics/cancer#tab=tab_1

2. National Cancer Institute. What is cancer? [Internet]. 2021 [cited 2021 Aug 16]. Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer

3. International Agency for Research on Cancer. World cancer report 2014. Stewart BW, Wild CP, editors. International Agency for Research on Cancer. Lyon, France: International Agency for Research on Cancer; 2014.

4. International Agency for Research on Cancer. Globocan 2020: All Cancers. [Internet]. 2020 [cited 2021 Aug 16]. Available from: https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf

5. International Agency for Research on Cancer - World Health Organization (IARC-WHO). GLOBOCAN 2020 [Internet]. 2020 [cited 2021 Oct 4]. Available from: https://gco.iarc.fr/today/home

6. Red Española de Registros de Cáncer. Estimaciones de la incidencia del cáncer en España, 2020 [Internet]. 2020 [cited 2021 Aug 16]. Available from: https://redecan.org/storage/documents/a1d352cb-5f15-49e9-99ef-07725d639aee.pdf

7. Instituto Nacional de Estadística. Fallecidos por cáncer en España [Internet]. 2021 [cited 2021 Aug 16]. Available from: https://www.ine.es/infografias/infografia_fallecidos_cancer.pdf

8. Ritchie H, Roser M. Causes of Death [Internet]. 2018 [cited 2021 Oct 4]. Available from: https://ourworldindata.org/causes-of-death

9. International Agency for Research on Cancer. Globocan 2020: Estimated age-standardized incidence rates (World) in 2020, worldwide, males, all ages [Internet]. 2021 [cited 2021 Oct 4]. Available from: https://gco.iarc.fr/today/home

10. International Agency for Research on Cancer. Globocan 2020: Estimated age-standardized incidence rates (World) in 2020, Spain, males, all ages [Internet]. 2021 [cited 2021 Oct 4]. Available from: https://gco.iarc.fr/today/home

11. International Agency for Research on Cancer. Globocan 2020: Estimated age-standardized incidence rates (World) in 2020, worldwide, females, all ages [Internet]. 2021 [cited 2021 Oct 4]. Available from: https://gco.iarc.fr/today/home

12. International Agency for Research on Cancer. Globocan 2020: Estimated age-standardized incidence rates (World) in 2020, Spain, females, all ages [Internet]. 2021 [cited 2021 Oct 4]. Available from: https://gco.iarc.fr/today/home

13. World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD) [Internet]. 2021 [cited 2021 Aug 16]. Available from: https://www.who.int/standards/classifications/classification-of-diseases

14. World Health Organization. International Classification of Diseases for Oncology. Third. Geneva: WHO Press; 2013.

15. SNOMED International. Home [Internet]. 2021 [cited 2021 Oct 11]. Available from: https://www.snomed.org/

16. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. Cancer [Internet]. 2005 Apr 1;103(7):1457–67. Available from: https://doi.org/10.1002/cncr.20910

17. United Kingdom Government. UK Read Code [Internet]. 2015 [cited 2021 Oct 11]. Available from: https://data.gov.uk/dataset/f262aa32-9c4e-44f1-99eb-4900deada7a4/uk-read-code

18. Medical Dictionary for Regulatory Activities. Home [Internet]. 2021 [cited 2021 Oct 11]. Available from: https://www.meddra.org/

19. World Health Organization. International Statistical Classification of Diseases and Related Health Problems: Tenth Revision. Geneva: World Health Organization; 2004.

20. Griffiths A, Miller J, Suzuki D, Lewontin R, Gelbart W. Mutation and cancer. An Introduction to Genetic Analysis. 7th ed. New York: W. H. Freeman; 2000.

21. American Cancer Society. Family cancer syndromes [Internet]. 2021 [cited 2021 Aug 16]. Available from: https://www.cancer.org/cancer/cancer-causes/genetics/family-cancer-syndromes.html

22. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. The Third Expert Report [Internet]. 2018 [cited 2021 Oct 10]. Available from: dietandcancerreport.org

23. National Cancer Institute. Risk factors for cancer [Internet]. 2015 [cited 2021 Oct 4]. Available from: https://www.cancer.gov/about-cancer/causes-prevention/risk

24. Brennan P, Davey-Smith G. Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed. JNCI: Journal of the National Cancer Institute [Internet]. 2022 Mar 1;114(3):353–60. Available from: https://doi.org/10.1093/jnci/djab204

25. World Health Organization. Overweight and obesity [Internet]. 2021 [cited 2021 Aug 17]. Available from: http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

26. Haslam DW, James WPT. Obesity. The Lancet [Internet]. 2005 Oct 1;366(9492):1197–209. Available from: https://doi.org/10.1016/S0140-6736(05)67483-1

27. The GBD 2015 Obesity Collaborators. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. New England Journal of Medicine [Internet]. 2017 Jun 12;377(1):13–27. Available from: https://doi.org/10.1056/NEJMoa1614362

28. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. International Journal of Obesity [Internet]. 2008;32(9):1431–7. Available from: https://doi.org/10.1038/ijo.2008.102

29. Instituto Nacional de Estadística. Sociedad / Salud / Encuesta Nacional de Salud / Determinantes de la salud. Cifras relativas / Características físicas [Internet]. 2017 [cited 2021 Aug 17]. Available from: https://www.ine.es/dynt3/inebase/es/index.htm?type=pcaxis&path=/t15/p419/a2017/p06/&file=pcaxis

30. Aranceta-Bartrina J, Pérez-Rodrigo C, Alberdi-Aresti G, Ramos-Carrera N, Lázaro-Masedo S. Prevalence of General Obesity and Abdominal Obesity in the Spanish Adult Population (Aged 25–64 Years) 2014–2015: The ENPE Study. Revista Española de Cardiología (English Edition) [Internet]. 2016;69(6):579–87. Available from: https://www.sciencedirect.com/science/article/pii/S1885585716001225

31. Gutiérrez-Fisac JL, Guallar-Castillón P, León-Muñoz LM, Graciani A, Banegas JR, Rodríguez-Artalejo F. Prevalence of general and abdominal obesity in the adult population of Spain, 2008–2010: the ENRICA study. Obesity Reviews [Internet]. 2012 Apr 1;13(4):388–92. Available from: https://doi.org/10.1111/j.1467-789X.2011.00964.x

32. World Health Organization. Noncommunicable diseases: Risk factors [Internet]. 2022 [cited 2021 Aug 17]. Available from: http://www.who.int/gho/ncd/risk_factors/overweight_obesity/bmi_trends_adults/en/

33. Valdés S, García-Torres F, Maldonado-Araque C, Goday A, Calle-Pascual A, Soriguer F, et al. Prevalencia de obesidad, diabetes mellitus y otros factores de riesgo cardiovascular en Andalucía. Comparación con datos de prevalencia nacionales. Estudio Di@bet.es. Revista Española de Cardiología [Internet]. 2014 Jun 1 [cited 2022 Mar 9];67(6):442–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0300893213005368

34. Fernández-Bergés D, Cabrera de León A, Sanz H, Elosua R, Guembe MJ, Alzamora M, et al. Síndrome metabólico en España: prevalencia y riesgo coronario asociado a la definición armonizada y a la propuesta por la OMS. Estudio DARIOS. Revista Española de Cardiología [Internet]. 2012 Mar 1 [cited 2022 Mar 9];65(3):241–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0300893211008840

35. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutrition Reviews [Internet]. 2012 Jan 1;70(1):3–21. Available from: https://doi.org/10.1111/j.1753-4887.2011.00456.x

36. Harvard T.H. Chan S of PH. Adult Obesity [Internet]. 2021 [cited 2021 Aug 17]. Available from: https://www.hsph.harvard.edu/obesity-prevention-source/obesity-trends/obesity-rates-worldwide/#References

37. Hu F. Obesity Epidemiology. Oxford University Press; 2008.

38. Du S, Lu B, Zhai F, Popkin BM. A new stage of the nutrition transition in China. Public Health Nutrition [Internet]. 2006/12/22. 2002;5(1a):169–74. Available from: https://www.cambridge.org/core/article/new-stage-of-the-nutrition-transition-in-china/6228807AB394233022BFDF08F36DB386

39. Monteiro CA, Conde WL, Popkin BM. Is obesity replacing or adding to undernutrition? Evidence from different social classes in Brazil. Public Health Nutr [Internet]. 2002;5(1A):105–12. Available from: http://europepmc.org/abstract/MED/12027272

40. Wang Y, Monteiro C, Popkin BM. Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China, and Russia. The American Journal of Clinical Nutrition [Internet]. 2002 Jun 1;75(6):971–7. Available from: https://doi.org/10.1093/ajcn/75.6.971

41. Kanter R, Caballero B. Global Gender Disparities in Obesity: A Review. Advances in Nutrition [Internet]. 2012 Jul 1;3(4):491–8. Available from: https://doi.org/10.3945/an.112.002063

42. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010. JAMA [Internet]. 2012 Feb 1;307(5):491–7. Available from: https://doi.org/10.1001/jama.2012.39

43. Duren DL, Sherwood RJ, Czerwinski SA, Lee M, Choh AC, Siervogel RM, et al. Body Composition Methods: Comparisons and Interpretation. Journal of Diabetes Science and Technology [Internet]. 2008 Nov 1;2(6):1139–46. Available from: https://doi.org/10.1177/193229680800200623

44. Cornier MA, Després JP, Davis N, Grossniklaus DA, Klein S, Lamarche B, et al. Assessing Adiposity. Circulation [Internet]. 2011 Nov 1;124(18):1996–2019. Available from: https://doi.org/10.1161/CIR.0b013e318233bc6a

45. National Institutes of Health, National Heart L and BI, North American association for the study of obesity. The Practical Guide Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. 2000.

46. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. Nutrition Today [Internet]. 2015;50(3). Available from:

https://journals.lww.com/nutritiontodayonline/Fulltext/2015/05000/Body_Mass_Index__Obesity,_BMI,_and_Health__A.5.aspx

47. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity*. International Journal of Epidemiology [Internet]. 2014 Jun 1;43(3):655–65. Available from: https://doi.org/10.1093/ije/dyu058

48. World Health Organization. Physical Status: The Use and Interpretation of Anthropometry: Report of a World Health Organization (WHO) Expert Committee. Geneva; 1995.

49. Gallagher D, Visser M, Sepúlveda D, Pierson RN, Harris T, Heymsfield SB. How Useful Is Body Mass Index for Comparison of Body Fatness across Age, Sex, and Ethnic Groups? American Journal of Epidemiology [Internet]. 1996 Feb 1;143(3):228–39. Available from: https://doi.org/10.1093/oxfordjournals.aje.a008733

50. Blew RM, Sardinha LB, Milliken LA, Teixeira PJ, Going SB, Ferreira DL, et al. Assessing the Validity of Body Mass Index Standards in Early Postmenopausal Women. Obesity Research [Internet]. 2002 Aug 1;10(8):799–808. Available from: https://doi.org/10.1038/oby.2002.108

51. Evans EM, Rowe DA, Racette SB, Ross KM, McAuley E. Is the current BMI obesity classification appropriate for black and white postmenopausal women? International Journal of Obesity [Internet]. 2006;30(5):837–43. Available from: https://doi.org/10.1038/sj.ijo.0803208

52. Garn SM, Leonard WR, Hawthorne VM. Three limitations of the body mass index. The American Journal of Clinical Nutrition [Internet]. 1986 Dec 1;44(6):996–7. Available from: https://doi.org/10.1093/ajcn/44.6.996

53. Malina R, Heymsfield S, Lohman T, Wang Z, Going S. Variation in body composition associated with sex and ethnicity. In: Human Body Composition. 2nd ed. Champaign, IL: Human Kinetics; 2005. p. 271–98.

54. Deurenberg P, Bray G, Bouchard C, James P. Ethnic and geographic influences on body composition. In: Handbook of Obesity. New York: Dekker; 1998. p. 81–92.

55. Lean MEJ, Han TS, Morrison CE. Waist circumference as a measure for indicating need for weight management. BMJ [Internet]. 1995 Jul 15;311(6998):158. Available from: http://www.bmj.com/content/311/6998/158.abstract

56. Pouliot MC, Després JP, Lemieux S, Moorjani S, Bouchard C, Tremblay A, et al. Waist circumference and abdominal sagittal diameter: Best simple anthropometric indexes of abdominal visceral adipose tissue accumulation and related cardiovascular risk in men and women. The American Journal of Cardiology [Internet]. 1994;73(7):460–8. Available from: https://www.sciencedirect.com/science/article/pii/0002914994906769

57. de Koning L, Merchant AT, Pogue J, Anand SS. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. European Heart Journal [Internet]. 2007 Apr 1;28(7):850–6. Available from: https://doi.org/10.1093/eurheartj/ehm026

58. Heymsfield S, Shen W, Wang J, Bray G, Bouchard C, James P. Evaluation of total and regional adiposity. In: Handbook of Obesity. New York: Dekker; 1998.

59. Misra A, Wasir JS, Vikram NK. Waist circumference criteria for the diagnosis of abdominal obesity are not applicable uniformly to all populations and ethnic groups. Nutrition [Internet]. 2005;21(9):969–76. Available from: https://www.sciencedirect.com/science/article/pii/S0899900705001620

60. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. The Lancet [Internet]. 2011;378(9793):804–14. Available from: https://www.sciencedirect.com/science/article/pii/S0140673611608131

61. Kumanyika S, Jeffery RW, Morabia A, Ritenbaugh C, Antipatis VJ. Obesity prevention: the case for action. International Journal of Obesity [Internet]. 2002;26(3):425–36. Available from: https://doi.org/10.1038/sj.ijo.0801938

62. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, et al. Foresight. tackling obesities: future choices - project report [Internet]. 2007 [cited 2021 Mar 9]. Available from: http://www.bis.gov.uk/assets/foresight/docs/obesity/17.pdf.

63. Swinburn B, Egger G, Raza F. Dissecting Obesogenic Environments: The Development and Application of a Framework for Identifying and Prioritizing Environmental Interventions for Obesity. Preventive Medicine [Internet]. 1999;29(6):563–70. Available from: https://www.sciencedirect.com/science/article/pii/S0091743599905856

64. Kitchen PJ, Brignell J, Li T, Jones GS. The Emergence of IMC: A Theoretical Perspective. Journal of Advertising Research [Internet]. 2004 Mar [cited 2022 Mar 9];44(1):19–30. Available from: http://www.journalofadvertisingresearch.com/cgi/doi/10.1017/S0021849904040048

65. Rush EC, Freitas I, Plank LD. Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. British Journal of Nutrition [Internet]. 2009 Aug 10 [cited 2022 Mar 9];102(04):632. Available from: http://www.journals.cambridge.org/abstract_S0007114508207221

66. Fox KR, Hillsdon M. Physical activity and obesity. Obesity Reviews [Internet]. 2007 Mar 1;8(s1):115–21. Available from: https://doi.org/10.1111/j.1467-789X.2007.00329.x

67. King AC, Jeffery RW, Fridinger F, Dusenbury L, Provence S, Hedlund SA, et al. Environmental and Policy Approaches to Cardiovascular Disease Prevention Through Physical Activity: Issues and Opportunities. Health Education Quarterly [Internet]. 1995 Nov 1;22(4):499–511. Available from: https://doi.org/10.1177/109019819502200407

68. Swinburn B, Egger G. Preventive strategies against weight gain and obesity. Obesity Reviews [Internet]. 2002 Nov 1;3(4):289–301. Available from: https://doi.org/10.1046/j.1467-789X.2002.00082.x

69. Singh G, Siahpush M. Ethnic-Immigrant Differentials in Health Behaviors, Morbidity, and Cause-Specific Mortality in the United States: An Analysis of Two National Data Bases. Human Biology. 2002;74(1):83–109.

70. Tremblay MS, Pérez CE, Ardern CI, Bryan SN, Katzmarzyk PT. Obesity, overweight and ethnicity. Health Reports. 2005;16(4):23–34.

71. Lindström M, Sundquist K. The impact of country of birth and time in Sweden on overweight and obesity: A population-based study. Scandinavian Journal of Public Health [Internet]. 2005 Aug 1;33(4):276–84. Available from: https://doi.org/10.1080/14034940510005653

72. Centers for disease Control and Prevention. Genes and obesity [Internet]. 2013 [cited 2021 Aug 21]. Available from: https://www.cdc.gov/genomics/resources/diseases/obesity/obesedit.htm

73. Oken E, Levitan EB, Gillman MW. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. Int J Obes (Lond) [Internet]. 2007/11/27. 2008 Feb;32(2):201–10. Available from: https://pubmed.ncbi.nlm.nih.gov/18278059

74. Oken E, Taveras EM, Kleinman KP, Rich-Edwards JW, Gillman MW. Gestational weight gain and child adiposity at age 3 years. American Journal of Obstetrics and Gynecology [Internet]. 2007;196(4):322.e1-322.e8. Available from: https://www.sciencedirect.com/science/article/pii/S0002937806023994

75. Gillman MW, Rifas-Shiman S, Berkey CS, Field AE, Colditz GA. Maternal Gestational Diabetes, Birth Weight, and Adolescent Obesity. Pediatrics [Internet]. 2003 Mar 1;111(3):e221–6. Available from: https://doi.org/10.1542/peds.111.3.e221

76. Quigley MA. RE: "Duration of breastfeeding and risk of overweight: a meta-analysis." American Journal of Epidemiology [Internet]. 2006 May 1;163(9):870–2. Available from: https://doi.org/10.1093/aje/kwj134

77. Baird J, Fisher D, Lucas P, Kleijnen J, Roberts H, Law C. Being big or growing fast: systematic review of size and growth in infancy and later obesity. BMJ [Internet]. 2005 Oct 20;331(7522):929. Available from: http://www.bmj.com/content/331/7522/929.abstract

78. Wardle J. Eating Behaviour and Obesity. Short Science Review .Foresight Tackling Obesities: Future Choices. [Internet]. [cited 2021 Mar 9]. Available from: http://www.foresight.gov.uk

79. Taveras EM, Rifas-Shiman SL, Oken E, Gunderson EP, Gillman MW. Short Sleep Duration in Infancy and Risk of Childhood Overweight. Archives of Pediatrics & Adolescent Medicine [Internet]. 2008 Apr 1;162(4):305–11. Available from: https://doi.org/10.1001/archpedi.162.4.305

80. Patel SR, Hu FB. Short Sleep Duration and Weight Gain: A Systematic Review. Obesity [Internet]. 2008 Mar 1;16(3):643–53. Available from: https://doi.org/10.1038/oby.2007.118

81. Shaw P, Kushida C. Thermoregulatory changes. In: Sleep Deprivation: Basic Science, Physiology, and Behavior. New York: Dekker; 2005. p. 319–38.

82. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep. 1997;20(4):267–77.

83. Patel SR, Malhotra A, Gottlieb DJ, White DP, Hu FB. Correlates of Long Sleep Duration. Sleep [Internet]. 2006 Jul 1;29(7):881–9. Available from: https://doi.org/10.1093/sleep/29.7.881

84. Patel SR, Malhotra A, White DP, Gottlieb DJ, Hu FB. Association between Reduced Sleep and Weight Gain in Women. American Journal of Epidemiology [Internet]. 2006 Nov 15;164(10):947–54. Available from: https://doi.org/10.1093/aje/kwj280

85. Mullington JM, Chan JL, van Dongen HPA, Szuba MP, Samaras J, Price NJ, et al. Sleep Loss Reduces Diurnal Rhythm Amplitude of Leptin in Healthy Men. Journal of Neuroendocrinology [Internet]. 2003 Sep 1;15(9):851–4. Available from: https://doi.org/10.1046/j.1365-2826.2003.01069.x

86. Maio G, Manstead A, Verplanken B. Lifestyle change. Evidence Review. Foresight Tackling Obesities: Future Choices [Internet]. 2007 [cited 2021 Mar 9]. Available from: http://www.foresight.gov.uk

87. Korkeila M, Kaprio J, Rissanen A, Koskenvuo M, Sörensen TIA. Predictors of major weight gain in adult Finns: Stress, life satisfaction and personality traits. International Journal of Obesity [Internet]. 1998;22(10):949–57. Available from: https://doi.org/10.1038/sj.ijo.0800694

88. Räikkönen K, Matthews KA, Kuller LH. Anthropometric and psychosocial determinants of visceral obesity in healthy postmenopausal women. International Journal of Obesity [Internet]. 1999;23(8):775–82. Available from: https://doi.org/10.1038/sj.ijo.0800917

89. Faith MS, Matz PE, Jorge MA. Obesity–depression associations in the population. Journal of Psychosomatic Research [Internet]. 2002;53(4):935–42. Available from: https://www.sciencedirect.com/science/article/pii/S0022399902003082

90. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. New England Journal of Medicine [Internet]. 2016 Aug 24;375(8):794–8. Available from: https://doi.org/10.1056/NEJMsr1606602

91. International Agency for Research on Cancer. Globocan 2020: Estimated age-standardized incidence rates (World) in 2020, Spain, both sexes, all ages [Internet]. 2021 [cited 2021 Aug 25]. Available from: https://gco.iarc.fr/today/home

92. Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, et al. Rare cancers are not so rare: The rare cancer burden in Europe. European Journal of Cancer [Internet]. 2011;47(17):2493–511. Available from: https://www.sciencedirect.com/science/article/pii/S0959804911006083

93. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. The Lancet [Internet]. 2008;371(9612):569–78. Available from: https://www.sciencedirect.com/science/article/pii/S014067360860269X

94. Engeland A, Tretli S, Bjørge T. Height and Body Mass Index in Relation to Esophageal Cancer; 23-year Follow-up of Two Million Norwegian Men and Women. Cancer Causes & Control [Internet]. 2004;15(8):837–43. Available from: https://doi.org/10.1023/B:CACO.0000043434.21558.ea

95. Engeland A, Tretli S, Bjørge T. Height, Body Mass Index, and Ovarian Cancer: A Follow-Up of 1.1 Million Norwegian Women. JNCI: Journal of the National Cancer Institute [Internet]. 2003 Aug 20;95(16):1244–8. Available from: https://doi.org/10.1093/jnci/djg010

96. Engeland A, Tretli S, Bjørge T. Height, body mass index, and prostate cancer: a follow-up of 950 000 Norwegian men. British Journal of Cancer [Internet]. 2003;89(7):1237–42. Available from: https://doi.org/10.1038/sj.bjc.6601206

97. Bjørge T, Tretli S, Engeland A. Relation of Height and Body Mass Index to Renal Cell Carcinoma in Two Million Norwegian Men and Women. American Journal of Epidemiology [Internet]. 2004 Dec 15;160(12):1168–76. Available from: https://doi.org/10.1093/aje/kwh345

98. Engeland A, Tretli S, Austad G, Bjørge T. Height and Body Mass Index in Relation to Colorectal and Gallbladder Cancer in Two Million Norwegian Men and Women. Cancer Causes & Control [Internet]. 2005;16(8):987–96. Available from: https://doi.org/10.1007/s10552-005-3638-3

99. Bjørge T, Tretli S, Engeland A. Height and body mass index in relation to cancer of the small intestine in two million Norwegian men and women. British Journal of Cancer [Internet]. 2005;93(7):807–10. Available from: https://doi.org/10.1038/sj.bjc.6602789

100. Engeland A, Tretli S, Akslen LA, Bjørge T. Body size and thyroid cancer in two million Norwegian men and women. British Journal of Cancer [Internet]. 2006;95(3):366–70. Available from: https://doi.org/10.1038/sj.bjc.6603249

101. Engeland A, Tretli S, Hansen S, Bjørge T. Height and Body Mass Index and Risk of Lymphohematopoietic Malignancies in Two Million Norwegian Men and Women. American Journal of Epidemiology [Internet]. 2007 Jan 1;165(1):44–52. Available from: https://doi.org/10.1093/aje/kwj353

102. Bjørge T, Tretli S, Lie AK, Engeland A. The impact of height and body mass index on the risk of testicular cancer in 600,000 Norwegian men. Cancer Causes & Control [Internet]. 2006;17(7):983–7. Available from: https://doi.org/10.1007/s10552-006-0032-8

103. Bjørge T, Engeland A, Tretli S, Weiderpass E. Body size in relation to cancer of the uterine corpus in 1 million Norwegian women. International Journal of Cancer [Internet]. 2007 Jan 15;120(2):378–83. Available from: https://doi.org/10.1002/ijc.22260

104. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ [Internet]. 2007 Nov 29;335(7630):1134. Available from: http://www.bmj.com/content/335/7630/1134.abstract

105. Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5.24 million UK adults. The Lancet. 2014;384(9945):755–65.

106. Song M, Giovannucci E. Estimating the Influence of Obesity on Cancer Risk: Stratification by Smoking Is Critical. Journal of Clinical Oncology [Internet]. 2016 Jul 25;34(27):3237–9. Available from: https://doi.org/10.1200/JCO.2016.67.6916

107. Hofstetter A, Schutz Y, Jéquier E, Wahren J. Increased 24-Hour Energy Expenditure in Cigarette Smokers. New England Journal of Medicine [Internet]. 1986 Jan 9;314(2):79–82. Available from: https://doi.org/10.1056/NEJM198601093140204

108. Jessen A, Buemann B, Toubro S, Skovgaard IM, Astrup A. The appetite-suppressant effect of nicotine is enhanced by caffeine*. Diabetes, Obesity and Metabolism [Internet]. 2005 Jul 1;7(4):327–33. Available from: https://doi.org/10.1111/j.1463-1326.2004.00389.x

109. van den Borst B, Koster A, Yu B, Gosker HR, Meibohm B, Bauer DC, et al. Is age-related decline in lean mass and physical function accelerated by obstructive lung

disease or smoking? Thorax [Internet]. 2011 Nov 1;66(11):961. Available from: http://thorax.bmj.com/content/66/11/961.abstract

110. Kim JH, Shim KW, Yoon YS, Lee SY, Kim SS, Oh SW. Cigarette Smoking Increases Abdominal and Visceral Obesity but Not Overall Fatness: An Observational Study. PLOS ONE [Internet]. 2012 Sep 24;7(9):e45815-. Available from: https://doi.org/10.1371/journal.pone.0045815

111. Lv J, Chen W, Sun D, Li S, Millwood IY, Smith M, et al. Gender-Specific Association between Tobacco Smoking and Central Obesity among 0.5 Million Chinese People: The China Kadoorie Biobank Study. PLOS ONE [Internet]. 2015 Apr 21;10(4):e0124586-. Available from: https://doi.org/10.1371/journal.pone.0124586

112. Canoy D, Wareham N, Luben R, Welch A, Bingham S, Day N, et al. Cigarette Smoking and Fat Distribution in 21, 828 British Men and Women: A Population-based Study. Obesity Research [Internet]. 2005 Aug 1;13(8):1466–75. Available from: https://doi.org/10.1038/oby.2005.177

113. Cena H, Fonte ML, Turconi G. Relationship between smoking and metabolic syndrome. Nutrition Reviews [Internet]. 2011 Dec 1;69(12):745–53. Available from: https://doi.org/10.1111/j.1753-4887.2011.00446.x

114. Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. Nature Reviews Cancer [Internet]. 2015;15(8):484–98. Available from: https://doi.org/10.1038/nrc3967

115. Balagopal P (Babu), de Ferranti SD, Cook S, Daniels SR, Gidding SS, Hayman LL, et al. Nontraditional Risk Factors and Biomarkers for Cardiovascular Disease: Mechanistic, Research, and Clinical Considerations for Youth. Circulation [Internet]. 2011 Jun 14;123(23):2749–69. Available from: https://doi.org/10.1161/CIR.0b013e31821c7c64

116. de Ridder J, Julián-Almárcegui C, Mullee A, Rinaldi S, van Herck K, Vicente-Rodríguez G, et al. Comparison of anthropometric measurements of adiposity in relation to cancer risk: a systematic review of prospective studies. Cancer Causes & Control [Internet]. 2016;27(3):291–300. Available from: https://doi.org/10.1007/s10552-015-0709-y

117. Aleksandrova K, Nimptsch K, Pischon T. Obesity and colorectal cancer. Frontiers in Bioscience Elite. 2013;5(1):61–77.

118. Perez-Cornago A, Appleby PN, Pischon T, Tsilidis KK, Tjønneland A, Olsen A, et al. Tall height and obesity are associated with an increased risk of aggressive prostate cancer: results from the EPIC cohort study. BMC Medicine [Internet]. 2017;15(1):115. Available from: https://doi.org/10.1186/s12916-017-0876-7

119. Moore LL, Bradlee ML, Singer MR, Splansky GL, Proctor MH, Ellison RC, et al. BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study

adults. International Journal of Obesity [Internet]. 2004;28(4):559–67. Available from: https://doi.org/10.1038/sj.ijo.0802606

120. White AJ, Nichols HB, Bradshaw PT, Sandler DP. Overall and central adiposity and breast cancer risk in the sister study. Cancer [Internet]. 2015 Oct 15;121(20):3700–8. Available from: https://doi.org/10.1002/cncr.29552

121. World Cancer Research Fund and American Institute for Cancer Research. Continuous Update Project Report. Food, Nutrition, Physical Activity, and the Prevention of Pancreatic Cancer. 2012.

122. World Cancer Research Fund and American Institute for Cancer Research. Continuous Update Project Report. Food, Nutrition, Physical Activity, and the Prevention of Endometrial Cancer. 2013.

123. World Cancer Research Fund and American Institute for Cancer Research. Continuous UpdateProject Report. Food, Nutrition, Physical Activity, and the Prevention of Ovarian Cancer. 2014.

124. World Cancer Research Fund and American Institute for Cancer Research. Continuous Update Project Report: Diet, Nutrition, Physical Activity, and Prostate Cancer . 2014.

125. Freisling H, Arnold M, Soerjomataram I, O'Doherty MG, Ordóñez-Mena JM, Bamia C, et al. Comparison of general obesity and measures of body fat distribution in older adults in relation to cancer risk: meta-analysis of individual participant data of seven prospective cohorts in Europe. British Journal of Cancer [Internet]. 2017;116(11):1486–97. Available from: https://doi.org/10.1038/bjc.2017.106

126. Barberio AM, Alareeki A, Viner B, Pader J, Vena JE, Arora P, et al. Central body fatness is a stronger predictor of cancer risk than overall body size. Nature Communications [Internet]. 2019;10(1):383. Available from: https://doi.org/10.1038/s41467-018-08159-w

127. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Future research directions. [Internet]. 2018 [cited 2022 Mar 9]. Available from: dietandcancerreport.org

128. Stolzenberg-Solomon RZ, Schairer C, Moore S, Hollenbeck A, Silverman DT. Lifetime adiposity and risk of pancreatic cancer in the NIH-AARP Diet and Health Study cohort. The American Journal of Clinical Nutrition [Internet]. 2013 Oct 1;98(4):1057–65. Available from: https://doi.org/10.3945/ajcn.113.058123

129. Arnold M, Freisling H, Stolzenberg-Solomon R, Kee F, O'Doherty MG, Ordóñez-Mena JM, et al. Overweight duration in older adults and cancer risk: a study of cohorts in Europe and the United States. European Journal of Epidemiology [Internet]. 2016;31(9):893–904. Available from: https://doi.org/10.1007/s10654-016-0169-z

130. Noh H, Charvat H, Freisling H, Ólafsdóttir GH, Ólafsdóttir EJ, Tryggvadóttir L, et al. Cumulative exposure to premenopausal obesity and risk of postmenopausal cancer: A population-based study in Icelandic women. International Journal of Cancer [Internet]. 2020 Aug 1;147(3):793–802. Available from: https://doi.org/10.1002/ijc.32805

131. Arnold M, Jiang L, Stefanick ML, Johnson KC, Lane DS, LeBlanc ES, et al. Duration of Adulthood Overweight, Obesity, and Cancer Risk in the Women's Health Initiative: A Longitudinal Study from the United States. PLOS Medicine [Internet]. 2016 Aug 16;13(8):e1002081-. Available from: https://doi.org/10.1371/journal.pmed.1002081

132. Marinac CR, Birmann BM, Lee IM, Rosner BA, Townsend MK, Giovannucci E, et al. Body mass index throughout adulthood, physical activity, and risk of multiple myeloma: a prospective analysis in three large cohorts. British Journal of Cancer [Internet]. 2018;118(7):1013–9. Available from: https://doi.org/10.1038/s41416-018-0010-4

133. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nature Reviews Cancer [Internet]. 2004;4(8):579–91. Available from: https://doi.org/10.1038/nrc1408

134. Khandekar MJ, Cohen P, Spiegelman BM. Molecular mechanisms of cancer development in obesity. Nature Reviews Cancer [Internet]. 2011;11(12):886–95. Available from: https://doi.org/10.1038/nrc3174

135. Roberts DL, Dive C, Renehan AG. Biological Mechanisms Linking Obesity and Cancer Risk: New Perspectives. Annual Review of Medicine [Internet]. 2010 Jan 8;61(1):301–16. Available from: https://doi.org/10.1146/annurev.med.080708.082713

136. van Kruijsdijk RCM, van der Wall E, Visseren FLJ. Obesity and Cancer: The Role of Dysfunctional Adipose Tissue. Cancer Epidemiology, Biomarkers & Prevention [Internet]. 2009 Oct 8;18(10):2569–78. Available from: https://doi.org/10.1158/1055-9965.EPI-09-0372

137. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: A meta-analysis of prospective studies. European Journal of Cancer [Internet]. 2011;47(16):2422–30. Available from: https://www.sciencedirect.com/science/article/pii/S0959804911004333

138. Wallin A, Larsson SC. Body mass index and risk of multiple myeloma: A meta-analysis of prospective studies. European Journal of Cancer [Internet]. 2011;47(11):1606–15. Available from: https://www.sciencedirect.com/science/article/pii/S0959804911000657

139. Poynter JN, Richardson M, Blair CK, Roesler MA, Hirsch BA, Nguyen P, et al. Obesity over the life course and risk of acute myeloid leukemia and myelodysplastic syndromes. Cancer Epidemiology [Internet]. 2016;40:134–40. Available from: https://www.sciencedirect.com/science/article/pii/S1877782115002830

140. Onat A, Avcı GŞ, Barlan MM, Uyarel H, Uzunlar B, Sansoy V. Measures of abdominal obesity assessed for visceral adiposity and relation to coronary risk. International Journal of Obesity [Internet]. 2004;28(8):1018–25. Available from: https://doi.org/10.1038/sj.ijo.0802695

141. Kahn BB, Flier JS. Obesity and insulin resistance. The Journal of Clinical Investigation [Internet]. 2000 Aug 15;106(4):473–81. Available from: https://doi.org/10.1172/JCI10842

142. Clayton PE, Banerjee I, Murray PG, Renehan AG. Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. Nature Reviews Endocrinology [Internet]. 2011;7(1):11–24. Available from: https://doi.org/10.1038/nrendo.2010.171

143. Abdullah A, Amin FA, Stoelwinder J, Tanamas SK, Wolfe R, Barendregt J, et al. Estimating the risk of cardiovascular disease using an obese-years metric. BMJ Open [Internet]. 2014 Sep 1;4(9):e005629. Available from: http://bmjopen.bmj.com/content/4/9/e005629.abstract

144. Abdullah A, Wolfe R, Mannan H, Stoelwinder JU, Stevenson C, Peeters A. Epidemiologic Merit of Obese-Years, the Combination of Degree and Duration of Obesity. American Journal of Epidemiology [Internet]. 2012 Jul 15;176(2):99–107. Available from: https://doi.org/10.1093/aje/kwr522

145. Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and Obesity as Determinants of Cardiovascular Risk: The Framingham Experience. Archives of Internal Medicine [Internet]. 2002 Sep 9;162(16):1867–72. Available from: https://doi.org/10.1001/archinte.162.16.1867

146. World Health Organization. Hypertension [Internet]. 2018 [cited 2021 Sep 7]. Available from: https://www.who.int/news-room/fact-sheets/detail/hypertension

147. World Health Organization. Prevalence of raised blood pressure, ages 18+, 1975-2015 (Age standardized estimate): both sexes: 2015 [Internet]. 2015 [cited 2021 Sep 30]. Available from: http://gamapserver.who.int/gho/interactive_charts/ncd/risk_factors/blood_pressure_prevalence/atlas.html

148. Menéndez E, Delgado E, Fernández-Vega F, Prieto MA, Bordiú E, Calle A, et al. Prevalencia, diagnóstico, tratamiento y control de la hipertensión arterial en España. Resultados del estudio Di@bet.es. Revista Española de Cardiología [Internet]. 2016;69(6):572–8. Available from: https://www.sciencedirect.com/science/article/pii/S030089321600035X

149. Carretero OA, Oparil S. Essential Hypertension. Circulation [Internet]. 2000 Jan 25;101(3):329–35. Available from: https://doi.org/10.1161/01.CIR.101.3.329

150. The INTERSALT Co-operative Research Group. Sodium, potassium, body mass, alcohol and blood pressure: the INTERSALT Study. . Journal of hypertension Supplement : official journal of the International Society of Hypertension. 1988;6(4):S584–6.

151. Sever PS, Poulter NR. A hypothesis for the pathogenesis of essential hypertension: the initiating factors. Journal of Hypertension [Internet]. 1989;7. Available from: https://journals.lww.com/jhypertension/Fulltext/1989/02001/A_hypothesis_for_the_pathogenesis_of_essential.4.aspx

152. Mayo Clinic. High blood pressure (hypertension) [Internet]. 2021 [cited 2021 Sep 30]. Available from: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410

153. Centers for disease control and prevention. High blood pressure symptoms and causes [Internet]. 2021 [cited 2021 Sep 30]. Available from: https://www.cdc.gov/bloodpressure/about.htm

154. Stocks T, van Hemelrijck M, Manjer J, Bjørge T, Ulmer H, Hallmans G, et al. Blood Pressure and Risk of Cancer Incidence and Mortality in the Metabolic Syndrome and Cancer Project. Hypertension [Internet]. 2012 Apr 1;59(4):802–10. Available from: https://doi.org/10.1161/HYPERTENSIONAHA.111.189258

155. Radišauskas R, Kuzmickienė I, Milinavičienė E, Everatt R. Hypertension, serum lipids and cancer risk: A review of epidemiological evidence. Medicina (B Aires) [Internet]. 2016;52(2):89–98. Available from: https://www.sciencedirect.com/science/article/pii/S1010660X16000276

156. Seretis A, Cividini S, Markozannes G, Tseretopoulou X, Lopez DS, Ntzani EE, et al. Association between blood pressure and risk of cancer development: a systematic review and meta-analysis of observational studies. Scientific Reports [Internet]. 2019;9(1):8565. Available from: https://doi.org/10.1038/s41598-019-45014-4

157. World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus [Internet]. Geneva: World Health Organization; 1999. Available from: https://apps.who.int/iris/handle/10665/66040

158. Institut Català de la Salut. Guies de pràctica clinica: Abordatge de la diabetis mellitus tipus 2 [Internet]. 2015 [cited 2019 Mar 5]. Available from: http://ics.gencat.cat/web/.content/documents/assistencia/gpc/GuiaDiabetis2015.pdf

159. World Health Organization. Global report on diabetes. Geneva; 2016.

160. Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. Diabetes Research and Clinical Practice [Internet]. 2014;103(2):150–60.

Available from: https://www.sciencedirect.com/science/article/pii/S0168822713003847

161. Gakidou E, Mallinger L, Abbott-Klafter J, Guerrero R, Villalpando S, Ridaura RL, et al. Management of diabetes and associated cardiovascular risk factors in seven countries: a comparison of data from national health examination surveys. Bull World Health Organ. 2011;89(3):172–83.

162. World Health Organization. Tracking universal health coverage: first global monitoring report. Geneva; 2015.

163. Zhou B, Lu Y, Hajifathalian K, Bentham J, di Cesare M, Danaei G, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. The Lancet [Internet]. 2016 Apr 9;387(10027):1513–30. Available from: https://doi.org/10.1016/S0140-6736(16)00618-8

164. Bullard K, Cowie C, Lessem S, Saydah S, Menke A, Geiss L, et al. Prevalence of Diagnosed Diabetes in Adults by Diabetes Type — United States,. MMWR Morb Mortal Wkly Rep. 2018;67:359–61.

165. Ruiz-Ramos M, Escolar-Pujolar A, Mayoral-Sánchez E, Corral-San Laureano F, Fernández-Fernández I. La diabetes mellitus en España: mortalidad, prevalencia, incidencia, costes económicos y desigualdades. Gaceta Sanitaria [Internet]. 2006;20:15–24. Available from: https://www.sciencedirect.com/science/article/pii/S021391110671562X

166. Mayo clinic. Type 2 Diabetes [Internet]. 2021 [cited 2021 Oct 4]. Available from: https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/syc-20351193

167. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ : British Medical Journal [Internet]. 2015 Jan 2;350:g7607. Available from: http://www.bmj.com/content/350/bmj.g7607.abstract

168. Starup-Linde J, Karlstad O, Eriksen SA, Vestergaard P, Bronsveld HK, de Vries F, et al. CARING (CAncer Risk and INsulin analoGues): the association of diabetes mellitus and cancer risk with focus on possible determinants - a systematic review and a meta-analysis. Curr Drug Saf [Internet]. 2013 Nov;8(5):296–332. Available from: https://pubmed.ncbi.nlm.nih.gov/24215312

169. Ohkuma T, Peters SAE, Woodward M. Sex differences in the association between diabetes and cancer: a systematic review and meta-analysis of 121 cohorts including 20 million individuals and one million events. Diabetologia [Internet]. 2018;61(10):2140–54. Available from: https://doi.org/10.1007/s00125-018-4664-5

170. Noto H, Osame K, Sasazuki T, Noda M. Substantially increased risk of cancer in patients with diabetes mellitus: A systematic review and meta-analysis of epidemiologic evidence in Japan. Journal of Diabetes and its Complications [Internet]. 2010;24(5):345–53. Available from: https://www.sciencedirect.com/science/article/pii/S1056872710000693

171. Zhenming G, Qiwen B, Junbo Q, Yamin W, Yuming L. Diabetes mellitus and risk of gastric cancer: a systematic review and meta-analysis of observational studies. European Journal of Gastroenterology & Hepatology [Internet]. 2011;23(12). Available from: https://journals.lww.com/eurojgh/Fulltext/2011/12000/Diabetes_mellitus_and_risk_of_gastric_cancer__a.6.aspx

172. Jiang Y, Ben Q, Shen H, Lu W, Zhang Y, Zhu J. Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies. European Journal of Epidemiology [Internet]. 2011;26(11):863–76. Available from: https://doi.org/10.1007/s10654-011-9617-y

173. Wang C, Wang X, Gong G, Ben Q, Qiu W, Chen Y, et al. Increased risk of hepatocellular carcinoma in patients with diabetes mellitus: A systematic review and meta-analysis of cohort studies. International Journal of Cancer [Internet]. 2012 Apr 1;130(7):1639–48. Available from: https://doi.org/10.1002/ijc.26165

174. Song S, Wang B, Zhang X, Hao L, Hu X, Li Z, et al. Long-Term Diabetes Mellitus Is Associated with an Increased Risk of Pancreatic Cancer: A Meta-Analysis. PLOS ONE [Internet]. 2015 Jul 29;10(7):e0134321-. Available from: https://doi.org/10.1371/journal.pone.0134321

175. Friberg E, Orsini N, Mantzoros CS, Wolk A. Diabetes mellitus and risk of endometrial cancer: a meta-analysis. Diabetologia [Internet]. 2007;50(7):1365–74. Available from: https://doi.org/10.1007/s00125-007-0681-5

176. Bao C, Yang X, Xu W, Luo H, Xu Z, Su C, et al. Diabetes mellitus and incidence and mortality of kidney cancer: A meta-analysis. Journal of Diabetes and its Complications [Internet]. 2013;27(4):357–64. Available from: https://www.sciencedirect.com/science/article/pii/S1056872713000056

177. Castillo JJ, Mull N, Reagan JL, Nemr S, Mitri J. Increased incidence of non-Hodgkin lymphoma, leukemia, and myeloma in patients with diabetes mellitus type 2: a meta-analysis of observational studies. Blood [Internet]. 2012 May 24;119(21):4845–50. Available from: https://doi.org/10.1182/blood-2011-06-362830

178. Bansal D, Bhansali A, Kapil G, Undela K, Tiwari P. Type 2 diabetes and risk of prostate cancer: a meta-analysis of observational studies. Prostate Cancer and Prostatic Diseases [Internet]. 2013;16(2):151–8. Available from: https://doi.org/10.1038/pcan.2012.40

179. National Health Service. Cardiovascular disease [Internet]. 2018 [cited 2021 Oct 7]. Available from: https://www.nhs.uk/conditions/cardiovascular-disease/

180. World Health Organization. Cardiovascular diseases [Internet]. 2021 [cited 2021 Oct 7]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

181. National Institutes of Health. In Brief: Your Guide To Living Well With Heart Disease. [Internet]. 2006 [cited 2021 Oct 7]. Available from: https://www.nhlbi.nih.gov/files/docs/public/heart/living_hd_fs.pdf

182. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol [Internet]. 2017;70(1):1–25. Available from: https://www.sciencedirect.com/science/article/pii/S0735109717372443

183. Banegas JR, Villar F, Graciani A, Rodríguez-Artalejo F. Epidemiología de las enfermedades cardiovasculares en España. Revista Española de Cardiología Suplementos [Internet]. 2006;6(7):3G-12G. Available from: https://www.sciencedirect.com/science/article/pii/S1131358706753249

184. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation [Internet]. 2016 Mar 15;133(11):1104–14. Available from: https://doi.org/10.1161/CIRCULATIONAHA.115.020406

185. Meijers WC, Maglione M, Bakker SJL, Oberhuber R, Kieneker LM, de Jong S, et al. Heart Failure Stimulates Tumor Growth by Circulating Factors. Circulation [Internet]. 2018 Aug 14;138(7):678–91. Available from: https://doi.org/10.1161/CIRCULATIONAHA.117.030816

186. Alexander RW. Hypertension and the Pathogenesis of Atherosclerosis. Hypertension [Internet]. 1995 Feb 1;25(2):155–61. Available from: https://doi.org/10.1161/01.HYP.25.2.155

187. Felmeden DC, Spencer CGC, Belgore FM, Blann AD, Beevers DG, Lip GYH. Endothelial damage and angiogenesis in hypertensive patients: relationship to cardiovascular risk factors and risk factor management*: American Journal of Hypertension [Internet]. 2003 Jan 1;16(1):11–20. Available from: https://doi.org/10.1016/S0895-7061(02)03149-7

188. Ferrara N. VEGF and the quest for tumour angiogenesis factors. Nature Reviews Cancer [Internet]. 2002;2(10):795–803. Available from: https://doi.org/10.1038/nrc909

189. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. The Lancet [Internet]. 2004 Apr 24;363(9418):1346–53. Available from: https://doi.org/10.1016/S0140-6736(04)16044-3

190. Chen W, Wang S, Tian T, Bai J, Hu Z, Xu Y, et al. Phenotypes and genotypes of insulin-like growth factor 1, IGF-binding protein-3 and cancer risk: evidence from 96 studies.

European Journal of Human Genetics [Internet]. 2009;17(12):1668–75. Available from: https://doi.org/10.1038/ejhg.2009.86

191. Kaaks R, Lukanova A, Kurzer MS. Obesity, Endogenous Hormones, and Endometrial Cancer Risk: A Synthetic Review. Cancer Epidemiology, Biomarkers & Prevention. 2002 Dec 1;11(12):1531–43.

192. Murphy E, Kelly DP. Estrogen Signaling and Cardiovascular Disease. Circulation Research [Internet]. 2011 Sep 2;109(6):687–96. Available from: https://doi.org/10.1161/CIRCRESAHA.110.236687

193. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. The Journal of Clinical Investigation [Internet]. 2006 Jul 3;116(7):1793–801. Available from: https://doi.org/10.1172/JCI29069

194. Masoudkabir F, Sarrafzadegan N, Gotay C, Ignaszewski A, Krahn AD, Davis MK, et al. Cardiovascular disease and cancer: Evidence for shared disease pathways and pharmacologic prevention. Atherosclerosis [Internet]. 2017;263:343–51. Available from: https://www.sciencedirect.com/science/article/pii/S0021915017302484

195. Gunter TD, Terry NP. The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions. J Med Internet Res [Internet]. 2005;7(1):e3. Available from: http://www.jmir.org/2005/1/e3/

196. Cook JA, Collins GS. The rise of big clinical databases. British Journal of Surgery. 2015;102(2):93–101.

197. Haynes K, Forde KA, Schinnar R, Wong P, Strom BL, Lewis JD. Cancer incidence in The Health Improvement Network. Pharmacoepidemiology and Drug Safety [Internet]. 2009 Aug 1;18(8):730–6. Available from: https://doi.org/10.1002/pds.1774

198. Nissen F, Quint JK, Morales DR, Douglas IJ. How to validate a diagnosis recorded in electronic health records. Breathe [Internet]. 2019 Mar 1;15(1):64. Available from: http://breathe.ersjournals.com/content/15/1/64.abstract

199. Bolíbar B, Fina Avilés F, Morros R, del Mar Garcia-Gil M, Hermosilla E, Ramos R, et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. Medicina Clinica. 2012;138(14):617–21.

200. Generalitat de Catalunya. Conjunt mínim bàsic de dades (CMBD) [Internet]. 2017 [cited 2019 Mar 5]. Available from: https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/index.html#googtrans(ca%7Ces)

201. Borràs JM, Izquierdo A, Vilardell L, Marcos-Gragera R, Ribes J, Galvez J, et al. Cancer Incidence in Girona (2008-2012). In: Bray F, Colombet M, Mery L, Piñeros M, Znaor

A, Zanetti R, et al., editors. Cancer Incidence in Five Continents. Vol XI (el. Lyon, France: International Agency for Research on Cancer; 2017.

202. Galceran J, Carulla M, Ameijide A, Jiménez A, Llauradó L, Mateu S, et al. Cancer Incidence in Tarragona (2008-2012). In: Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al., editors. Cancer Incidence in Five Continents. Vol. XI (e. Lyon, France: International Agency for Research on Cancer; 2017.

203. Unitat d'Epidemiologia i Registre de Càncer de Girona. El càncer a Girona 2010-12: Projeccions de la incidència 2017 [Internet]. Girona; 2016 [cited 2019 Jan 14]. Available from: http://ico.gencat.cat/ca/professionals/serveis_i_programes/registre_del_cancer/

204. Navarro C, Martos C, Ardanaz E, Galceran J, Izarzugaza I, Peris-Bonet R, et al. Population-based cancer registries in Spain and their role in cancer control. Annals of Oncology. 2010 May;21(Supplement 3):iii3–13.

205. International Agency for Research on Cancer. Indices of data Quality: All sites except non-melanoma skin (C00-96 exc. C44) [Internet]. Cancer Incidence in Five Continents XI. 2017 [cited 2018 Mar 10]. Available from: http://ci5.iarc.fr/CI5-XI/PDF/INDICES/21.pdf

206. International Agency for Research on Cancer. Chapter 5: Data Comparability and Quality [Internet]. Cancer Incidence in Five Continents Volume XI. 2017 [cited 2018 Mar 10]. Available from: http://ci5.iarc.fr/CI5-XI/Pages/Chapter5.aspx

207. Lecube A, Monereo S, Rubio MÁ, Martínez-de-Icaya P, Martí A, Salvador J, et al. Prevención, diagnóstico y tratamiento de la obesidad. Posicionamiento de la Sociedad Española para el Estudio de la Obesidad de 2016. Endocrinología, Diabetes y Nutrición. 2017;64:15–22.

208. Abdullah A, Wolfe R, Stoelwinder JU, de Courten M, Stevenson C, Walls HL, et al. The number of years lived with obesity and the risk of all-cause and cause-specific mortality. International Journal of Epidemiology [Internet]. 2011 Aug 1;40(4):985–96. Available from: https://doi.org/10.1093/ije/dyr018

209. World Health Organization. Ninth revision of the International Classification of Diseases. 1976.

210. Arnold M, Ferlay J, van Berge Henegouwen MI, Soerjomataram I. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. Gut [Internet]. 2020 Sep 1;69(9):1564. Available from: http://gut.bmj.com/content/69/9/1564.abstract

211. Orsini N, Greenland S. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. Stata Journal. 2011;11(1):1–29.

212. Harrell FEJ. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2001.

213. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLOS Medicine [Internet]. 2007 Oct 16;4(10):e297-. Available from: https://doi.org/10.1371/journal.pmed.0040297

214. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarød K, Holmen J. Obesity, Smoking, and Physical Inactivity as Risk Factors for CKD: Are Men More Vulnerable? American Journal of Kidney Diseases [Internet]. 2006 Mar 1;47(3):396–405. Available from: https://doi.org/10.1053/j.ajkd.2005.11.027

215. Discacciati A, Orsini N, Wolk A. Body mass index and incidence of localized and advanced prostate cancer—a dose–response meta-analysis of prospective studies. Annals of Oncology [Internet]. 2012;23(7):1665–71. Available from: https://www.sciencedirect.com/science/article/pii/S0923753419380238

216. Molina-Montes E, Ubago-Guisado E, Petrova D, Amiano P, Chirlaque MD, Agudo A, et al. The Role of Diet, Alcohol, BMI, and Physical Activity in Cancer Mortality: Summary Findings of the EPIC Study. . Nutrients. 2021;13(12).

217. Ford ES. Body Mass index and Colon Cancer in a National Sample of Adult US Men and Women. American Journal of Epidemiology [Internet]. 1999 Aug 15;150(4):390–8. Available from: https://doi.org/10.1093/oxfordjournals.aje.a010018

218. García-Gil M del M, Hermosilla E, Prieto-Alhambra D, Fina F, Rosell M, Ramos R, et al. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). Journal of Innovation in Health Informatics. 2011;19(3):135–45.

219. Garcia-Gil M, Elorza JM, Banque M, Comas-Cufí M, Blanch J, Ramos R, et al. Linking of Primary Care Records to Census Data to Study the Association between Socioeconomic Status and Cancer Incidence in Southern Europe: A Nation-Wide Ecological Study. PLoS ONE. 2014;9(10):e109706.

220. Margulis A v., Fortuny J, Kaye JA, Calingaert B, Reynolds M, Plana E, et al. Validation of Cancer Cases Using Primary Care, Cancer Registry, and Hospitalization Data in the United Kingdom. Epidemiology. 2018;29(2):308–13.

221. Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. Cancer Epidemiology. 2012;36(5):425–9.

222. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. British Journal of Clinical Pharmacology [Internet]. 2010 Jan 1;69(1):4–14. Available from: https://doi.org/10.1111/j.1365-2125.2009.03537.x

223. Abar L, Sobiecki JG, Cariolou M, Nanu N, Vieira AR, Stevens C, et al. Body size and obesity during adulthood, and risk of lympho-haematopoietic cancers: an update of the WCRF-AICR systematic review of published prospective studies. Annals of Oncology [Internet]. 2019;30(4):528–41. Available from: https://www.sciencedirect.com/science/article/pii/S0923753419311263

224. Larsson SC, Wolk A. Overweight and obesity and incidence of leukemia: A meta-analysis of cohort studies. International Journal of Cancer [Internet]. 2008 Mar 15;122(6):1418–21. Available from: https://doi.org/10.1002/ijc.23176

225. Willett E v, Morton LM, Hartge P, Becker N, Bernstein L, Boffetta P, et al. Non-Hodgkin lymphoma and obesity: A pooled analysis from the InterLymph Consortium. International Journal of Cancer [Internet]. 2008 May 1;122(9):2062–70. Available from: https://doi.org/10.1002/ijc.23344

226. Strongman H, Brown A, Smeeth L, Bhaskaran K. Body mass index and Hodgkin's lymphoma: UK population-based cohort study of 5.8 million individuals. British Journal of Cancer [Internet]. 2019;120(7):768–70. Available from: https://doi.org/10.1038/s41416-019-0401-1

227. Harvard T.H. Chan S of PH. Healthy Food Environment [Internet]. 2022 [cited 2022 Jan 17]. Available from: https://www.hsph.harvard.edu/obesity-prevention-source/obesity-prevention/food-environment/

228. Chan R, Woo J. Prevention of Overweight and Obesity: How Effective is the Current Public Health Approach. Journal of Environmental Research and Public Health. 2010;7(3):765–83.

229. Sacks G, Swinburn B, Lawrence M. Obesity Policy Action framework and analysis grids for a comprehensive policy approach to reducing obesity. Obesity Reviews [Internet]. 2009 Jan 1;10(1):76–86. Available from: https://doi.org/10.1111/j.1467-789X.2008.00524.x

230. Dietz WH, Benken DE, Hunter AS. Public Health Law and the Prevention and Control of Obesity. The Milbank Quarterly [Internet]. 2009 Mar 1;87(1):215–27. Available from: https://doi.org/10.1111/j.1468-0009.2009.00553.x

231. Khan LK, Sobush K, Keener D, Goodman K, Lowry A, Kakietek J, et al. Recommended community strategies and measurements to prevent obesity in the United States. . MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports,. 2009;58(RR7):1–26.

232. Corporate Europe Observatory. A red light for consumer information: the food industry's €1-billion campaign to block health warnings on food. Brussels; 2010.

233. Brownell KD, Warner KE. The Perils of Ignoring History: Big Tobacco Played Dirty and Millions Died. How Similar Is Big Food? The Milbank Quarterly [Internet]. 2009

Mar 1;87(1):259–94. Available from: https://doi.org/10.1111/j.1468-0009.2009.00555.x

234. Shekelle PG, Morton SC, Maglione M, Suttorp M, Tu W, Li Z, et al. Pharmacological and surgical treatment of obesity. Evid Rep Technol Assess (Summ) [Internet]. 2004;(103):1–6. Available from: http://europepmc.org/abstract/MED/15526396

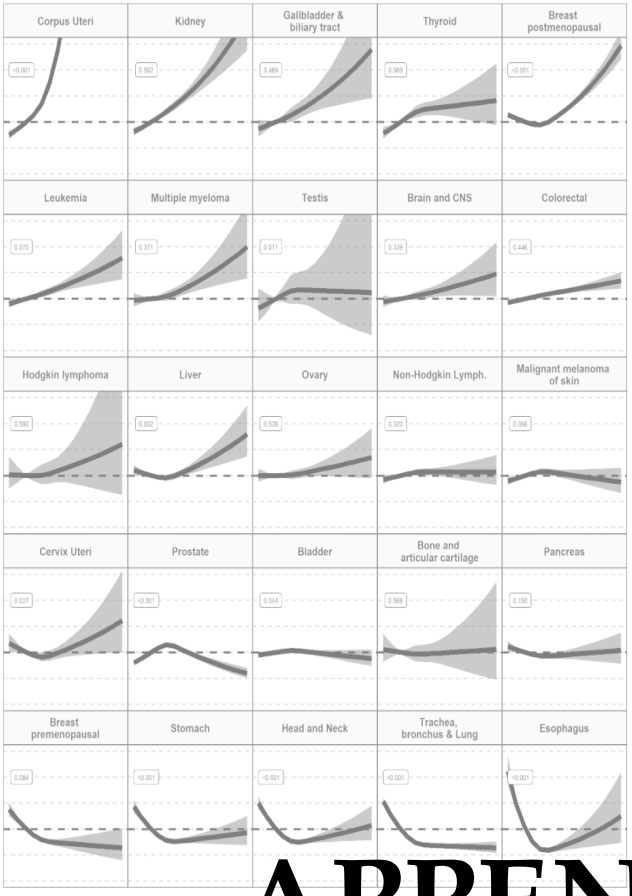
235. Observational Health Data Sciences and Informatics. The Book of OHDSI [Internet]. 2019. 1–470 p. Available from: https://ohdsi.github.io/TheBookOfOhdsi/TheBookOfOhdsi.pdf

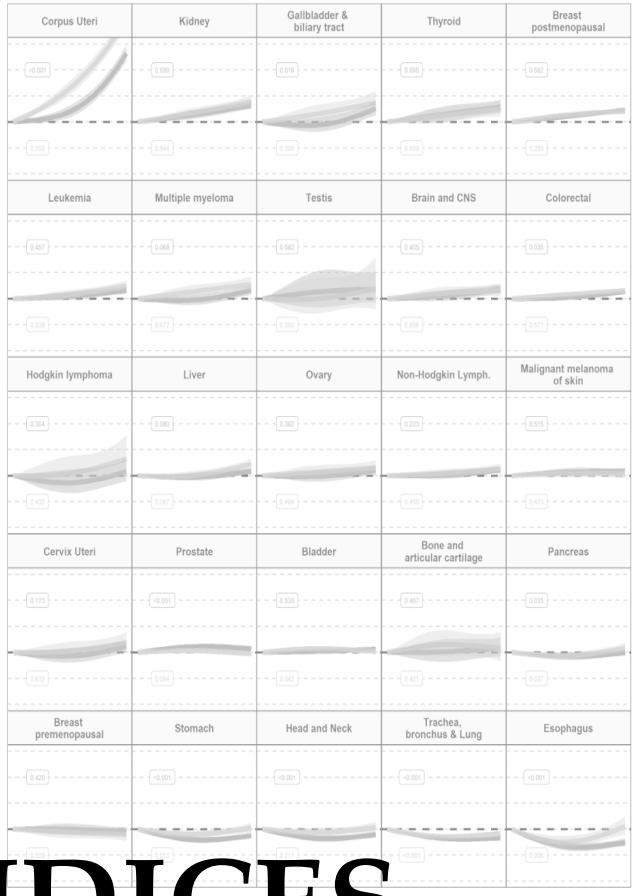
236. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574–8.

237. Gonzalez MC, Correia MITD, Heymsfield SB. A requiem for BMI in the clinical setting. Current Opinion in Clinical Nutrition & Metabolic Care [Internet]. 2017;20(5). Available from: https://journals.lww.com/co-clinicalnutrition/Fulltext/2017/09000/A_requiem_for_BMI_in_the_clinical_setting.3.aspx

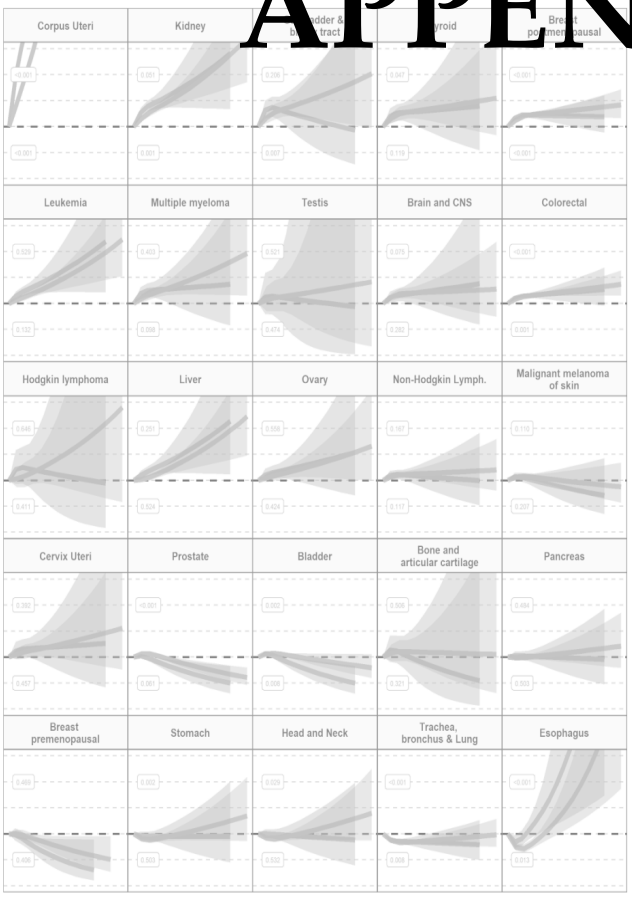
238. Oza S, Thun MJ, Henley SJ, Lopez AD, Ezzati M. How many deaths are attributable to smoking in the United States? Comparison of methods for estimating smoking-attributable mortality when smoking prevalence changes. Preventive Medicine [Internet]. 2011;52(6):428–33. Available from: https://www.sciencedirect.com/science/article/pii/S0091743511001605

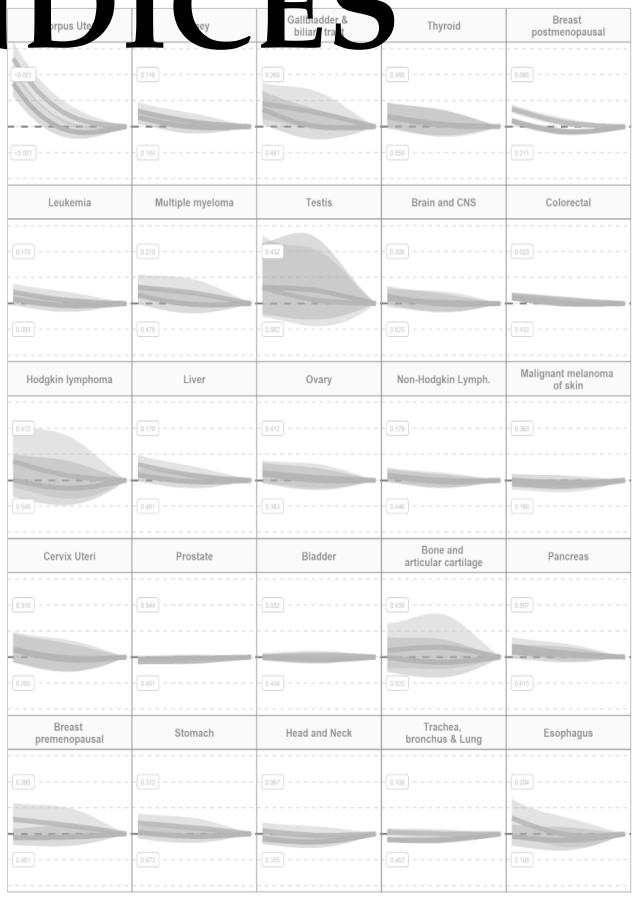
239. Markozannes G, Kanellopoulou A, Dimopoulou O, Kosmidis D, Zhang X, Wang L, et al. Systematic review of Mendelian randomization studies on risk of cancer. BMC Medicine [Internet]. 2022;20(1):41. Available from: https://doi.org/10.1186/s12916-022-02246-y


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Kidney
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

APPENDICES


Corpus Uteri
Kidney
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus


Corpus Uteri
Gallbladder & biliary tract
Thyroid
Breast postmenopausal
Leukemia
Multiple myeloma
Testis
Brain and CNS
Colorectal
Hodgkin lymphoma
Liver
Ovary
Non-Hodgkin Lymph.
Malignant melanoma of skin
Cervix Uteri
Prostate
Bladder
Bone and articular cartilage
Pancreas
Breast premenopausal
Stomach
Head and Neck
Trachea, bronchus & Lung
Esophagus

9. Appendices

In this section, we included the manuscripts corresponding to Studies IV and V which have been submitted to scientific journals. These studies are the continuation of Studies I-III which are manuscripts published in scientific journals.

9.1. Appendix 1: Study IV

Recalde M, Pistillo A, Davila-Batista V, Leitzmann M, Romieu I, Viallon V, Freisling H, Duarte-Salles T.

Longitudinal body mass index-derived exposures during early adulthood and risk of 26 types of cancer: a cohort study of 2.6 million adults in Catalonia, Spain

Submitted to a scientific journal

Longitudinal body mass index-derived exposures during early adulthood and risk of 26 types of cancer: a cohort study of 2.6 million adults in Catalonia, Spain

Martina Recalde[1,2,3†], Andrea Pistillo[2†], Veronica Davila-Batista[1,4], Michael Leitzmann[5], Isabelle Romieu[6], Vivian Viallon[1], Heinz Freisling[1††], Talita Duarte-Salles[2††]

†Joint first authors
††Joint senior authorship

1. International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert Thomas, 69008 Lyon, France
2. Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
3. Universitat Autònoma de Barcelona, Spain
4. Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, 28029, Madrid, Spain
5. Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany
6. Center for Research on Population Health, National Institute of Public Health, Mexico City, Mexico

Corresponding authors:
Talita Duarte-Salles
tduarte@idiapjgol.org
Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
Gran Via Corts Catalanes, 587 àtic
08007 Barcelona, Spain

Heinz Freisling
FreislingH@iarc.fr
International Agency for Research on Cancer (IARC-WHO)
150 Cours Albert Thomas
69008 Lyon, France

Veronica Davila-Batista
vdavbat@gobiernodecanarias.org
Endocrinology Department, Complejo Hospitalario Universitario Insular Materno Infantil, Las Palmas de GC, Canary Health Service
Av. Marítima s/n
35016 - Las Palmas de Gran Canaria, Spain

Keywords: cumulative exposure, duration, age of onset, overweight, obesity, electronic health records, body mass index, trajectories, cancer, cancer incidence

ABSTRACT

Background: Single body mass index (BMI) measurements have been associated with an increased risk of 13 cancers. Whether life course adiposity-related exposures are more relevant cancer risk factors than baseline BMI remains unclear. We aimed to investigate the association of baseline and longitudinal BMI-derived exposures during early adulthood with risk of 26 cancers.

Methods: We conducted a cohort study from 2009 until 2018 with population-based electronic health records in Catalonia, Spain. We included 2,645,885 individuals aged ≥40 years and free of cancer in 2009. The exposures were baseline BMI (as of 2009) and six longitudinal BMI-derived exposures for overweight and obesity (duration of, cumulative exposure to, and age of onset of BMI≥25 and ≥30kg/m^2, respectively) calculated between 18 and 40 years. The main outcome measures were cause-specific hazard ratios (HRs) for incident cancer at 26 anatomical sites.

Results: After 9 years of follow-up, 225,396 participants were diagnosed with cancer. Baseline BMI and longitudinal adiposity exposures were positively related to the risk of cancers of the *corpus uteri* (HR, 95%CI per 10-year increment of duration of BMI≥25: 1.46, 1.42-1.51), *kidney*, *gallbladder*, *thyroid*, *breast (postmenopausal)*, *brain*, *leukemia, multiple myeloma*, *colorectal*, *liver* (1.04, 1.01-1.07) (in decreasing order of HRs) and, among never smokers, of *head and neck* and *bladder* cancers. Longitudinal exposures, but not baseline BMI, were positively associated with the risk of *non-Hodgkin lymphoma, malignant melanoma of the skin,* and *ovary*, *prostate*, *pancreas*, and *stomach* cancers.

Conclusions: Longer duration, greater degree, and younger age of onset of overweight and obesity during early adulthood are positively associated with the risk of 18 cancers, including *leukemia*, *non-Hodgkin lymphoma*, and *head and neck,* and *bladder* cancers which are not yet considered as obesity-related cancers in the literature. Our findings support public health strategies for cancer prevention focussing on preventing and reducing early overweight and obesity.

INTRODUCTION

In 2016, 1.9 billion and 650 million adults were living with overweight and obesity, respectively.[1] Body mass index (BMI), the most common indicator to capture overweight (BMI≥25kg/m^2) and obesity (BMI≥30kg/m^2), has been convincingly associated with the risk of at least 13 cancer types.[2] However, previous studies have mostly focussed on single BMI measurements assessed at study baseline, which are measures of current BMI status. Whether overweight and obesity over the life course are more relevant risk factors for cancer remains unclear.[3–5] Capturing longitudinal BMI-derived exposures might better reflect the underlying biological mechanisms between long-term exposure to adiposity and cancer development. At an epidemiological level, this could translate into stronger associations between adiposity and obesity-related cancer risk and into adiposity being linked to a larger number of cancer types than currently recognized.

Few studies have investigated the association between longitudinal BMI-derived exposures and cancer risk. These exposures included duration of years lived with a BMI≥25 or ≥30kg/m^2 and cumulative exposure (an indicator considering degree and duration of overweight/obesity) to a BMI≥25 or ≥30kg/m^2, which have been positively associated with risk of cancers of the colorectum, postmenopausal breast, endometrium, kidney, pancreas, and multiple myeloma.[2,6–10] Studies investigating age of onset of a BMI≥25 or ≥30kg/m^2 in relation to cancer risk are currently lacking. Yet, such knowledge could identify periods of age, when overweight/obesity are most relevant to cancer risk.

Prior studies have provided insights into the longitudinal BMI-derived exposures-cancer association but did not formally compare cancer risk estimates of longitudinal exposures to those of baseline BMI. Other limitations involve excluding individuals without BMI information (increasing the risk of selection bias), having limited sample sizes that preclude the analysis of a wider range of cancers, or relying on self-reported and recalled weight and height, which could increase the likelihood of exposure misclassification. A study with BMI data measured by health professionals, capturing incident cancer cases from a large and representative population, and using advanced multiple imputation techniques to BMI for all eligible participants could help gain understanding of the adiposity–cancer association through a life-course perspective.

We investigated the association between duration of years lived with a BMI≥25 and ≥30kg/m^2, cumulative exposure to a BMI≥25 and ≥30kg/m^2, age of onset of a BMI≥25 and ≥30kg/m^2 during early adulthood (18 to 40 years) and BMI at baseline in relation to the risk of 26 cancer types.

METHODS

Study design, setting, and data sources

We conducted a population-based cohort study from January 1st, 2009 (index date or baseline date) to December 31st, 2018, using prospectively collected primary care records from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org) in Catalonia, Spain. SIDIAP contains pseudo-anonymized records for >8 million people since 2006.[11] It covers >75% of the population of Catalonia and is representative of the general population of Catalonia by age, sex, and geographic distribution.[11] SIDIAP contains longitudinal data on anthropometric measurements, disease diagnoses (International Classification for Diseases, 10th revision [ICD-10]), sociodemographic and lifestyle information, among others. SIDIAP can be linked to the Minimum Basic Dataset (CMBD), a national population-based registry that includes hospital discharge information of mandatory registration.[12]

Participants

We included individuals aged ≥40 years on January 1st, 2009. We excluded individuals without one year of history in SIDIAP (to capture their baseline characteristics), and/or with a cancer diagnosis prior to index date (Figure S1). We followed up participants from one year after index date (to minimize the possibility of reverse causality [ie, BMI affected by undiagnosed cancer]) until the earliest of cancer diagnosis (any cancer, except other cancer and unspecified malignant neoplasm of the skin), death, transferral out of the SIDIAP catchment area, or end of the study period (31st December 2018), whichever occurred first.

Assessment of variables

To calculate BMI trajectories we extracted data on BMI measurements (before applying multiple imputations). These were calculated using the weight and height of individuals assessed in a standardized manner by general practitioners or nurses in clinical practice.[13]

The outcomes were incident diagnoses of 26 cancer types (*head and neck*; *esophagus*; *stomach*; *colorectal*; *liver*; *gallbladder and biliary tract*; *pancreas*; *larynx*; *trachea, bronchus, and lung*; *bone and articular cartilage*; *malignant melanoma of skin*; *breast* [categorized into pre and postmenopausal due to well-established evidence indicating different relations with BMI];[14] *cervix uteri*; *corpus uteri*; *ovary*; *prostate*; *testis*; *kidney*; *bladder*; *brain and central nervous system* [CNS]; *thyroid*; *Hodgkin lymphoma*; *non-Hodgkin lymphoma*; *multiple myeloma*; and *leukemia*) that have been previously validated in SIDIAP (including the CMBD).[15] We identified cancer diagnoses using ICD-10 and ICD-9 codes recorded in the SIDIAP and CMBD databases, respectively (Table S1).

Potential confounding variables that we were able to consider were age (in 5-year categories) at index date, sex (female, male), geographic region of nationality (Spanish, Global North, or Global South),[16] the *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales* (MEDEA) deprivation index (an ecological index calculated in urban census tracts, categorized into quintiles by SIDIAP to which we added a rural category since the index was unavailable for participants living in those areas),[17] smoking status (never, former, or current smoker), and alcohol intake (no, low or high risk) which is constructed based on type of alcoholic drink, amount, situation, and frequency of consumption.[18]

Statistical analyses

We used a two-step approach for the statistical analyses. Firstly, we estimated life-course BMI trajectories among individuals aged ≥18 years (we excluded those without one year of history or follow-up before and after, respectively, their entry into SIDIAP, n=5,279,567). Secondly, we used these trajectories to construct longitudinal BMI-derived exposures among the study participants and we investigated their association with cancer risk using survival models.

Calculation of BMI trajectories

We applied multilevel time raster multiple imputation to BMI at six age points to obtain BMI trajectories.[19] We used a linear mixed-effects model and 5 imputations to obtain imputed BMI measurements for all eligible participants (a detailed methodological explanation is available in Appendix 1). To construct the life-course trajectories, we joined two contiguous BMI measurements (ie, between two consecutive age points) with a straight line. This method has previously been used to assess longitudinal changes of BMI in SIDIAP.[20]

Calculation of exposures

We used the BMI trajectories to calculate the exposures and we subsequently analyzed their associations with cancer risk (time-to-event analysis). The window to capture longitudinal exposures was between the ages of 18 and 40 years and was separated from the time-to-event window, which extended from the age of an individual (≥40 years for everyone) one year after index date until the age at end of follow-up (Figure S2). We generated six longitudinal exposures. The *duration of BMI≥25* kg/m^2 (and of *≥30*, respectively) was the sum of years lived with a BMI≥25 (≥30) kg/m^2. *Cumulative exposure to a BMI≥25* kg/m^2 (and *≥30*) was calculated by summing the differences between the BMI measurements that were ≥25 (≥30) kg/m^2 and 24.9 (29.9) kg/m^2 for every year lived with a BMI≥25 (≥30) kg/m^2. For all other years, the value of the cumulative exposure was set to 0.[21,22] *Age of onset of a BMI≥25* (and *≥30*) kg/m^2 was the age at which a person had a BMI measurement ≥25 (≥30) kg/m^2 for the first time in the trajectory and was only available for individuals who ever had a BMI≥25 (≥30) kg/m^2. Figure S3 shows graphical representations of the exposures. For comparability, we also considered BMI at index date (or at baseline, on January 1st, 2009) as an exposure.

Association between BMI-derived exposures and cancer risk

We investigated the association between each of the exposures with the risk of the 26 cancer types by running Cox proportional hazard models with age as the underlying time metric.[23] The minimally-adjusted models included one exposure at a time and were adjusted for sex and stratified by age (5-year categories). The fully-adjusted models were further adjusted for the geographic region of nationality, MEDEA deprivation index, smoking status, and alcohol intake. We guided our decisions on the control for confounding by using a directed acyclic graph (Figure S4).[24] We multiply imputed covariates with missing data at baseline (using predictive mean matching, with 5 imputations drawn) (Appendix 1) and we checked the proportional hazard assumptions for the variables included in the models by visual inspection of survival curves. We calculated the hazard ratios (HRs) and their respective 95% confidence intervals (CIs) for each cancer type per 1 standard deviation (SD) increment of each exposure to allow comparability between the different HRs.[25] We checked whether the 95% CIs of the HR of each longitudinal exposure overlapped with that of BMI at index date to assess differences in the strength of the associations between the longitudinal exposures and BMI at index date. For better interpretability, we inverted the HRs of the models including age of onset as the main exposure (ie, HRs >1 indicate greater risk at younger ages). We also fitted models

using restricted cubic splines for the exposures with 3 knots (placed at the 10th, 50th, and 90th percentiles).[26,27] We evaluated linearity by comparing the difference in log-likelihood of the models with each exposure as a linear and non-linear term.

We conducted four secondary analyses to contextualize our findings. We stratified the analyses by age at index date at two arbitrarily selected age points (<65 or ≥65) and sex. We mutually adjusted the models for the association of age of onset of a BMI≥25 (and ≥30) and duration of BMI≥25 (and ≥30) kg/m^2 and cancer risk. We restricted the analyses to never smokers to account for possible residual confounding by smoking.[28] We compared the Harrell's C-indices of the models with BMI at index date as the main exposure to the same models further adjusted for each longitudinal exposure separately to assess if the longitudinal exposures improve cancer risk discrimination compared to the standard baseline BMI criterion.[29]

We conducted three sensitivity analyses to assess the robustness of our findings. We i)further adjusted our models by the difference between the BMI at index date and at 40 years to account for changes in BMI between the start of follow-up and the end of the longitudinal exposure window (see graphical representation in Figure S2), ii)restricted the analyses to individuals with ≥1 BMI assessment in their health records, and iii)extended the start of the follow-up period from one to three years after index date to account for potential reverse causality.

RESULTS

Of the 3,247,244 individuals who were eligible to enter the study, we excluded 172,800, 190,171, and 238,388 persons who had <1 year of history in SIDIAP, prior history of cancer, and <1 year of follow-up, respectively (Figure S1).

Among 2,645,885 participants followed up for a median time of 9 (interquartile range [IQR]: 8-9) years, 225,396 (9%) individuals were diagnosed with any of the 26 cancers of interest (Table 1). The median age of the participants was 56 (IQR: 47-68) years, the median BMI at index date (baseline) was 28 (24-31)kg/m^2, and 47% were males. The median duration of BMI≥25 and ≥30kg/m^2, respectively, were 12 (0-23) and 0 (0-4) years. The median cumulative exposure to BMI≥25 and to BMI ≥30m/kg^2 were 16 (0-74) cumulative overweight-years and 0 (0-2) cumulative obese-years, respectively. Of all participants, 1,833,516 (69%) ever had a BMI≥25kg/m^2 (median age of onset of BMI≥25 was 20 [IQR: 18-29] years), of which 801,612 (30% of all participants) ever had a BMI≥30kg/m^2 (median age of onset of BMI≥30 was 29 [21-35] years). Those who never had a BMI≥25kg/m^2 were more likely to live in the least

deprived areas of Catalonia, to be current smokers, and to have fewer comorbidities than those who ever had a BMI≥25kg/m^2 (Table 1).

Association between BMI-derived exposures and cancer risk

In fully adjusted models, longer duration of a BMI≥25 (≥30) kg/m^2 was positively associated with the risk of 14 (12) cancers, higher cumulative exposure to a BMI≥25 (≥30) kg/m^2 with 13 (11), age of onset of a BMI≥25 (≥30) kg/m^2 with 11 (10), and BMI at index date with 10 cancers (Figure 1, Table S2 & S3). All exposures were positively associated with the risk of the following eight cancer types: *corpus uteri* (eg, HR, 95%CI per 10-year [1-SD] increment of duration of a BMI≥25: 1.46, 1.42-1.51), *kidney*, *gallbladder and biliary tract*, *breast postmenopausal*, *leukemia, multiple myeloma*, *colorectal*, and *liver* (HR of duration of a BMI≥25: 1.04, 1.01-1.07) cancers. All exposures except age of onset of a BMI≥25 and/or ≥30, were also positively associated with the risk of two cancers: *thyroid* (eg, HR, 95% CI per 70-cumulative overweight-year [1-SD] increment of cumulative exposure of a BMI≥25: 1.08, 1.04-1.12), and *brain and CNS* (same eg,: 1.06, 1.02-1.10). There were nuances in the shape of the relationship of some of the exposures with the risk of six of these cancers (p-value for non-linearity <0.05) (Figures 2, 3, and 4, Figure S5). For instance, there was a stronger association between cumulative exposure to a BMI≥25 and/or ≥30 and the risk of *colorectal*, *gallbladder and biliary tract*, *breast postmenopausal*, *thyroid*, and *kidney* cancers at lower values of these exposures, after which the increase in risk diminished. For *corpus uteri* cancer, the risk increased faster than linear at higher values of most exposures. The longitudinal exposures had a similar strength of association with the above mentioned 10 cancer types compared to BMI at index date (in linear models), except for *corpus uteri* cancer which was stronger for the latter (eg, BMI at index date 1.55 [1.51-1.58] vs cumulative exposure to a BMI≥30: 1.29 [1.27-1.31]) (Figure 1, Table S2). The results of the minimally- and fully-adjusted models were similar (Figure S6).

Contrary to BMI at index date, one or more of these longitudinal exposures were also positively associated with the risk of seven cancer types including cancers of the *ovary*, *non-Hodgkin lymphoma*, *bladder*, *malignant melanoma of skin*, *prostate*, *pancreas*, and *stomach* (Figure 1, Table S2). Duration of BMI≥25 and ≥30, cumulative exposure to a BMI≥25, and age of onset of a BMI≥25 were all positively associated with the risk of *non-Hodgkin lymphoma*. Duration of BMI≥30 and cumulative exposure to a BMI≥25 and ≥30 were positively related to the risk of *ovarian* cancer. Longer duration of BMI≥25 and higher cumulative exposure to a BMI≥25

were positively related to the risk of *bladder* cancer. Although in non-linear analyses only lower levels of cumulative exposure to a BMI≥25 were positively linked to *bladder* cancer (Figure 3). Duration of BMI≥25 was further associated with risk of *malignant melanoma of the skin*, and *prostate* (for which the association had an attenuated, inverted U-shape in non-linear analyses) cancers (Figure 3). Age of onset of a BMI≥25 and ≥30 were both related to a higher risk of *pancreatic* cancer, whereas only BMI≥30 was associated with a greater risk of *stomach* cancer.

A higher BMI at index date was inversely associated with the risk of six cancer types, of which five were also inversely linked to duration of a BMI≥25 kg/m^2, including cancers of the *stomach* and respiratory tract (*esophagus* [HR, 95% CI: 0.88, 0.82-0.93], *larynx*, *trachea, bronchus, and lung*, and *head and neck* [0.95, 0.92-0.98]) cancers (Figure 1). These associations were found to be non-linear (Figures 2 and Figure S5), but while the relationships were L-shaped for BMI at index date, they had an attenuated inverted U-shape for duration of a BMI≥25 (which were similarly shaped for BMI≥30, albeit closer to 1). In addition, although cumulative exposure to a BMI≥25 and/or ≥30 was only inversely related to the risk of cancers of the *larynx* and *trachea, bronchus, and lung* cancers in linear analyses, in non-linear models these exposures were related to the risk of *stomach* and the four respiratory tract cancers in a J-shaped fashion (Figures 1 and 3). Age of onset of a BMI≥25 was inversely associated with the risk of *larynx* cancer).

The results of the supplementary and sensitivity analyses are described in Appendix 2 and reported in Figures S7, S8, S9, S10, S11, and S12. The inverse associations (for *stomach* and respiratory tract cancers) became null when we restricted the analyses to never smokers. Moreover, BMI at index date, duration of, and cumulative exposure to a BMI ≥25 (≥30) became positively and more pronouncedly, respectively, associated with *head and neck* and *bladder* cancers (Figure S10). Our results were similar to those from three sensitivity analyses (Figure S12).

DISCUSSION

Main findings

In this population-based cohort study that included 2,645,885 individuals living in Catalonia, Spain, we found that longitudinal BMI-derived exposures and BMI at index date were positively associated with the risk of 12 cancers (*corpus uteri*, *kidney*, *gallbladder and biliary*

tract, *multiple myeloma*, *leukemia*, *breast postmenopausal*, *colorectal*, *liver, thyroid*, *brain and CNS,* as well as *head and neck* and *bladder* [among never smokers]). Some longitudinal exposures, but not BMI at index date, were additionally positively associated with the risk of six cancer types (*ovary*, *non-Hodgkin lymphoma*, *malignant melanoma of skin*, *prostate*, *pancreas*, and *stomach* cancers). BMI at index date and overweight duration were inversely associated with the risk of *stomach* and respiratory tract cancers, which likely indicates residual confounding by smoking since these associations were attenuated towards unity when we restricted these analyses to individuals who never smoked.

Interpretation

A single measurement of BMI at study baseline has been convincingly associated with the risk of 13 cancer types in previous studies, of which 10 (*colorectum*, *liver*, *gallbladder and biliary tract*, *post-menopausal breast*, *corpus uteri*, *ovary*, *kidney*, *brain and CNS*, *thyroid*, and *multiple myeloma)* cancers were also positively associated with the longitudinal BMI-derived exposures we investigated.[2] Thus, our findings seem to indicate that longer exposures to overweight and obesity (with or without accounting for the degree of overweight and obesity), as well as developing overweight and obesity at younger ages in early adulthood might increase cancer risk. This suggests that overweight and obesity prevention should start in early adulthood and that weight management and weight loss interventions leading to shorter durations of overweight and obesity might reduce cancer incidence.

We also provide novel evidence that longitudinal BMI-derived exposures and/or BMI at index date are positively associated with the risk of *leukemia*, *non-Hodgkin lymphoma*, *malignant melanoma of the skin*, *prostate*, and among never smokers only and more pronouncedly, respectively, with *head and neck,* and *bladder* cancers, all of which are not yet considered as obesity-related cancers in the literature.[2] Furthermore, for some of these cancers (*non-Hodgkin lymphoma*, *malignant melanoma of the skin*, *prostate*, and -in the main analysis- *bladder* cancers) we only found positive associations for the longitudinal exposures (not for BMI at index date), which highlights that these longitudinal adiposity-related exposures provide additional information compared to a single measure of BMI in time. These additional associations might also indicate that the longitudinal exposures we considered better reflect, than baseline BMI, the underlying biological mechanisms between long-term exposure to adiposity and cancer development.

The IARC viewpoint on excess body fatness and cancer risk considered the evidence as inadequate for *leukemia* and *non-Hodgkin lymphoma*.[2] However, four meta-analyses have reported the association between BMI and higher risk of *leukemia* and *non-Hodgkin lymphoma* (or only of diffuse large B cell lymphoma).[3,30–32] Our findings support and extend these results by providing evidence that higher levels of adiposity through a life course perspective are consistently associated with the risk of hematological cancers, including *multiple myeloma*, *leukemia,* and *non-Hodgkin lymphoma*. Furthermore, we showed that among individuals who never smoked, higher levels of baseline BMI, and longer exposures to overweight and obesity (with or without accounting for the degree of overweight/obesity) are positively associated with the risk of *head and neck* and *bladder* cancers which expands on the extent to which adiposity can affect cancer risk. The three mechanisms by which greater overall adiposity may increase cancer risk have been extensively reported in the literature (sex hormone metabolism, insulin and insulin-like growth factors (IGF) signaling, and adipokine pathways) and could also explain some of the associations between longitudinal BMI-derived exposures and cancer risk (eg, *corpus uteri*, *breast postmenopausal*, *colorectal* cancers).[14,33–39] However, other pathways may be involved in the risk of cancer types not yet considered obesity-related and require further research.

Moreover, duration of BMI≥25 kg/m^2 was the only exposure positively associated with the risk of *malignant melanoma of the skin* and *prostate* cancers. This is in line with what was observed in the non-linear analysis of the association between BMI at index date and risk of these cancers (in this and other studies), where an inverted U-shaped association was found, indicating a higher risk of these cancers only for BMI in the overweight range.[40,41] Future research should focus on confirming these associations and on understanding the pathways by which only being overweight (and a longer duration of it) could have a harmful effect on the risk of these cancers. On the other hand, while higher levels of BMI have been convincingly associated with risk of *pancreatic* and gastric cardia cancers,[2] in our study we only found a positive association with respect to age of onset of a BMI≥25 (≥30) kg/m^2. The lack of association with *stomach* cancer for other exposures could be due to our inability to distinguish gastric cardia (obesity-related) from non-cardia cancers (in Spain, the incidence of the non-obesity-related subsite of this cancer is higher than the obesity-related one).[42] Finally, greater levels of baseline BMI and longer duration of overweight and obesity were inversely associated with risk of respiratory tract cancers in the main analyses, but were attenuated towards unity in the analysis restricted to never smokers. Previous studies have also reported these inverse associations for baseline

BMI,[40,41] which have hypothesized that this might be due to residual confounding by smoking (lower BMI levels as an approximation of heavy smoking). We hypothesize that the findings for duration of overweight and obesity might also be due to residual confounding by smoking (no or short periods with a BMI≥25 as an approximation of heavy smoking), but further research specifically focussing on this is needed.

Strengths and limitations

This study has several strengths. To our knowledge, this is the first study to analyze the associations between several BMI-derived longitudinal exposures and the risk of numerous (26) cancer types in a single and sufficiently powered data set, including systematic investigation of non-linearity. SIDIAP is representative of the general population of Catalonia in terms of age, sex, and geographic distribution, which lends external validity to our results.[11] Thanks to the advanced multiple imputation approach for the BMI trajectories, we were able to include all individuals eligible to enter the study, likely minimizing the possibility of selection bias. The diagnoses of the cancer types considered as outcomes have been previously validated and used for BMI and cancer-related research.[15,20,40] While we cannot discard the possibility of outcome misclassification, this was likely not differential according to the exposures, thus, this probably did not greatly affect our results.

Our findings should be interpreted in light of some limitations. The major limitation is that due to the length of follow-up available (12 years), we exclusively relied on multiple imputed BMI measurements for the exposure window. This could have introduced exposure misclassification bias. We aimed to reduce this bias by using high-quality BMI measurements (measured by health professionals and with a distribution shown to be similar to representative studies of the Spanish population)[13,40] and by including data on all adults in SIDIAP (n=5,279,567) for the multiple imputations. We were also empirically reassured about the quality of our exposures given that we found similar associations between BMI-derived longitudinal exposures and risk of specific cancer types that have been previously studied (colorectal, breast postmenopausal, endometrium, kidney, and multiple myeloma cancers).[2,7–10] Another limitation is that the observed BMI measurements of the participants were very close in time between each other difficulting the capture of granularity (eg, weight cycling) in the trajectories. Also, the dispersion of duration of BMI≥30 and cumulative exposure to a BMI≥30 was modest in the overall population; which could explain why for certain cancers (eg, non-Hodgkin lymphoma or bladder cancers) we only observed statistically-significant associations for the respective

exposures of BMI≥25. Finally, for certain potential confounding factors we had limited information (ie, smoking amount or individual-level SES). While we had access to related indicators such as smoking status or the MEDEA deprivation index, we cannot exclude the possibility of some residual confounding.

CONCLUSION

In this large Southern European study, we found that longer duration and greater degree of overweight and obesity during early adulthood as well as younger age of onset of a high BMI are associated with a higher risk of 18 cancer types. We provide novel evidence that adiposity over the life course is positively associated with the risk of *leukemia*, *non-Hodgkin lymphoma*, as well as *head and neck* and *bladder* cancers (among never smokers) and we confirm associations that have been reported in studies focusing on single BMI measurements at study baseline. Our findings reinforce the need for public health strategies for overweight and obesity prevention and reduction in early adulthood for cancer prevention.

Acknowledgments

We would like to thank all healthcare professionals of Catalonia who daily register information in the populations' electronic health records. This study was carried out as part of the Doctoral Program in Biomedical Research Methodology and Public Health at the Autonomous University of Barcelona.

Role of the funding source

Funding was obtained from Wereld Kanker Onderzoek Fonds (WKOF grant number: 2017/1630), as part of the international grants from the World Cancer Research Fund. MR was also funded by World Cancer Research Fund (UK) (grant number: IIG_2019_1978), as part of the World Cancer Research Fund International grant program. TDS acknowledges receiving financial support from the Instituto de Salud Carlos III (ISCIII; Miguel Servet 2021: CP21/00023). The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer and World Health Organization, the authors alone are responsible for the views expressed in this Article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer and World Health Organization.

Declaration of interests

The authors declare no conflicts of interest.

Data sharing

In accordance with current European and national law, the data used in this study is only available for the researchers participating in this study. Thus, we are not allowed to distribute or make publicly available the data to other parties. However, researchers from public institutions can request data from SIDIAP if they comply with certain requirements. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudesen/application-proccedure) or by contacting Anna Moleras (amoleras@idiapjgol.org).

Author contributions

MR performed the literature review. AP did the data management and data analysis with contributions from all authors. MR wrote the first draft with insightful contributions from AP, HF, and TDS. All

authors were involved in the study conception and design, interpretation of the results, manuscript preparation, and approved the final version of the manuscript.

Ethical approval

We obtained approval from the Clinical Research Ethics Committee of the IDIAPJGol (project code: P14/074) to perform this study.

REFERENCES

1. World Health Organization. Overweight and obesity. Published 2021. Accessed August 17, 2021. http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

2. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine*. 2016;375(8):794-798. doi:10.1056/NEJMsr1606602

3. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*. 2008;371(9612):569-578. doi:https://doi.org/10.1016/S0140-6736(08)60269-X

4. Brennan P, Davey-Smith G. Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed. *JNCI: Journal of the National Cancer Institute*. 2022;114(3):353-360. doi:10.1093/jnci/djab204

5. World Cancer Research Fund/American Institute for Cancer Research. *Continuous Update Project Expert Report 2018. Future Research Directions.*; 2018. Accessed March 9, 2022. dietandcancerreport.org

6. Stolzenberg-Solomon RZ, Schairer C, Moore S, Hollenbeck A, Silverman DT. Lifetime adiposity and risk of pancreatic cancer in the NIH-AARP Diet and Health Study cohort. *The American Journal of Clinical Nutrition*. 2013;98(4):1057-1065. doi:10.3945/ajcn.113.058123

7. Arnold M, Freisling H, Stolzenberg-Solomon R, et al. Overweight duration in older adults and cancer risk: a study of cohorts in Europe and the United States. *European Journal of Epidemiology*. 2016;31(9):893-904. doi:10.1007/s10654-016-0169-z

8. Noh H, Charvat H, Freisling H, et al. Cumulative exposure to premenopausal obesity and risk of postmenopausal cancer: A population-based study in Icelandic women. *International Journal of Cancer*. 2020;147(3):793-802. doi:https://doi.org/10.1002/ijc.32805

9. Arnold M, Jiang L, Stefanick ML, et al. Duration of Adulthood Overweight, Obesity, and Cancer Risk in the Women's Health Initiative: A Longitudinal Study from the

United States. *PLOS Medicine*. 2016;13(8):e1002081-. https://doi.org/10.1371/journal.pmed.1002081

10. Marinac CR, Birmann BM, Lee IM, et al. Body mass index throughout adulthood, physical activity, and risk of multiple myeloma: a prospective analysis in three large cohorts. *British Journal of Cancer*. 2018;118(7):1013-1019. doi:10.1038/s41416-018-0010-4

11. Bolíbar B, Fina Avilés F, Morros R, et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Medicina Clinica*. 2012;138(14):617-621.

12. Generalitat de Catalunya. Conjunt mínim bàsic de dades (CMBD). Published 2017. Accessed March 5, 2019. https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/index.html#googtrans(ca%7Ces)

13. Lecube A, Monereo S, Rubio MÁ, et al. Prevención, diagnóstico y tratamiento de la obesidad. Posicionamiento de la Sociedad Española para el Estudio de la Obesidad de 2016. *Endocrinología, Diabetes y Nutrición*. 2017;64:15-22. doi:https://doi.org/10.1016/j.endonu.2016.07.002

14. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Reviews Cancer*. 2004;4(8):579-591. doi:10.1038/nrc1408

15. Recalde M, Manzano-Salgado C, Díaz Y, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. *Clinical Epidemiology*. 2019;Volume 11:1015-1024. doi:10.2147/CLEP.S225568

16. Brandt W. *North-South: A Program for Survival*. MIT Press; 1990.

17. Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). *Gaceta Sanitaria*. 2008;22(3):179-187.

18. Generalitat de Catalunya. Registre del consum d'alcohol a l'e-CAP. Accessed March 22, 2022.

http://www.gencat.cat/salut/butlletins/butlleti_beveu_menys/arxius/pdf/registre_consum_alcohol.pdf

19. van Buuren S. *Flexible Imputation of Missing Data: Time Raster Imputation*. 2nd Edition. Chapman & Hall/CRC; 2012.

20. Recalde M, Pistillo A, Viallon V, Fontvieille E, Duarte-Salles T, Freisling H. Body Mass Index and Incident Cardiometabolic Conditions in Relation to Cancer Risk: A Population-Based Cohort Study in Catalonia, Spain. *SSRN*. Published online 2022.

21. Abdullah A, Wolfe R, Stoelwinder JU, et al. The number of years lived with obesity and the risk of all-cause and cause-specific mortality. *International Journal of Epidemiology*. 2011;40(4):985-996. doi:10.1093/ije/dyr018

22. Abdullah A, Amin FA, Stoelwinder J, et al. Estimating the risk of cardiovascular disease using an obese-years metric. *BMJ Open*. 2014;4(9):e005629. doi:10.1136/bmjopen-2014-005629

23. Kom EL, Graubard BI, Midthune D. Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. *American Journal of Epidemiology*. 1997;145(1):72-80. doi:10.1093/oxfordjournals.aje.a009034

24. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. *Epidemiology*. 1999;10(1).

25. Freisling H, Arnold M, Soerjomataram I, et al. Comparison of general obesity and measures of body fat distribution in older adults in relation to cancer risk: meta-analysis of individual participant data of seven prospective cohorts in Europe. *British Journal of Cancer*. 2017;116(11):1486-1497. doi:10.1038/bjc.2017.106

26. Orsini N, Greenland S. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal*. 2011;11(1):1-29. doi:10.1177/1536867x1101100101

27. Harrell FEJ. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer; 2001.

28. Song M, Giovannucci E. Estimating the Influence of Obesity on Cancer Risk: Stratification by Smoking Is Critical. *Journal of Clinical Oncology*. 2016;34(27):3237-3239. doi:10.1200/JCO.2016.67.6916

29. The Fibrinogen Studies Collaboration. Measures to assess the prognostic ability of the stratified Cox proportional hazards model. *Statistics in Medicine*. 2009;28(3):389-411. doi:https://doi.org/10.1002/sim.3378

30. Larsson SC, Wolk A. Overweight and obesity and incidence of leukemia: A meta-analysis of cohort studies. *International Journal of Cancer*. 2008;122(6):1418-1421. doi:https://doi.org/10.1002/ijc.23176

31. Willett E v, Morton LM, Hartge P, et al. Non-Hodgkin lymphoma and obesity: A pooled analysis from the InterLymph Consortium. *International Journal of Cancer*. 2008;122(9):2062-2070. doi:https://doi.org/10.1002/ijc.23344

32. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: A meta-analysis of prospective studies. *European Journal of Cancer*. 2011;47(16):2422-2430. doi:https://doi.org/10.1016/j.ejca.2011.06.029

33. Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. *Nature Reviews Cancer*. 2015;15(8):484-498. doi:10.1038/nrc3967

34. Khandekar MJ, Cohen P, Spiegelman BM. Molecular mechanisms of cancer development in obesity. *Nature Reviews Cancer*. 2011;11(12):886-895. doi:10.1038/nrc3174

35. Roberts DL, Dive C, Renehan AG. Biological Mechanisms Linking Obesity and Cancer Risk: New Perspectives. *Annual Review of Medicine*. 2010;61(1):301-316. doi:10.1146/annurev.med.080708.082713

36. van Kruijsdijk RCM, van der Wall E, Visseren FLJ. Obesity and Cancer: The Role of Dysfunctional Adipose Tissue. *Cancer Epidemiology, Biomarkers & Prevention*. 2009;18(10):2569-2578. doi:10.1158/1055-9965.EPI-09-0372

37. Dashti SG, Simpson JA, Viallon V, et al. Adiposity and breast, endometrial, and colorectal cancer risk in postmenopausal women: Quantification of the mediating effects

of leptin, C-reactive protein, fasting insulin, and estradiol. *Cancer Medicine*. 2022;11(4):1145-1159. doi:https://doi.org/10.1002/cam4.4434

38. Dashti SG, Simpson JA, Karahalios A, et al. Adiposity and estrogen receptor-positive, postmenopausal breast cancer risk: Quantification of the mediating effects of fasting insulin and free estradiol. *International Journal of Cancer*. 2020;146(6):1541-1552. doi:https://doi.org/10.1002/ijc.32504

39. Dashti SG, English DR, Simpson JA, et al. Adiposity and Endometrial Cancer Risk in Postmenopausal Women: A Sequential Causal Mediation Analysis. *Cancer Epidemiology, Biomarkers & Prevention*. 2021;30(1):104-113. doi:10.1158/1055-9965.EPI-20-0965

40. Recalde M, Davila-Batista V, Díaz Y, et al. Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain. *BMC Medicine*. 2021;19(1):10. doi:10.1186/s12916-020-01877-3

41. Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5.24 million UK adults. *The Lancet*. 2014;384(9945):755-765. doi:10.1016/S0140-6736(14)60892-8

42. Arnold M, Ferlay J, van Berge Henegouwen MI, Soerjomataram I. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. *Gut*. 2020;69(9):1564. doi:10.1136/gutjnl-2020-321600

Tables and Figures

Table 1. Baseline characteristics of the study population, overall and by having ever had a body mass index ≥25 or ≥30 kg/m2, after multiple imputations

Figure 1. Forest plot of hazard ratios of 26 cancer types related to one standard deviation increment of BMI at index date and other longitudinal BMI-derived exposures, with 95% CIs

Figure 2. Hazard ratios of 26 cancer types related to duration of BMI ≥25 and ≥30 kg/m2 in years, with 95% CIs, allowing for non-linearity

Figure 3. Hazard ratios of 26 cancer types related to cumulative exposure to BMI ≥25 and ≥30 kg/m2 in cumulative overweight and obese-years, respectively, with 95% CIs, allowing for non-linearity

Figure 4. Hazard ratios of 26 cancer types related to age of onset of BMI ≥25 and ≥30 kg/m2 (among people who ever had a BMI ≥25 and ≥30 kg/m2, respectively) in years

Table 1. Baseline characteristics of the study population, overall and by having ever had a body mass index ≥25 or ≥30 kg/m^2, after multiple imputations

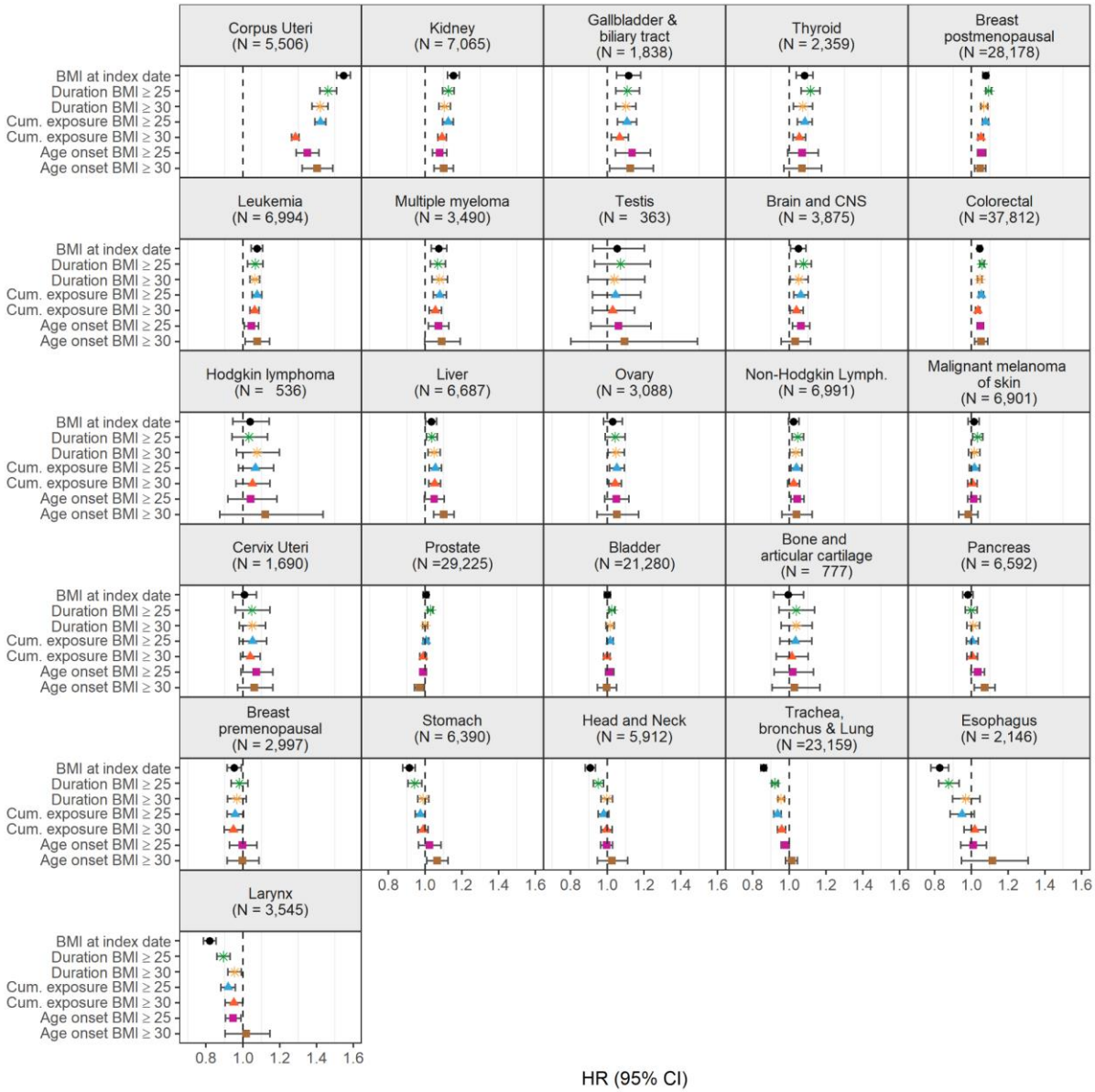
	Overall N (%)	Never overweight (BMI<25 kg/m^2) N (%)[1]	Ever overweight (BMI≥25 kg/m^2) N (%)[1]	Ever obese (BMI≥30 kg/m^2) N (%)[1]
	2,645,885	812,369 (30.7)	1,833,516 (69.3)	801,612 (30.3)
Follow-up time in years, median (IQR)	9.0 (7.7, 9.0)	9.0 (7.9, 9.0)	9.0 (7.7, 9.0)	9.0 (7.5, 9.0)
Duration of BMI ≥25 kg/m^2 in years, median (IQR)[2]	12.0 (0.0, 23.0)	0.0 (0.0, 0.0)	20.0 (10.0, 23.0)	23.0 (23.0, 23.0)
Duration of BMI ≥30 kg/m^2 in years, median (IQR)[2]	0.0 (0.0, 4.0)	0.0 (0.0, 0.0)	0.0 (0.0, 9.0)	11.0 (5.0, 20.0)
Cumulative exposure to BMI ≥25 kg/m^2 in cumulative overweight-years, median (IQR)[2,3]	16.4 (0.0, 73.7)	0.0 (0.0, 0.0)	45.9 (13.7, 103.3)	113.5 (78.3, 163.3)
Cumulative exposure to BMI ≥30 kg/m^2 in cumulative obese-years, median (IQR)[2,3]	0.0 (0.0, 2.2)	0.0 (0.0, 0.0)	0.0 (0.0, 12.4)	17.4 (4.2, 51.9)
Age of onset of BMI ≥25 kg/m^2 in years, median (IQR)[2,4]	20.0 (18.0, 29.0)	-	20.0 (18.0, 29.0)	18.0 (18.0, 18.0)
Age of onset of BMI ≥30 kg/m^2 in years, median (IQR)[2,4]	29.0 (21.0, 35.0)	-	29.0 (21.0, 35.0)	29.0 (21.0, 35.0)
BMI at index date in kg/m^2, median (IQR)[2,5]	27.6 (24.2, 31.1)	23.0 (20.7, 24.9)	29.4 (26.8, 32.5)	32.5 (30.5, 35.2)
Age in years, median (IQR)	56.0 (47.0, 68.0)	55.0 (46.0, 66.0)	57.0 (47.0, 70.0)	58.0 (48.0, 71.0)
Male sex, n (%)	1,241,523 (46.9)	362,147 (44.6)	879,376 (48.0)	358,172 (44.7)
Nationality				
Spanish	2,495,536 (94.3)	766,176 (94.3)	1,729,360 (94.3)	756,163 (94.3)
Global North	51,320 (1.9)	17,049 (2.1)	34,271 (1.9)	14,834 (1.9)
Global South	99,029 (3.7)	29,145 (3.6)	69,884 (3.8)	30,616 (3.8)
MEDEA deprivation index, n (%)[2]				
Quintile 1 (least deprived)	472,049 (17.8)	170,403 (21.0)	301,646 (16.5)	120,028 (15.0)
Quintile 2	429,823 (16.2)	136,784 (16.8)	293,039 (16.0)	124,672 (15.6)
Quintile 3	416,465 (15.7)	123,903 (15.3)	292,562 (16.0)	128,865 (16.1)
Quintile 4	401,681 (15.2)	112,463 (13.8)	289,218 (15.8)	131,747 (16.4)
Quintile 5 (most deprived)	361,665 (13.7)	96,963 (11.9)	264,702 (14.4)	125,063 (15.6)
Rural	564,201 (21.3)	171,853 (21.2)	392,348 (21.4)	171,237 (21.4)
Smoking status, n (%)[2]				

Never smoker	1,663,154 (62.9)	486,100 (59.8)	1,177,054 (64.2)	529,478 (66.1)
Former smoker	390,711 (14.8)	110,853 (13.6)	279,858 (15.3)	122,089 (15.2)
Current smoker	592,020 (22.4)	215,416 (26.5)	376,604 (20.5)	150,046 (18.7)
Alcohol intake, n (%)[2]				
No risk	1,663,281 (62.9)	501,729 (61.8)	1,161,553 (63.4)	526,049 (65.6)
Low risk	894,238 (33.8)	283,422 (34.9)	610,816 (33.3)	249,140 (31.1)
High risk	88,366 (3.3)	27,218 (3.4)	61,147 (3.3)	26,423 (3.3)
Charlson comorbidity index, n (%)				
0	1,250,781 (47.3)	439,775 (54.1)	811,006 (44.2)	323,454 (40.4)
1	892,103 (33.7)	243,404 (30.0)	648,699 (35.4)	302,468 (37.7)
2	357,217 (13.5)	94,760 (11.7)	262,457 (14.3)	121,705 (15.2)
≥ 3	145,784 (5.5)	34,430 (4.2)	111,354 (6.1)	53,986 (6.7)
Cause of exit from the study, n (%)				
End of study	1,865,496 (70.5)	577,856 (71.1)	1,287,640 (70.2)	557,511 (69.5)
Transferred out of the SIDIAP	291,641 (11.0)	94,850 (11.7)	196,791 (10.7)	85,001 (10.6)
Death	250,914 (9.5)	71,727 (8.8)	179,187 (9.8)	83,541 (10.4)
Any cancer[6]	237,834 (9.0)	67,935 (8.4)	169,899 (9.3)	75,559 (9.4)
Cancer outcomes, n (%)	225,396 (8.5)	64,466 (7.9)	160,930 (8.8)	71,456 (8.9)

Notes: 1) This categorization was done in the 5 datasets obtained after performing the multiple imputations. For visualization purposes and in order for the categorical variables to add up to 2,645,885 we divided the n for the categorical variables by 5. 2) The exposures of interest, the MEDEA deprivation index, smoking status, and alcohol intake were calculated using the multiple imputation approach, with 5 data sets created. For visualization purposes, we divided the n for the categorical variables by 5. 3) This indicator was calculated by adding the difference between the BMI measurements that were ≥25 (≥30, for obesity) kg/m^2 and 24.9 (29.9) kg/m^2 for every year lived with a BMI ≥25 and ≥30, respectively. 4) Age of onset of a BMI ≥25 (and ≥30) kg/m^2 is only available for individuals who ever had a BMI ≥25 (≥30) kg/m^2. 5) BMI assessment at the start of the time-to-event analysis (baseline BMI). 6) Any cancer does not include non-melanoma skin cancer.

Abbreviations: BMI: Body Mass Index; IQR: Interquartile range; MEDEA: "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales"; SIDIAP: Information System for Research in Primary Care.

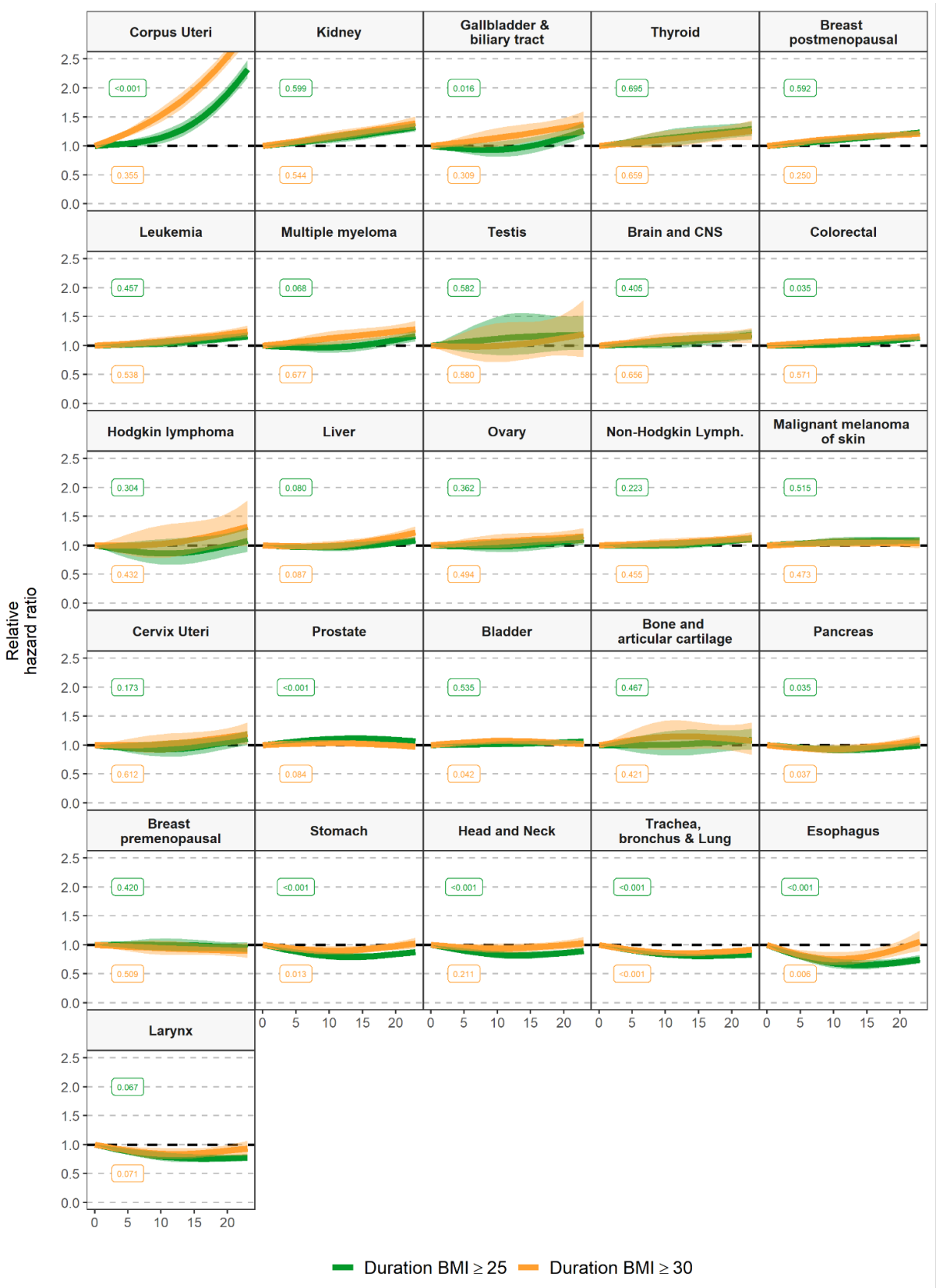
Figure 1. Forest plot of hazard ratios of 26 cancer types related to one standard deviation increment of BMI at index date and other longitudinal BMI-derived exposures, with 95% CIs



Notes: Models are adjusted for geographic region of nationality, the MEDEA deprivation index, smoking status, and alcohol intake and stratified by age (5-year categories). Cumulative exposure is an exposure considering both degree and duration of overweight/obesity which is obtained by adding the difference between the BMI measurements that were ≥25 (≥30) kg/m^2 and 24.9 (29.9) kg/m^2 for every year lived with a BMI ≥25 and ≥30, respectively. Age of onset of a BMI ≥25 (and ≥30) kg/m^2 is only available for individuals who ever had a BMI ≥25 (≥30) kg/m^2 (N of cases are in Table S4) and the HRs of these exposures were inverted for visualization purposes (ie, an HR>1 means a greater risk at younger ages). Cancer types are ordered by descending ranking of the HRs for BMI at index date. The SD for each exposure were: 10 years for duration of BMI≥25 and 7 years of BMI≥30 kg/m^2, 69 cumulative overweight-years for cumulative exposure to a BMI≥25 and 36 cumulative obese-years to a BMI≥30kg/m^2, 7 years for age of onset of a BMI ≥25 and 8 years ≥30kg/m^2. Models for ovary, cervix, and corpus uteri cancers were only computed in females, for breast pre-menopausal only in pre-menopausal females, for breast post-menopausal only in post-menopausal females, and for prostate and

testis only computed in males (their respective SDs can be consulted in Table S3). Brain and CNS includes pituitary gland and pineal gland tumors.
Abbreviations: BMI: Body Mass Index; CI: Confidence Interval; CNS: central nervous system; Cum: Cumulative; HR: Hazard Ratio; Lymph: lymphoma.

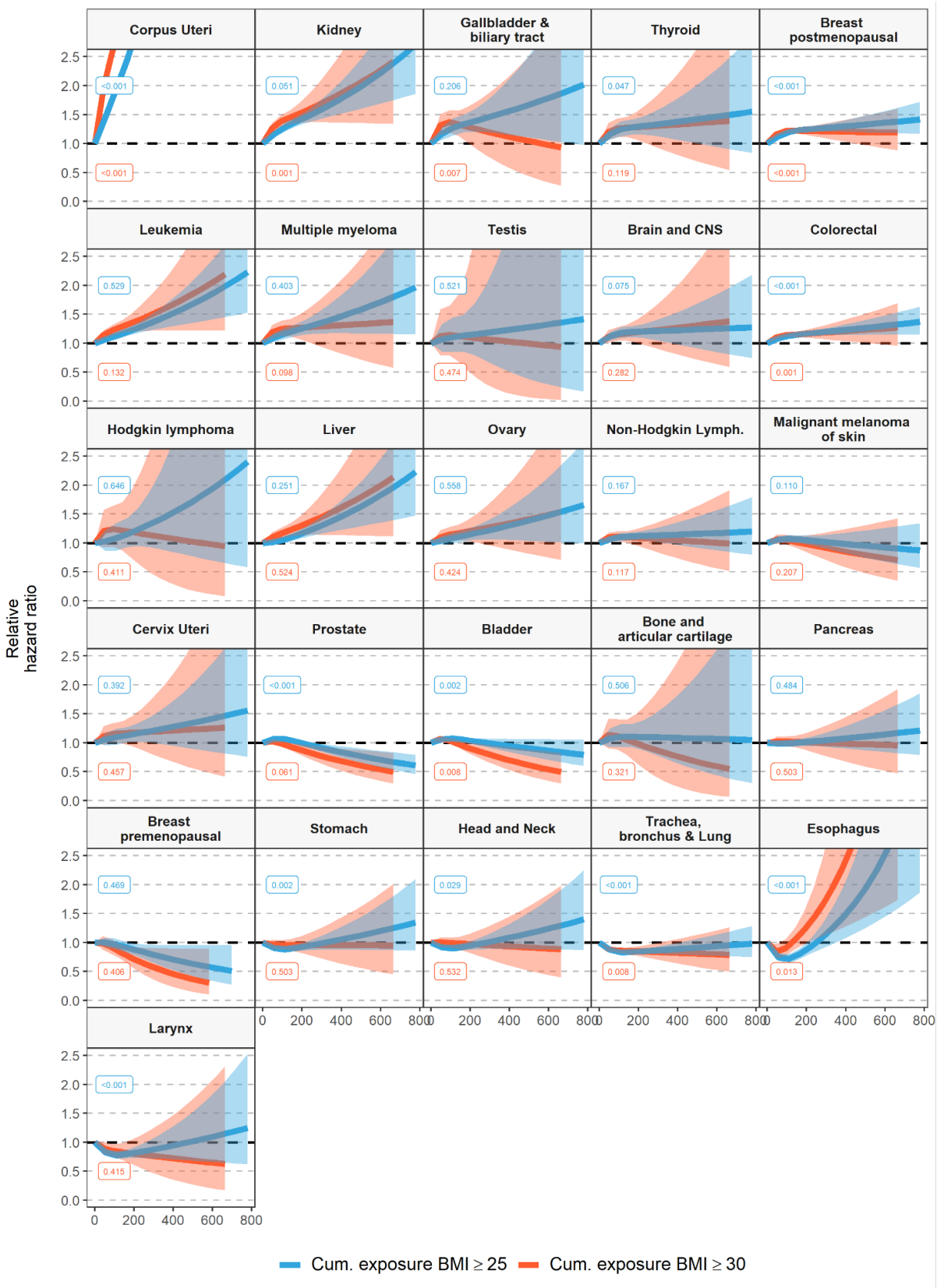
Figure 2. Hazard ratios of 26 cancer types related to duration of BMI ≥25 and ≥30 kg/m² in years, with 95% CIs, allowing for non-linearity



Notes: Models are adjusted for geographic region of nationality, the MEDEA deprivation index, smoking status, and alcohol intake and stratified by age (5-year categories). These graphs were obtained

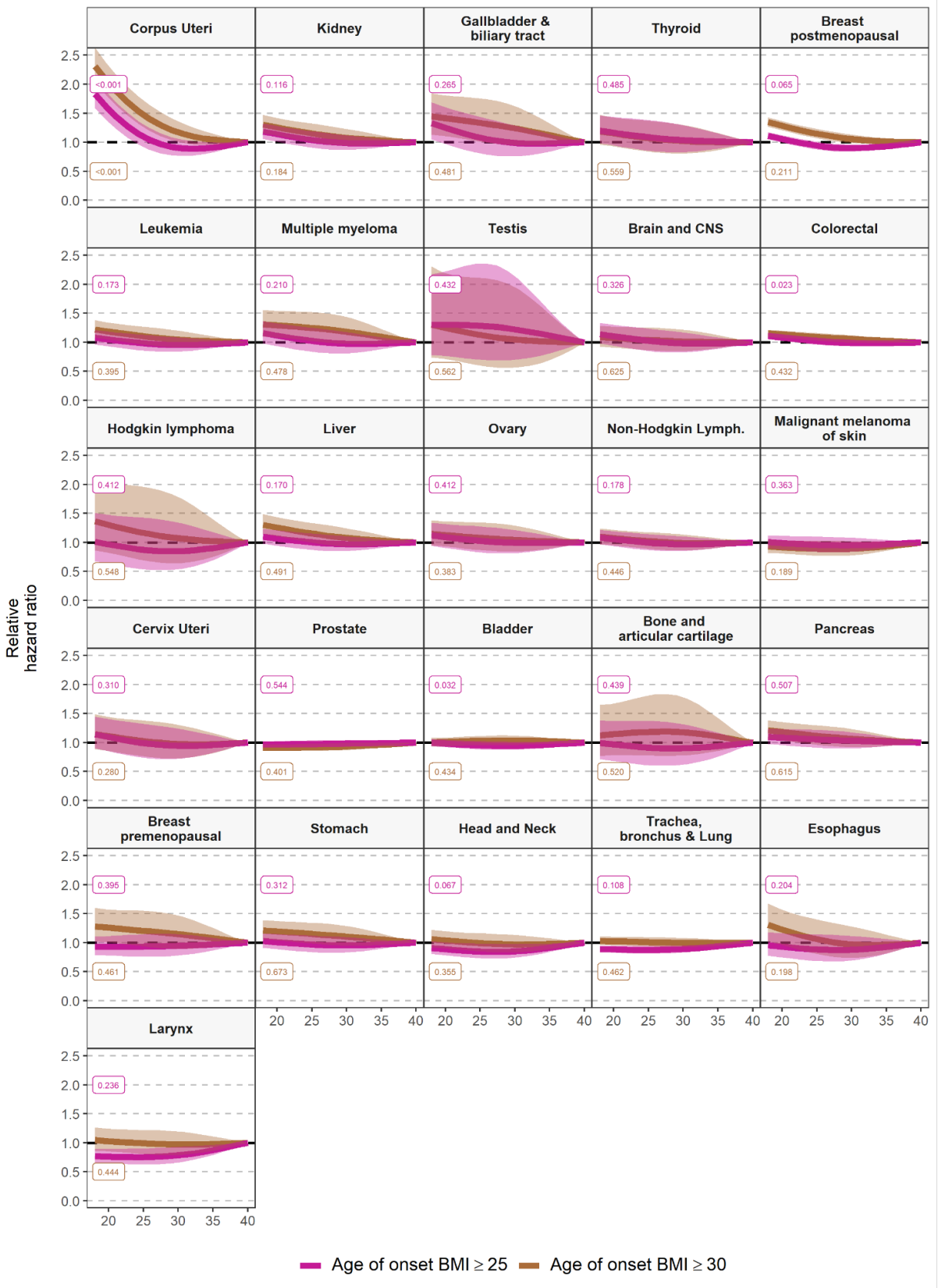
using restricted cubic splines with 3 knots for the exposures of interest with 0 years as the reference point. P-values for nonlinearity were obtained by comparing the model where the exposures were fitted with a nonlinear term against a linear model using a likelihood ratio test. Cancer types are ordered by descending ranking of the HRs for BMI at index date of Figure 1. Models for ovary, cervix, and corpus uteri cancers were only computed in females, for breast pre-menopausal only in pre-menopausal females, for breast post-menopausal only in post-menopausal females, and for prostate and testis only computed in males. Brain and CNS includes pituitary gland and pineal gland tumors.
Abbreviations: BMI: Body Mass Index; CI: Confidence Interval; CNS: central nervous system; Cum: Cumulative; Lymph: lymphoma.

Figure 3. Hazard ratios of 26 cancer types related to cumulative exposure to BMI ≥25 and ≥30 kg/m^2 in cumulative overweight and obese-years, respectively, with 95% CIs, allowing for non-linearity



Notes: Models are adjusted for geographic region of nationality, the MEDEA deprivation index, smoking status, and alcohol intake and stratified by age (5-year categories). These graphs were obtained using restricted cubic splines with 3 knots for the exposures of interest with 0 years as the reference point. P-values for nonlinearity were obtained by comparing the model where the exposures were fitted with a nonlinear term against a linear model using a likelihood ratio test. Cumulative exposure is an exposure considering both degree and duration of overweight/obesity which is obtained by adding the difference between the BMI measurements that were ≥25 (≥30) kg/m^2 and 24.9 (29.9) kg/m^2 for every year lived with a BMI ≥25 and ≥30, respectively. Cancer types are ordered by descending ranking of the HRs for BMI at index date of Figure 1. Models for ovary, cervix, and corpus uteri cancers were only computed in females, for breast pre-menopausal only in pre-menopausal females, for breast post-menopausal only in post-menopausal females, and for prostate and testis only computed in males. Brain and CNS includes pituitary gland and pineal gland tumors.
Abbreviations: BMI: Body Mass Index; CI: Confidence Interval; CNS: central nervous system; Cum: Cumulative; Lymph: lymphoma.

Figure 4. Hazard ratios of 26 cancer types related to age of onset of BMI ≥25 and ≥30 kg/m² (among people who ever had a BMI ≥25 and ≥30 kg/m², respectively) in years



Notes: Models are adjusted for geographic region of nationality, the MEDEA deprivation index, smoking status, and alcohol intake and stratified by age (5-year categories). These graphs were obtained

using restricted cubic splines with 3 knots for the exposures of interest with 40 years as the reference point (thus an HR>1 means a greater risk at younger ages). P-values for nonlinearity were obtained by comparing the model where the exposures were fitted with a nonlinear term against a linear model using a likelihood ratio test. Cancer types are ordered by descending ranking of the HRs for BMI at index date of Figure 1. Models for ovary, cervix, and corpus uteri cancers were only computed in females, for breast pre-menopausal only in pre-menopausal females, for breast post-menopausal only in post-menopausal females, and for prostate and testis only computed in males. Brain and CNS includes pituitary gland and pineal gland tumors.
Abbreviations: BMI: Body mass index; CI: confidence interval; CNS: central nervous system; Lymph: Lymphoma.

9.2. Appendix 2: Study V

Recalde M, Pistillo A, Fontvieille E, Viallon V, Duarte-Salles T, Freisling H.

Body mass index and incident cardiometabolic conditions in relation to cancer risk: a population-based cohort study in Catalonia, Spain

Submitted to a scientific journal

Body mass index and incident cardiometabolic conditions in relation to cancer risk: a population-based cohort study in Catalonia, Spain

Martina Recalde, MPH[1,2,3], Andrea Pistillo, MSc[1], Vivian Viallon, PhD[3], Emma Fontvieille, MPH[3], Talita Duarte-Salles, PhD[1]*, Heinz Freisling, PhD[3]*

*Joint senior authorship

1. Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Gran Via Corts Catalanes, 587 àtic, 08007, Barcelona, Spain
2. Universitat Autònoma de Barcelona, 08193, Bellaterra (Cerdanyola del Vallès), Barcelona, Spain
3. International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert Thomas, 69008 Lyon, France

Corresponding authors:

Heinz Freisling; FreislingH@iarc.fr

International Agency for Research on Cancer (IARC-WHO)

150 Cours Albert Thomas

69008 Lyon, France

Talita Duarte-Salles; tduarte@idiapjgol.org

Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)

Gran Via Corts Catalanes, 587 àtic

08007 Barcelona, Spain

Keywords: adiposity; cancer; type 2 diabetes; hypertension; cardiovascular disease; electronic health record.

Word count: 3977

ABSTRACT

Background: High BMI has been associated with an increased risk of 13 cancer types and cardiometabolic conditions (e.g., hypertension [HTN], type 2 diabetes mellitus [T2DM], and cardiovascular disease [CVD]). The extent to which these conditions modify the BMI-cancer association and the combined effect of adiposity and cardiometabolic conditions remains unknown. We investigated the association between BMI and obesity-related cancer risk among individuals with and without incident HTN, T2DM, and CVD and the joint associations of overweight/obesity (BMI>25kg/m^2) and each cardiometabolic condition with obesity-related cancer risk.

Methods: We conducted a population-based cohort study between 2010 and 2018 with electronic health records from Catalonia, Spain. We included 1,774,904 individuals aged ≥40 years and free of cancer and cardiometabolic conditions at baseline. Our main outcome measures were hazard ratios (HRs) for incident obesity-related cancers and relative excess risk due to interaction (RERI).

Results: After a median follow-up of 8 years, 38,082 individuals developed obesity-related cancers. The positive association between BMI and obesity-related cancer risk was similar among individuals free of cardiometabolic conditions (HR: 1.08, 95%CI: 1.06–1.10) and with incident *HTN* (1.05, 1.01-1.08) or *CVD* (1.08, 0.97-1.21). The association among those with incident *T2DM* was null (0.98, 0.93-1.03). There was a positive additive interaction between overweight/obesity and CVD (RERI: 0.19 [0.09, 0.30]), meaning that the combined association was 0.19 more than the sum of the individual associations (the combined HR was 2.07 instead of 1.87). In contrast, a RERI of -0.24 (-0.28, -0.20) was observed for the combined association between overweight/obesity and *T2DM*.

Conclusions: Public health strategies to reduce overweight can help prevent cancer cases among the general population and individuals with incident HTN or CVD. Our findings further suggest that weight-loss interventions would lead to a greater cancer risk reduction among population subgroups with CVD.

INTRODUCTION

The prevalence of overweight (body mass index, BMI ≥25 & <30kg/m^2) and obesity (BMI ≥30kg/m^2) has rapidly risen over the past decades, reaching more than 1.9 billion and 650 million adults in 2016, respectively.[1] A high BMI (the most common indicator of general adiposity) has been convincingly associated with at least 13 cancer types (labeled as obesity-related cancers).[2] It has also been associated with a higher risk of cardiometabolic conditions such as hypertension (HTN), type 2 diabetes mellitus (T2DM), and cardiovascular diseases (CVD).[3] The prevalence of these conditions has highly increased over the past decades.[4–6] Moreover, HTN and T2DM have been proposed as risk factors for certain cancers.[7] CVD and cancer have been shown to share common molecular pathways,[7,8] and emerging evidence also suggests CVD might be an independent risk factor for cancer.[9]

Incident HTN, T2DM, and CVD may modify cancer processes associated with obesity through shared or additionally triggered (non-shared) biological pathways. Inflammation, oxidative stress, or hormonal processes (sex hormones, insulin and insulin-like growth factor [IGF] signaling, and adipokines) are shared biological pathways with obesity.[7,8] Cardiac proteins excreted into the bloodstream after myocardial infarction are emerging as novel factors promoting tumor growth,[9] and are likely non-shared with obesity.

However, the extent to which these cardiometabolic conditions modify the BMI-cancer association is unclear given that prior studies have mostly focused on healthy or general populations.[10–12] From a public health and clinical perspective, it is important to address whether BMI-cancer associations differ among population groups affected by cardiometabolic conditions, especially given the rise in their prevalence. In addition, prior studies have not investigated the combination of component risk factors (ie, adiposity and incident cardiometabolic conditions) in relation to cancer risk. Evaluation of such interactions is important to guide interventions, clinical decision-making, and health planning. How co-present, or sequential diseases and related risk factors promote negative effects of disease interaction has been referred to as the syndemics model of health.[13] A study conducted with comprehensive patient-level data containing detailed individuals' BMI information and capturing incident HTN, T2DM, CVD, and obesity-related cancer cases from a large and representative population could address these gaps in knowledge.

Our primary aim was to investigate whether incident HTN, T2DM, or CVD modify the association between BMI and the risk of developing obesity-related cancers, using electronic health record (EHR) data from Catalonia, Spain. Our secondary aim was to study the joint associations of overweight/obesity and incident cardiometabolic conditions with obesity-related cancer risk.

METHODS

Study design, setting, and data sources

We conducted a population-based cohort study from January 1st, 2010 to December 31st, 2018. This study was underpinned by prospectively collected primary care records from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org) in Catalonia, Spain. The SIDIAP is a pseudo-anonymized database of EHRs containing data from 5.8 million people living in Catalonia since 2006. This database covers >75% of the population of Catalonia and is representative of the overall population in terms of age, sex, and geographic distribution.[14] The SIDIAP contains data on anthropometric measurements, disease diagnoses (International Classification for Diseases, 10th revision [ICD-10]), and demographic and lifestyle information, among others. Further, SIDIAP can be linked to the Minimum Basic Dataset (CMBD in Spanish), a population-based registry of hospital discharge information including diagnoses and procedures.[15]

Participants

We included all individuals aged ≥40 years registered in SIDIAP on January 1st, 2010 (index date). We excluded participants who had been registered in the database for less than one year (to have sufficient time to capture participants' characteristics before study entry), who had been diagnosed with any cancer type (except other and unspecified malignant neoplasm of skin), HTN, T2DM and/or CVD prior to index date (Figure 1). The follow-up period extended between one year after index date (to minimize the possibility of reverse causality [ie, BMI affected by undiagnosed cancer]) and exit from the database, death, cancer diagnosis (any except other and unspecified malignant neoplasm of skin), or the end of study period (31st December 2018), whichever occurred first.

Outcome assessment

The outcome was a binary indicator of incident diagnoses of a first primary obesity-related cancer which we identified with ICD-10 and ICD-9 codes in the SIDIAP and CMBD hospital discharge databases, respectively (Table S1). Obesity-related cancers comprised cancers of the colorectum; liver; gallbladder and biliary tract; pancreas; post-menopausal breast; corpus uteri; ovary; kidney; brain and central nervous system; thyroid; and multiple myeloma; due to well-established evidence indicating that the presence of excess body fatness increases the risk of these cancers.[2] We omitted esophagus and stomach cancers from this list given that with the available data we could not differentiate esophageal adenocarcinoma (obesity-related) from squamous cell carcinoma or gastric cardia (obesity-related) from non-cardia cancers. Furthermore, the incidence of the non-obesity-related subtypes/subsites of

these cancers is higher in Spain.[16] Cancer diagnoses registered in the SIDIAP including the CMBD have been previously validated.[17]

Exposures assessment

For our primary objective, the exposure was BMI (continuous variable in kg/m^2). BMI values were calculated using the weight (kg) and height (cm) assessed in a standardized manner by general practitioners or nurses.[18] We implemented a multilevel time raster multiple imputation approach to have complete information on BMI for all study participants and to update BMI values every time a participant was diagnosed with a cardiometabolic condition (ie, HTN, T2DM, or CVD) or a combination of these conditions.[19] This method is described in the "Statistical Analyses" section and Appendix 1.

For our secondary objective, the exposure was a composite variable of 16 categories combining binary BMI (< or ≥25kg/^{2}) and cardiometabolic conditions, coded as a time-varying variable with eight categories ("*healthy"*; *HTN*; *T2DM*; *CVD*; *HTN & T2DM*; *HTN & CVD*; *T2DM & CVD*; *HTN, T2DM, & CVD*). All study participants were in the "*healthy"* category at index date, and during follow-up, they could change states to one (*HTN*; *T2DM*; or *CVD*), two (*HTN & T2DM*; *HTN & CVD*; or *T2DM & CVD*), or three (*HTN, T2DM, & CVD*) cardiometabolic conditions (the framework for the variable definition is available in Figure Sl). HTN and T2DM were identified using diagnostic codes recorded in the SIDIAP database (Table S1). CVD was defined as any diagnosis of coronary or cerebrovascular disease which we identified using data from the CMBD hospital discharge and SIDIAP (Table S1).[20]

Covariates of interest

The covariates were cardiometabolic conditions (only for the primary objective), sex, age, geographic region of nationality, socioeconomic status, smoking status, and alcohol intake. We extracted participants' sex (*female*, *male*), age (in years and 5-year categories) at index date (and updated at the moment of diagnosis of a cardiometabolic condition), and geographic region of nationality (*Spanish*, *Global North*, or *Global South*).[21] Socioeconomic status in urban areas was assessed using the *Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales* (MEDEA) deprivation index (calculated at the census tract level and categorized into quintiles by the SIDIAP for anonymization purposes).[22] We included a rural category since the index was unavailable for participants living in rural areas. We also extracted information on smoking status (*never*, *former*, or *current smoker*) and alcohol intake (*no*, *low* or *high risk*) (the closest record to the index date within 5 years before or at the index date was selected).

Statistical analyses

We applied multilevel time raster multiple imputation to BMI at several time points (2006, 2010, 2013, 2016, 2018).[19] We used a linear mixed-effects model with 5 imputations to obtain imputed trajectories of BMI for the study participants (a detailed explanation is available in Appendix 1).[19] BMI at baseline was defined as the corresponding value to the "2010" time point. For participants diagnosed with one or more cardiometabolic conditions, we updated their BMI measurement using the closest prior time point to the date of diagnosis.

We described the number of individuals excluded in each step of the study population definition. We reported the baseline characteristics of the study participants overall, by assessment (or lack) of BMI, and by World Health Organization (WHO) BMI categories: underweight or normal weight (BMI<25kg/m^2), overweight (≥25 and <30kg/m^2), and obesity (≥30kg/m^2).

To investigate if incident HTN, T2DM, or CVD modify the association between BMI and obesity-related cancer risk (primary aim), we fitted Cox proportional hazard models with age as the time metric including BMI, cardiometabolic conditions as a time-varying variable, and an interaction of those with BMI. We estimated two types of models, a model adjusted by sex and stratified by age at index date (5-year categories) (minimally-adjusted) and one further adjusted by geographic region of nationality, MEDEA deprivation index, smoking status, and alcohol intake (fully-adjusted or main model). We used a directed acyclic graph (DAG) to guide our decisions on the control for confounding (Figure S2).[23] We did multiple imputations for the covariates with missing data at baseline (using predictive mean matching, with 5 imputations drawn) (Appendix 1). We accounted for potential non-linearity in the BMI-obesity-related cancer association by fitting models with BMI as a linear term, with a polynomial of degree 2, and with restricted cubic splines (3, 4, or 5 knots).[24] We calculated the Bayesian Information Criterion and we favored the model with the lowest BIC value. We estimated hazard ratios (HRs) and their 95% confidence intervals (CIs) per 5 kg/m^2 increment of BMI. We evaluated the multiplicative interaction between BMI and the variable of cardiometabolic conditions by comparing the difference in log-likelihood of models with and without the interaction term (to facilitate interpretation, we also reported the p-values for interaction between BMI and each cardiometabolic condition). We checked the proportional hazard assumptions by visual inspection of survival curves. We conducted two supplementary analyses to contextualize our findings: stratification of the results by sex and age groups (aged <65 or ≥65 years) to assess potential effect modification and re-running the main model analyzing site-specific cancers (with ≥100 cancer cases) as outcomes. As sensitivity analyses, we re-ran the main model i) without updating the BMI and age of participants, ii) including only individuals with a real BMI assessment at baseline iii) or also during follow-up. We iv) added as an adjustment variable the

number of visits to primary care centers (year before study entry or upon diagnosis of cardiometabolic condition[s]) to account for potentially different health attitudes of the participants.

For our secondary aim, we assessed the relative excess risk due to interaction (RERI) of obesity-related cancers between overweight/obesity (BMI≥25 kg/m^2) and incident cardiometabolic conditions, as recommended in the STROBE statement (joint effects analysis).[25] We fitted a Cox proportional hazard model with age as the time metric including the composite variable adjusted by sex, geographic region of nationality, MEDEA deprivation index, smoking status, alcohol intake, and stratified by age.[26] The RERI was calculated as $RERI_{RR} = RR_{11} - RR_{10} - RR_{01} + 1$, where $_{11}$ denotes being exposed to both factors (eg, overweight/obesity and *HTN*), $_{10}$ to one factor (eg, overweight/obesity), and $_{01}$ to the other one (eg, *HTN*). A RERI of 0 was considered a lack of additive interaction and 95%CIs were calculated as proposed by Hosmer and Lemeshow.[27] To provide a better understanding of this joint association, we also performed a model for the association between the cardiometabolic conditions (8-category variable) and obesity-related cancers separately (supplementary analysis).

We used R version 4.0.1 for all the analyses. We obtained approval from the Clinical Research Ethics Committee of the IDIAPJGol (project code: 20/237-P) to perform this study.

RESULTS

There were 3,097,073 adults aged ≥40 years at index date eligible to enter the study. We excluded 144,772 individuals due to having less than one year of prior clinical history; 1,133,231 to prevalent cancer, HTN, T2DM, or CVD; and 44,166 to less than one year of follow-up (Figure 1).

Of the 1,774,904 study participants, 681,386 (39%) had a BMI assessment at baseline, 589,319 (33%) had at least one BMI assessment during follow-up and 504,199 (28%) did not have any BMI assessment available (Table S2). Age was similarly distributed in the three groups (median age was 53, 51, and 49 years, respectively) and so was BMI among those with an assessment at baseline and only during follow-up (median of 27 kg/m^2 for both). However, those without any BMI measurement had a higher representation of males, non-Spanish, individuals living in the least deprived areas of Catalonia, presenting with fewer comorbidities, and transferred out of SIDIAP than those with a BMI assessment at baseline.

Across all study participants, the median BMI at baseline was 27 (interquartile range [IQR]: 24-30) kg/m^2 (after multiple imputations), the median age was 51 (44-60) years and 53% were females (Table 1). In total, 34% were categorized as living with normal or underweight, 41% with overweight, and

25% with obesity. Compared to those living with obesity, those with normal or underweight, were more frequently females, living in the least deprived areas of Catalonia, and current smokers.

After a median follow-up of 8 years of the 1,774,904 ("*healthy"*) study participants, 38,082 (2.1%) were diagnosed with obesity-related cancers (Figure S1 and Table S3). The number of individuals diagnosed with obesity-related cancers was 6816 (2.3%) among the 296,445 participants diagnosed with incident *HTN* (follow up: 5 years), 1519 (2.3%) among the 65,000 with *TD2M* (5 years), 1140 (2.0%) among the 56,573 with *CVD* (4 years), 1020 (2.6%) among the 39,143 with *HTN & T2DM* (4 years), 497 (1.9%) among the 26,139 with *HTN & CVD* (3 years), 124 (2.3%) among the 6297 with *T2DM & CVD* (3 years), and 114 (1.9%) among the 6069 individuals diagnosed with *HTN, T2DM, & CVD* (2 years).

Association of BMI with obesity-related cancer risk by cardiometabolic conditions

There was multiplicative interaction between BMI and cardiometabolic conditions (p-value from log-likelihood ratio test=0.007) in the association with obesity-related cancers (Figure 2). We did not find evidence of non-linearity between BMI and cancer risk. A BMI increment of 5 kg/m^2 in the main models was positively associated with the risk of obesity-related cancers among "*healthy"* (free of cardiometabolic conditions) individuals (HR: 1.08, 95%CI: 1.06-1.10) and those diagnosed with *HTN* (1.05, 1.01-1.08) (Figure 2). Even though the CIs of the HRs for *CVD* (1.08, 0.97-1.21), *HTN, T2DM, & CVD* (1.05, 0.82-1.33), *HTN & CVD* (1.03, 0.92-1.15), and *T2DM & CVD* (1.02, 0.84-1.24) (in descending order of estimates) overlapped with 1, we did not find evidence of interaction between these and BMI (p-values for interaction >0.05); thus, we cannot conclude that the BMI-cancer association among individuals with these conditions differs from the *"healthy"*. On the contrary, the associations for those with *T2DM* (0.98, 0.93-1.03) and *HTN & T2DM* (1.00, 0.93-1.07) were attenuated with respect to "*healthy*" (p-values for interaction were 0.001 and 0.034, respectively). The effect estimates in the minimally-adjusted models were similar to those of the fully-adjusted models (Figure S3). The results of the supplementary and sensitivity analyses are described in Appendix 2 and reported in (Figures S4A, S4B, S5, and S6).

Joint associations of overweight/obesity and incident cardiometabolic conditions with obesity-related cancer risk

In Table 2 we present the results of the relative excess risk of obesity-related cancers due to additive interaction between overweight/obesity (BMI≥25 kg/m^2) and incident cardiometabolic conditions. The association between overweight/obesity with obesity-related cancer risk in absence of cardiometabolic

conditions (among the *"healthy"*) was 1.11 (95%CI: 1.06-1.16) and was lower than that of the conditions in absence of overweight/obesity. The most pronounced associations between cardiometabolic conditions and obesity-related cancer risk in absence of overweight/obesity were those including *T2DM* (eg, HR: 2.25 [95%CI: 1.82-2.77] for *HTN & T2DM* and 2.18 [1.91-2.50] for *T2DM*) while the least pronounced were those including *HTN* and/or *CVD* (eg, 1.40 [1.29-1.53] for *HTN* and 1.63 [1.35-1.96] for *HTN & CVD*). These results were consistent (but stronger in magnitude) with those of the supplementary analysis investigating the association between cardiometabolic conditions and obesity-related cancers (adjusting for continuous BMI) (Figure S7). There was evidence of additive interaction (RERI ≠0) for five out of seven joint associations of overweight/obesity and cardiometabolic conditions with obesity-related cancer risk. RERIs were >0 for the joint effect of overweight/obesity and *T2DM & CVD* (eg, joint effect: 2.57 [2.08-3.17], RERI: 0.36 [0.32, 0.40]), *HTN, T2DM, & CVD* as well as *CVD* while RERIs were <0 for *HTN & T2DM* (eg, joint effect: 2.12 [1.97-2.27], RERI: -0.24 [-0.28, -0.20]), and *T2DM* (in descending order of RERIs).

DISCUSSION

Main findings

In this large cohort study of 1,774,904 individuals in Catalonia, we found that the positive association between a higher BMI and obesity-related cancer risk was similar among "*healthy*" (free of cardiometabolic conditions) individuals and those with incident diagnoses of *HTN* and/or *CVD*. In contrast, among individuals with incident *T2DM,* this association was null. We also found that single cardiometabolic conditions, and combinations thereof, were each independently and positively associated with obesity-related cancer risk. A striking finding was that the association of overweight/obesity and *CVD* (only or in combination with more conditions) with obesity-related cancer risk was greater than the sum of their separate associations. On the contrary, the observed joint association of overweight/obesity and *T2DM* (only or in combination with other conditions) with obesity-related cancer risk was lower than the sum of their separate associations.

Interpretation and comparison with previous studies

The positive association between BMI and obesity-related cancers among "*healthy"* individuals is in line with well-established evidence.[2,10–12] Three mechanisms by which higher general adiposity can increase cancer risk have been extensively reported in the literature: sex hormonal metabolism, insulin and insulin-like growth factors (IGF) signaling, and adipokine pathways.[28–32] It has also been suggested that other factors, such as cardiometabolic conditions, could be mediators in the association between body fatness and cancer risk.[7,28,29] However, since the "*healthy"* population did not include individuals

with *HTN*, *T2DM*, and *CVD* by definition, our results support the existence of pathways between body fatness and cancer risk independent of these conditions.

Our results revealed that the BMI-obesity-related cancer association still remains present among individuals with an incident diagnosis of *HTN* and/or *CVD*. This observation could be explained by an independent (from *HTN* and *CVD*) pathway between BMI and cancer risk, but also by a weaker (compared to, for example, *T2DM*) association between *HTN* or *CVD* and obesity-related cancers (pathway being blocked by conditioning on *HTN* or *CVD*). The latter hypothesis is sustained by the more "modest" effect of *HTN* and *CVD* on cancer risk among individuals with a BMI<25 kg/m^2, compared to that of *T2DM*. Moreover, the evidence linking *HTN* and *CVD* to cancer risk is yet to be well-established. Studies investigating *HTN* as a risk factor for cancer have mostly focussed on kidney cancer, and although a link has also been suggested for risk of stomach, colorectal, pancreas, postmenopausal breast, brain, and malignant melanoma cancers (cancers also associated with higher adiposity levels) as well as lung cancer, the evidence supporting these associations is not yet well-established.[2,33–35] Similarly, the evidence suggesting that *CVD* might be an independent risk factor for cancer is still at very early stages.[9] While we did not observe an interaction between BMI and *CVD* (nor in combination with T2DM or HTN/T2DM together) on the multiplicative scale, we did find ones on the additive scale (ie, in terms of absolute risks). This suggests that a higher incidence of obesity-related cancers can be expected among population sub-groups affected by both overweight/obesity and these (combination of) cardiometabolic conditions.

The association between BMI and obesity-related cancers among people with *T2DM* was null. This could be explained by shared biological pathways underlying the carcinogenesis of obesity-related cancers between adiposity and *T2DM* as well as by a strong association between *T2DM* and cancer risk (direct association blocked by conditioning on *T2DM*). In meta-analyses, *T2DM* has been positively associated with the risk of cancers of the stomach, colorectal, liver, gallbladder and biliary tract, pancreas, breast, corpus uteri, kidney, bladder, thyroid, non-Hodgkin lymphoma, multiple myeloma, and leukemia.[36–46] Except for bladder cancer, all other associations are concordant with those described for adiposity.[2,35] Our findings (both for the interactions on the multiplicative and additive scale) in combination with what has been reported in the literature reinforce the existence of shared mechanisms between adiposity and *T2DM* in relation to cancer risk. In fact, the pathways that have been proposed to explain the T2DM-cancer associations (hyperinsulinemia, hyperglycemia, IGF signaling, and inflammation) have also been proposed as possible mediators for the BMI-cancer one.[7,47–51]

Strengths and limitations

This study has several strengths. To our knowledge, this is the largest study to date to investigate the association between BMI and risk of obesity-related cancers accounting for incident cardiometabolic

conditions. Individuals included in SIDIAP are representative of the general population living in Catalonia in terms of age, sex, and geographic distribution which favors the external validity of these findings.[14] We implemented an advanced multiple imputation methodology to include the individuals eligible to enter the study (with or without a BMI assessment at baseline) and to update their BMI levels during follow-up, minimizing the possibility of selection bias and exposure misclassification, respectively. While we cannot discard the possibility of outcome misclassification (cancer outcomes were ascertained in primary care and hospital databases), we do not think this was an important limitation. Cancer diagnoses registered in SIDIAP have been validated using population-based cancer registries' data and priorly used for epidemiological research.[12,17,52] We also do not expect that the potential misclassification would have been differential according to the exposure, therefore, this likely did not affect our results. Finally, our findings were robust in sensitivity analyses.

Our findings should be interpreted in light of some limitations. There was a high proportion of individuals who did not have a BMI assessment at baseline. While we used information from any recording in the individuals' health records (eg, also during follow-up) for the time-raster multiple imputations, 28% of the study participants did not have any BMI assessment which likely introduced high variability in their BMI estimations among the imputed datasets. However, when we ran sensitivity analyses only including individuals with a BMI at baseline or with any real BMI assessment, our results were consistent with those of the main analyses. In addition, we did not have enough statistical power to look at specific cancer types as separate outcomes given that the number of at-risk individuals was modest. Most obesity-related cancers were cases of breast postmenopausal (34%) and colorectal (32%) cancers, therefore our results seem to be highly driven by these cancer types. Nevertheless, when we explored associations for other specific cancer types in the secondary analyses (for single cardiometabolic conditions or also for *HTN & CVD* or *HTN & T2DM* for the most frequent cancer types), the associations were consistent (although with wider CIs) with those of the obesity-related cancers combined. We also lacked information on the histological subtypes (esophageal adenocarcinoma) and subsites (gastric cardia) of cancers which prevented us from including two other obesity-related cancers in the outcome definition. This could have biased our results towards the null, however, we do not expect this to have greatly impacted our results as the incidence of the non-obesity-related subtype/subsite of these cancers is higher in Spain.[16] We were also limited in terms of covariate data availability: for socioeconomic status, we only had data on the MEDEA deprivation index, an ecological indicator of deprivation, therefore there could have also been residual confounding (Figure S2).

In this large Southern European study, we found that the positive association between BMI and obesity-related cancers was similar among individuals free of cardiometabolic conditions and those with incident HTN and/or CVD, but was attenuated among individuals with T2DM. Furthermore, individuals

with both overweight/obesity and incident CVD accounted for the highest number of obesity-related cancer cases compared to those with other individual or combined cardiometabolic conditions. Our findings reinforce the need for public health strategies focusing on the reduction of overweight and obesity, which cannot only help prevent cancer cases among individuals from the general population or newly diagnosed with HTN or CVD but also prevent cardiometabolic conditions, such as T2DM, which can also increase obesity-related cancer risk. In case of limited resources, our findings highlight the need for weight loss interventions among individuals newly diagnosed with CVD who have overweight or obesity.

REFERENCES

1. World Health Organization. Overweight and obesity. Published 2021. Accessed August 17, 2021. http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight
2. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine*. 2016;375(8):794-798. doi:10.1056/NEJMsr1606602
3. Wilson PWF, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and Obesity as Determinants of Cardiovascular Risk: The Framingham Experience. *Archives of Internal Medicine*. 2002;162(16):1867-1872. doi:10.1001/archinte.162.16.1867
4. World Health Organization. Hypertension. Published 2018. Accessed September 7, 2021. https://www.who.int/news-room/fact-sheets/detail/hypertension
5. Zhou B, Lu Y, Hajifathalian K, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *The Lancet*. 2016;387(10027):1513-1530. doi:10.1016/S0140-6736(16)00618-8
6. Roth GA, Johnson C, Abajobir A, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. *J Am Coll Cardiol*. 2017;70(1):1-25. doi:https://doi.org/10.1016/j.jacc.2017.04.052
7. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. *Circulation*. 2016;133(11):1104-1114. doi:10.1161/CIRCULATIONAHA.115.020406
8. Masoudkabir F, Sarrafzadegan N, Gotay C, et al. Cardiovascular disease and cancer: Evidence for shared disease pathways and pharmacologic prevention. *Atherosclerosis*. 2017;263:343-351. doi:https://doi.org/10.1016/j.atherosclerosis.2017.06.001
9. Meijers WC, Maglione M, Bakker SJL, et al. Heart Failure Stimulates Tumor Growth by Circulating Factors. *Circulation*. 2018;138(7):678-691. doi:10.1161/CIRCULATIONAHA.117.030816
10. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ*. 2007;335(7630):1134. doi:10.1136/bmj.39367.495995.AE
11. Bhaskaran K, Douglas I, Forbes H, Dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5.24 million UK adults. *The Lancet*. 2014;384(9945):755-765. doi:10.1016/S0140-6736(14)60892-8
12. Recalde M, Davila-Batista V, Díaz Y, et al. Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain. *BMC Medicine*. 2021;19(1):10. doi:10.1186/s12916-020-01877-3

13. Singer M, Bulled N, Ostrach B, Mendenhall E. Syndemics and the biosocial conception of health. *The Lancet*. 2017;389(10072):941-950. doi:https://doi.org/10.1016/S0140-6736(17)30003-X
14. Bolíbar B, Fina Avilés F, Morros R, et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Medicina Clinica*. 2012;138(14):617-621.
15. Generalitat de Catalunya. Conjunt mínim bàsic de dades (CMBD). Published 2017. Accessed March 5, 2019. https://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/index.html#googtrans(ca%7Ces)
16. Arnold M, Ferlay J, van Berge Henegouwen MI, Soerjomataram I. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. *Gut*. 2020;69(9):1564. doi:10.1136/gutjnl-2020-321600
17. Recalde M, Manzano-Salgado C, Díaz Y, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. *Clinical Epidemiology*. 2019;Volume 11:1015-1024. doi:10.2147/CLEP.S225568
18. Lecube A, Monereo S, Rubio MÁ, et al. Prevención, diagnóstico y tratamiento de la obesidad. Posicionamiento de la Sociedad Española para el Estudio de la Obesidad de 2016. *Endocrinología, Diabetes y Nutrición*. 2017;64:15-22. doi:https://doi.org/10.1016/j.endonu.2016.07.002
19. van Buuren S. *Flexible Imputation of Missing Data: Time Raster Imputation*. 2nd Edition. Chapman & Hall/CRC; 2012.
20. ENCePP. Protocol for a multinational, multi-database cohort study to assess adverse cardiovascular and cerebrovascular outcomes and mortality in association with inhaled NVA237 in Europe.
21. Brandt W. *North-South: A Program for Survival*. MIT Press; 1990.
22. Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). *Gaceta Sanitaria*. 2008;22(3):179-187.
23. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. *Epidemiology*. 1999;10(1).
24. Harrell FEJ. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer; 2001.
25. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine*. 2007;4(10):e297-. https://doi.org/10.1371/journal.pmed.0040297
26. Rothman KJ. *Modern Epidemiology*. Little, Brown and Company; 1986.

27. Hallan S, de Mutsert R, Carlsen S, Dekker FW, Aasarød K, Holmen J. Obesity, Smoking, and Physical Inactivity as Risk Factors for CKD: Are Men More Vulnerable? *American Journal of Kidney Diseases*. 2006;47(3):396-405. doi:10.1053/j.ajkd.2005.11.027
28. Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. *Nature Reviews Cancer*. 2015;15(8):484-498. doi:10.1038/nrc3967
29. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Reviews Cancer*. 2004;4(8):579-591. doi:10.1038/nrc1408
30. Khandekar MJ, Cohen P, Spiegelman BM. Molecular mechanisms of cancer development in obesity. *Nature Reviews Cancer*. 2011;11(12):886-895. doi:10.1038/nrc3174
31. Roberts DL, Dive C, Renehan AG. Biological Mechanisms Linking Obesity and Cancer Risk: New Perspectives. *Annual Review of Medicine*. 2010;61(1):301-316. doi:10.1146/annurev.med.080708.082713
32. van Kruijsdijk RCM, van der Wall E, Visseren FLJ. Obesity and Cancer: The Role of Dysfunctional Adipose Tissue. *Cancer Epidemiology, Biomarkers & Prevention*. 2009;18(10):2569-2578. doi:10.1158/1055-9965.EPI-09-0372
33. Radišauskas R, Kuzmickienė I, Milinavičienė E, Everatt R. Hypertension, serum lipids and cancer risk: A review of epidemiological evidence. *Medicina (B Aires)*. 2016;52(2):89-98. doi:https://doi.org/10.1016/j.medici.2016.03.002
34. Seretis A, Cividini S, Markozannes G, et al. Association between blood pressure and risk of cancer development: a systematic review and meta-analysis of observational studies. *Scientific Reports*. 2019;9(1):8565. doi:10.1038/s41598-019-45014-4
35. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*. 2008;371(9612):569-578. doi:https://doi.org/10.1016/S0140-6736(08)60269-X
36. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. *BMJ : British Medical Journal*. 2015;350:g7607. doi:10.1136/bmj.g7607
37. Starup-Linde J, Karlstad O, Eriksen SA, et al. CARING (CAncer Risk and INsulin analoGues): the association of diabetes mellitus and cancer risk with focus on possible determinants - a systematic review and a meta-analysis. *Curr Drug Saf*. 2013;8(5):296-332. doi:10.2174/15748863113086660071
38. Ohkuma T, Peters SAE, Woodward M. Sex differences in the association between diabetes and cancer: a systematic review and meta-analysis of 121 cohorts including 20 million individuals and one million events. *Diabetologia*. 2018;61(10):2140-2154. doi:10.1007/s00125-018-4664-5
39. Noto H, Osame K, Sasazuki T, Noda M. Substantially increased risk of cancer in patients with diabetes mellitus: A systematic review and meta-analysis of epidemiologic evidence in Japan.

Journal of Diabetes and its Complications. 2010;24(5):345-353. doi:https://doi.org/10.1016/j.jdiacomp.2010.06.004

40. Zhenming G, Qiwen B, Junbo Q, Yamin W, Yuming L. Diabetes mellitus and risk of gastric cancer: a systematic review and meta-analysis of observational studies. *European Journal of Gastroenterology & Hepatology*. 2011;23(12). https://journals.lww.com/eurojgh/Fulltext/2011/12000/Diabetes_mellitus_and_risk_of_gastric_cancer__a.6.aspx
41. Jiang Y, Ben Q, Shen H, Lu W, Zhang Y, Zhu J. Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies. *European Journal of Epidemiology*. 2011;26(11):863-876. doi:10.1007/s10654-011-9617-y
42. Wang C, Wang X, Gong G, et al. Increased risk of hepatocellular carcinoma in patients with diabetes mellitus: A systematic review and meta-analysis of cohort studies. *International Journal of Cancer*. 2012;130(7):1639-1648. doi:https://doi.org/10.1002/ijc.26165
43. Song S, Wang B, Zhang X, et al. Long-Term Diabetes Mellitus Is Associated with an Increased Risk of Pancreatic Cancer: A Meta-Analysis. *PLOS ONE*. 2015;10(7):e0134321-. https://doi.org/10.1371/journal.pone.0134321
44. Friberg E, Orsini N, Mantzoros CS, Wolk A. Diabetes mellitus and risk of endometrial cancer: a meta-analysis. *Diabetologia*. 2007;50(7):1365-1374. doi:10.1007/s00125-007-0681-5
45. Bao C, Yang X, Xu W, et al. Diabetes mellitus and incidence and mortality of kidney cancer: A meta-analysis. *Journal of Diabetes and its Complications*. 2013;27(4):357-364. doi:https://doi.org/10.1016/j.jdiacomp.2013.01.004
46. Castillo JJ, Mull N, Reagan JL, Nemr S, Mitri J. Increased incidence of non-Hodgkin lymphoma, leukemia, and myeloma in patients with diabetes mellitus type 2: a meta-analysis of observational studies. *Blood*. 2012;119(21):4845-4850. doi:10.1182/blood-2011-06-362830
47. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. *The Lancet*. 2004;363(9418):1346-1353. doi:10.1016/S0140-6736(04)16044-3
48. Chen W, Wang S, Tian T, et al. Phenotypes and genotypes of insulin-like growth factor 1, IGF-binding protein-3 and cancer risk: evidence from 96 studies. *European Journal of Human Genetics*. 2009;17(12):1668-1675. doi:10.1038/ejhg.2009.86
49. Kaaks R, Lukanova A, Kurzer MS. Obesity, Endogenous Hormones, and Endometrial Cancer Risk: A Synthetic Review. *Cancer Epidemiology, Biomarkers & Prevention*. 2002;11(12):1531-1543.
50. Murphy E, Kelly DP. Estrogen Signaling and Cardiovascular Disease. *Circulation Research*. 2011;109(6):687-696. doi:10.1161/CIRCRESAHA.110.236687
51. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. *The Journal of Clinical Investigation*. 2006;116(7):1793-1801. doi:10.1172/JCI29069

52. Roel E, Pistillo A, Recalde M, et al. Cancer and the risk of coronavirus disease 2019 diagnosis, hospitalisation and death: A population-based multistate cohort study including 4 618 377 adults in Catalonia, Spain. *International Journal of Cancer*. 2022;150(5):782-794. doi:https://doi.org/10.1002/ijc.33846

Acknowledgments

We would like to thank all healthcare professionals of Catalonia who daily register information in the populations' electronic health records. We would also like to thank the Real World Epidemiology (RW-Epi) group from *Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina* (IDIAP Jordi Gol) and the Nutrition, Cancer, and Multimorbidity (NCM) group from the International Agency for Research on Cancer (IARC-WHO) for their insightful contributions during the development of this study. This study was carried out as part of the Doctoral Program in Biomedical Research Methodology and Public Health at the Autonomous University of Barcelona.

Role of the funding source

Funding [grant number: IIG_2019_1978] was obtained from World Cancer Research Fund (UK), as part of the World Cancer Research Fund International grant program. MR was also funded by Wereld Kanker Onderzoek Fonds (WKOF, grant number: 2017/1630), as part of the international grants from the World Cancer Research Fund. TDS acknowledges receiving financial support from the Instituto de Salud Carlos III (ISCIII; Miguel Servet 2021: CP21/00023). The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer and World Health Organization, the authors alone are responsible for the views expressed in this Article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer and World Health Organization.

Declaration of interests

The authors declare no conflicts of interest.

Data sharing

In accordance with current European and national law, the data used in this study is only available for the researchers participating in this study. Thus, we are not allowed to distribute or make publicly available the data to other parties. However, researchers from public institutions can request data from SIDIAP if they comply with certain requirements. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudesen/application-proccedure) or by contacting Anna Moleras (amoleras@idiapjgol.org).

Author contributions

MR performed the literature review. MR and AP did the data management and led the data analysis with contributions from all authors. MR wrote the first draft with insightful contributions from HF and TDS. All authors were involved in the study conception and design, interpretation of the results, manuscript preparation, and approved the final version of the manuscript.

Ethical approval

We obtained approval from the Clinical Research Ethics Committee of the IDIAPJGol (project code: 20/237-P) to perform this study.

TABLES AND FIGURES

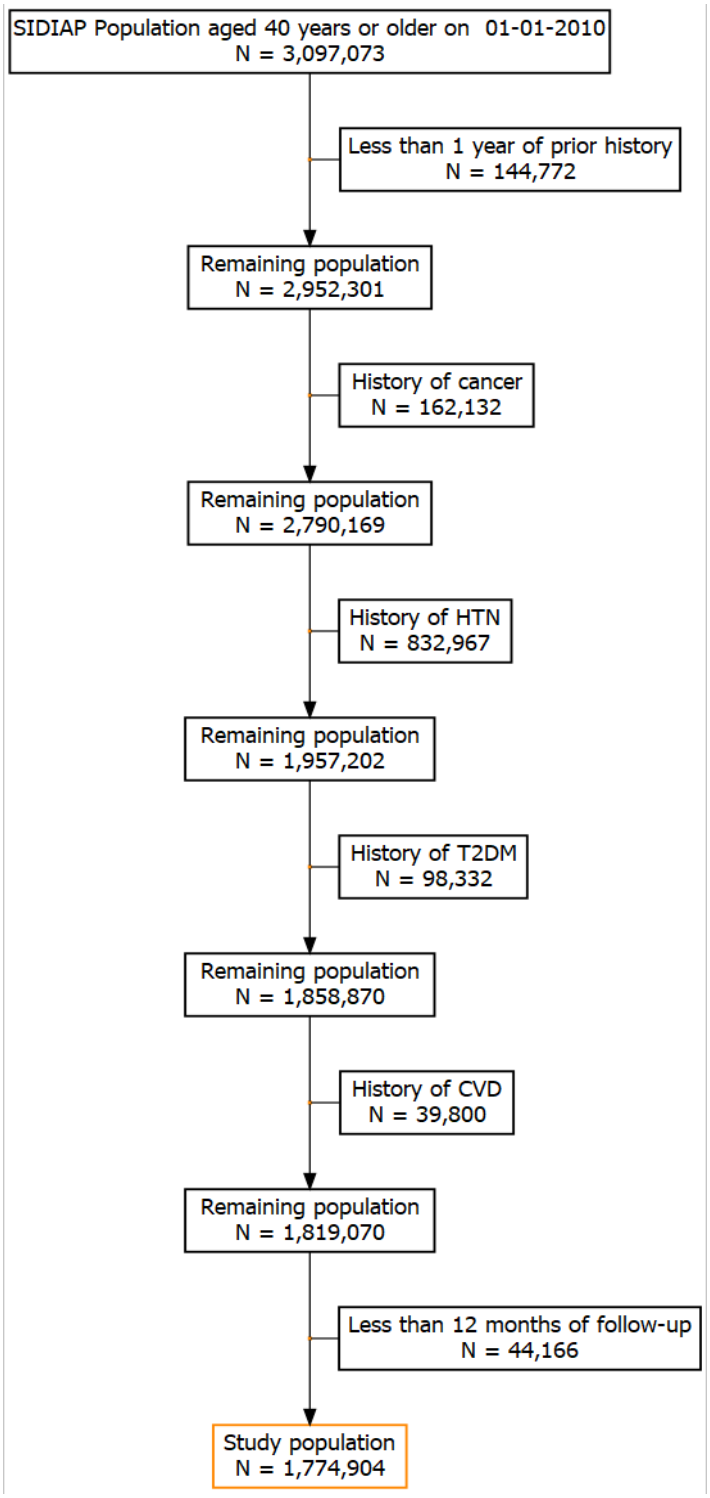
Figure 1. Flowchart with the inclusion and exclusion criteria of the study participants

Table 1. Baseline characteristics of the study participants by body mass index categories, after multiple imputations

Figure 2. Association between body mass index and the risk of obesity-related cancers by ascertainment of incident cardiometabolic conditions, with 95% CIs

Table 2. Relative excess risk of obesity-related cancers due to interaction between overweight/obesity (BMI≥25 kg/m^2) and incident cardiometabolic conditions 198

Figure 1. Flowchart with the inclusion and exclusion criteria of the study participants



Notes: History of cancer considers any type of cancer (C00-C97) except other and unspecified malignant neoplasm of skin (C44). Causes of end-of-follow-up include transferral out of SIDIAP, cancer diagnosis, death, or end-of-study period. Individuals with less than 12 months of follow-up were excluded because the follow-up of the participants started 1 year after study entry to avoid potential reverse causality (eg, BMI affected by undiagnosed cancer).

Abbreviations: CVD: Cardiovascular disease; HTN: Hypertension; SIDIAP: Information System for Research in Primary Care; T2DM: Type 2 diabetes mellitus.

Table 1. Baseline characteristics of the study participants by body mass index categories, after multiple imputations

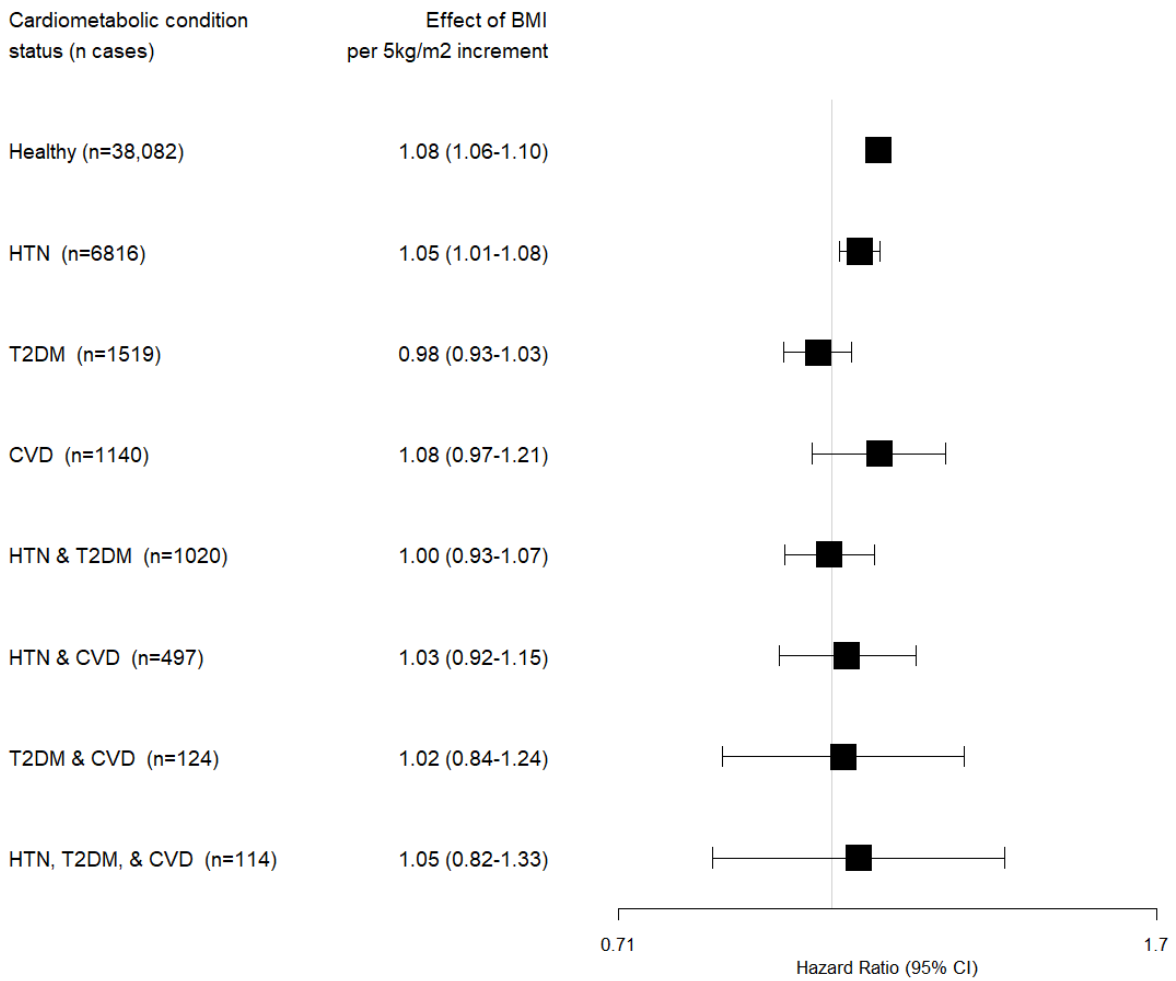
	Overall N (%)	By WHO categories of BMI[1] N (%)		
		Normal or underweight	Overweight	Obesity
	1,774,904 (100.0)	606,249 (34.0)	722,839 (41.0)	445,816 (25.0)
Follow-up time in years, median (IQR)	8.0 (8.0, 8.0)	8.0 (8.0, 8.0)	8.0 (8.0, 8.0)	8.0 (8.0, 8.0)
N of visits to primary care centers, median (IQR)	3.0 (0.0, 7.0)	2.0 (0.0, 6.0)	3.0 (0.0, 7.0)	3.0 (0.0, 8.0)
BMI in kg/m^2, median (IQR)[2]	27.0 (23.9, 30.0)	23.0 (20.9, 23.9)	27.0 (26.2, 28.6)	32.0 (31.1, 34.6)
Age in years, median (IQR)	51.0 (44.0, 60.0)	50.0 (44.0, 59.0)	51.0 (45.0, 61.0)	51.0 (45.0, 61.0)
Female sex, n (%)	931,239 (52.5)	354,019 (58.4)	351,476 (48.6)	225,744 (50.6)
Nationality				
Spanish	1,632,639 (92.0)	561,797 (92.7)	665,784 (92.1)	405,057 (90.9)
Global North	48,735 (2.7)	17,109 (2.8)	18,982 (2.6)	12,643 (2.8)
Global South	93,530 (5.3)	27,342 (4.5)	38,072 (5.3)	28,116 (6.3)
MEDEA deprivation index, n (%)[2]				
Quintile 1 (least deprived)	334,723 (18.9)	131,872 (21.8)	133,284 (18.4)	69,567 (15.6)
Quintile 2	294,506 (16.6)	102,920 (17.0)	120,508 (16.7)	71,078 (15.9)
Quintile 3	278,367 (15.7)	91,423 (15.1)	114,662 (15.9)	72,281 (16.2)
Quintile 4	263,856 (14.9)	82,432 (13.6)	108,559 (15.0)	72,865 (16.3)
Quintile 5 (most deprived)	236,249 (13.3)	71,723 (11.8)	95,716 (13.2)	68,811 (15.4)
Rural	367,203 (20.7)	125,880 (20.8)	150,109 (20.8)	91,215 (20.5)
Smoking status, n (%)[2]				
Never smoker	1,090,923 (61.5)	362,602 (59.8)	448,072 (62.0)	280,249 (62.9)
Former smoker	205,295 (11.6)	67,046 (11.1)	84,940 (11.8)	53,308 (12.0)
Current smoker	478,686 (27.0)	176,601 (29.1)	189,827 (26.3)	112,259 (25.2)
Alcohol intake, n (%)[2]				
No risk	1,099,308 (61.9)	383,154 (63.2)	438,409 (60.7)	277,745 (62.3)
Low risk	602,673 (34.0)	199,364 (32.9)	254,167 (35.2)	149,142 (33.5)
High risk	72,923 (4.1)	23,732 (3.9)	30,262 (4.2)	18,929 (4.2)
Charlson comorbidity index, n (%)				

0	1,522,931 (85.8)	525,679 (86.7)	620,395 (85.8)	376,857 (84.5)
1	210,730 (11.9)	67,111 (11.1)	85,846 (11.9)	57,773 (13.0)
2	31,995 (1.8)	10,449 (1.7)	12,929 (1.8)	8617 (1.9)
≥3	9,248 (0.5)	3010 (0.5)	3669 (0.5)	2570 (0.6)
Cause of exit from the study, n (%)				
End of study	1,373,650 (77.4)	467,137 (77.1)	562,394 (77.8)	344,119 (77.2)
Transferred out of the SIDIAP	219,024 (12.3)	77,669 (12.8)	868,880 (12.0)	54,474 (12.2)
Death	78,456 (4.4)	28,055 (4.6)	30,574 (4.2)	19,827 (4.4)
Obesity related cancers	49,312 (2.8)	15,318 (2.5)	20,016 (2.8)	13,978 (3.1)
Non-obesity related cancers	54,462 (3.1)	18,069 (3.0)	22,974 (3.2)	13,419 (3.0)

Notes: 1) This categorization was done in the 5 datasets with the multiple imputations. For visualization purposes and in order for the categorical variables to add up to 1,774,904 we divided the n for the categorical variables by 5. 2) The statistics of BMI, the MEDEA deprivation index, smoking status, and alcohol intake were calculated using the multiple imputation approach, with 5 data sets created. For visualization purposes, we divided the n for the categorical variables by 5. BMI categories: underweight or normal weight [BMI <25 kg/m^2], overweight [BMI ≥25 and <30 kg/m^2], and obesity [BMI ≥30 kg/m^2]). Non-obesity related cancers do not include non-melanoma skin cancer.

Abbreviations: BMI: Body Mass Index; IQR: Interquartile range; MEDEA: "Mortalidad en áreas pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales"; SIDIAP: Information System for Research in Primary Care; WHO: World Health Organization.

Figure 2. Association between body mass index and the risk of obesity-related cancers by ascertainment of incident cardiometabolic conditions, with 95% CIs



Notes: The model included BMI as a continuous variable with an interaction term with the time-varying "cardiometabolic conditions" variable and was adjusted by sex, the geographic region of nationality, the MEDEA deprivation index, smoking status, alcohol intake, and stratified by age (5-year categories). We evaluated the interaction between BMI and the variable of cardiometabolic conditions by comparing the difference in log-likelihood of models with and without the interaction term (p=0.007). The p-values for the interaction between BMI and each cardiometabolic condition (as extracted directly from the model output) were: 0.067 (HTN), 0.001 (T2DM), 0.980 (CVD), 0.034 (HTN & T2DM), 0.373 (HTN & CVD), 0.577 (T2DM & CVD), 0.790 (HTN, T2DM, & CVD).

Abbreviations: BMI: body mass index; CI: confidence interval; CVD: Cardiovascular disease; HTN: Hypertension; KG: kilograms; M: meters; T2DM: Type 2 diabetes mellitus.

Table 2. Relative excess risk of obesity-related cancers due to interaction between overweight/obesity (BMI≥25 kg/^{2}) and incident cardiometabolic conditions

	HTN		T2DM		CVD		HTN & T2DM		HTN & CVD		T2DM & CVD		HTN, T2DM, & CVD	
	n at risk (n cases)	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**	**n at risk (n cases)**	**HR (95% CI)**
BMI<25 kg/m^2, *"healthy"*	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)	606,249 (12,861)	1 (ref)
BMI ≥25 kg/m^2, *"healthy"*	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)	1,168,655 (25,221)	1.11 (1.06-1.16)
BMI<25 kg/m^2, with condition	67,488 (1544)	1.40 (1.29-1.53)	10,345 (262)	2.18 (1.91-2.50)	17,056 (310)	1.76 (1.44-2.16)	4638 (132)	2.25 (1.82-2.77)	6590 (116)	1.63 (1.35-1.96)	1206 (20)	2.09 (1.18-3.71)	1023 (17)	1.76 (0.87-3.54)
BMI ≥25 kg/m^2, with condition (joint effect)	228,957 (5272)	1.49 (1.44-1.55)	54,655 (1257)	2.07 (1.93-2.22)	39,517 (830)	2.07 (1.84-2.32)	34,505 (888)	2.12 (1.97-2.27)	19,549 (381)	1.78 (1.60-1.98)	5091 (104)	2.57 (2.08-3.17)	5046 (97)	2.10 (1.70-2.59)
RERI		-0.02 (-0.06, 0.02)		-0.22 (-0.29, -0.16)		0.19 (0.09, 0.30)		-0.24 (-0.28, -0.20)		0.04 (-0.07, 0.15)		0.36 (0.32, 0.40)		0.23 (0.02, 0.44)

Notes: The model was adjusted by sex, the geographic region of nationality, the MEDEA deprivation index, smoking status, alcohol intake, and stratified by age (5-year categories). A RERI of 0 denotes lack of additive interaction.

Abbreviations: BMI: Body mass index; CI, confidence interval; CVD: Cardiovascular disease; HR: hazard ratio; HTN: Hypertension; RERI: relative excess risk due to interaction; T2DM: Type 2 diabetes mellitus.

Dr. Talita Duarte-Salles, Senior Epidemiologist at Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol) and,

Dr. Heinz Freisling, Scientist at International Agency for Research on Cancer (IARC-WHO),

Certify,

That the doctoral Thesis entitled "Adiposity, cardiometabolic conditions, and cancer risk: evidence from electronic health records in Catalonia" presented by Martina Recalde, and co-directed by them comply with the merits to be presented and defended in front of the corresponding Reviewing committee to opt for the title of Doctor in Methodology of Biomedical Research Methodology and Public Health.

In witness thereof, the following document was signed in Barcelona, March 2022.

Dr. Talita Duarte-Salles Dr. Heinz Freisling

Adiposity, cardiometabolic conditions, and cancer risk: evidence from electronic health records in Catalonia

PhD Thesis

Doctoral program in Methodology of Biomedical Research and Public Health

Author:
Martina Recalde
martirecalde@gmail.com

Cover figure:
Andrea Pistillo

UAB, Universitat Autònoma de Barcelona
Department of Paediatrics, Obstetrics & Gynaecology, and Preventive Medicine and Public Health
1st Edition Barcelona, March 2022