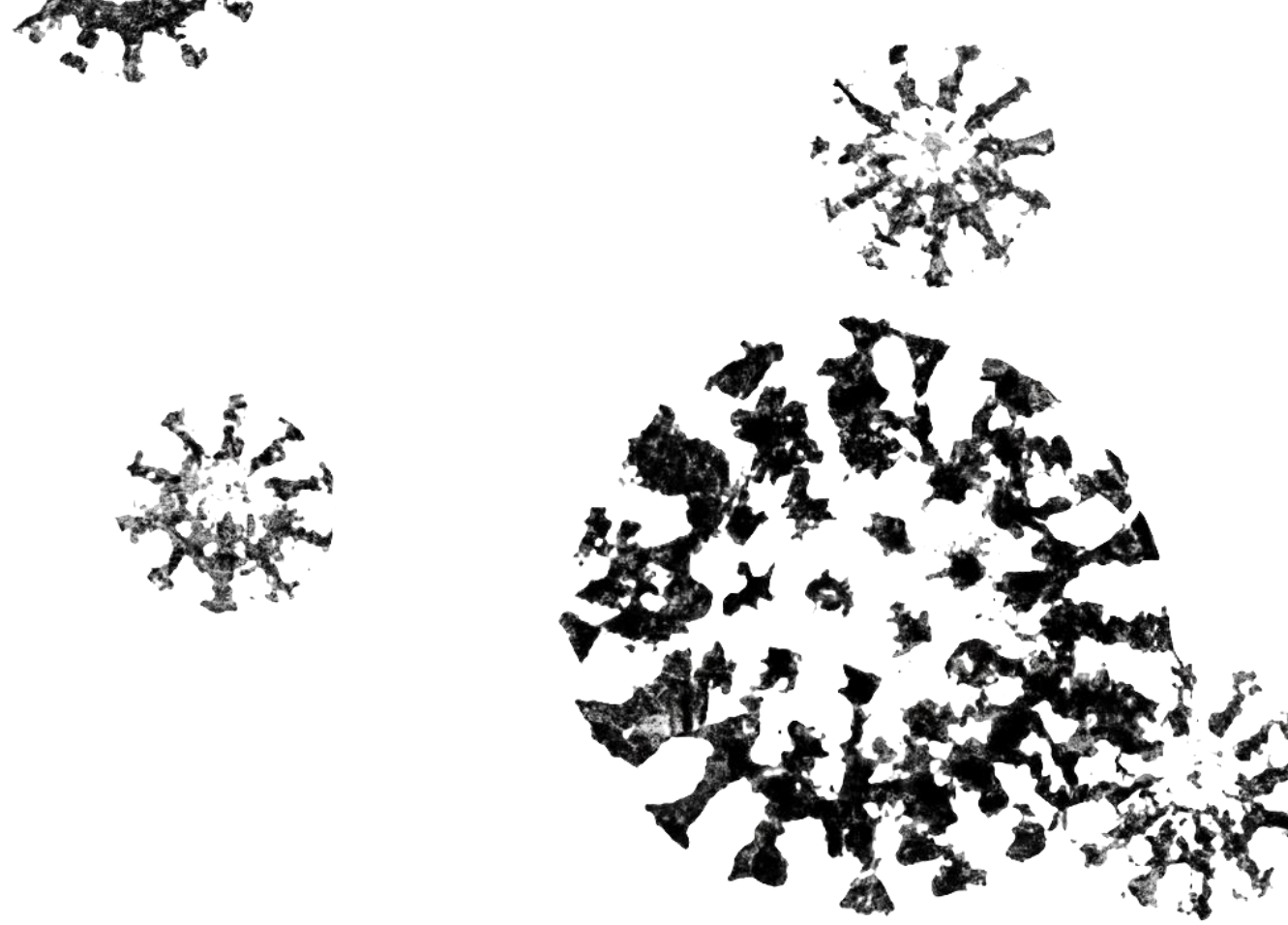
# Timely insights into the COVID-19 pandemic using real-world data

**Elena Roel**



# Timely insights into the COVID-19 pandemic using real-world data

Elena Roel

Doctoral program in Methodology of Biomedical Research and Public Health, Department of Paediatrics, Obstetrics, and Gynaecology, and Preventive Medicine and Public Health, Universitat Autònoma de Barcelona, 2023.

Doctoral Thesis

# Timely insights into the COVID-19 pandemic using real-world data

By Elena Roel

## 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 Daniel Prieto Alhambra
Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford

## Thesis tutor

Dr Ignasi Bolíbar Ribas
Department of Clinical Epidemiology and Public Health, Hospital de la Santa Creu i Sant Pau

"Disease only treats humans equally
when our social orders treat humans equally."

John Green

# Preface

This Thesis has been developed at the Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol) in Barcelona, Spain, and at the Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, in Oxford, United Kingdom; under the co-supervision of Dr Talita Duarte Salles (IDIAPJGol) and Dr Daniel Prieto Alhambra (University of Oxford). This Thesis started three months after the start of the coronavirus disease 2019 (COVID-19) pandemic and was written between June 2020 and January 2023. It consists of a compilation of five original published studies and complies with the procedures and regulations of the Doctoral program in Methodology of Biomedical Research and Public Health from the Department of Paediatrics, Obstetrics, Gynaecology, and Preventive Medicine and Public Health from Universitat Autònoma de Barcelona, Spain. During the conduction of this Thesis, Elena Roel was supported by Instituto de Salud Carlos III (ISCIII; Río Hortega 2020, CM20/00174).

This Thesis aimed to provide timely answers to some of the research questions that arose during the COVID-19 pandemic using real-world data. Briefly, we aimed to describe the baseline characteristics and the occurrence of outcomes among people with underlying conditions of interest with COVID-19 and to investigate factors associated with COVID-19 infection and COVID-19 vaccine uptake and safety. The Thesis begins with an introductory chapter that contextualises the studies carried out, followed by an overview of the objectives and methods of the studies included. Then, the five articles of the thesis are enclosed.

**Studies I** and **II** were multinational cohort studies, whereas **Studies III, IV,** and **V** were population-based cohort studies underpinned by electronic health records from Catalonia, Spain. In **Study I**, we described the characteristics and adverse

outcomes among individuals diagnosed and hospitalised with COVID-19 living with and without obesity. In **Study II**, we characterised the features and health outcomes observed amongst patients with a history of cancer diagnosed and hospitalised with COVID-19 and compared the latter to patients with a history of cancer hospitalised with seasonal influenza in 2017-2018. In **Study III**, we analysed the associations between cancer and COVID-19-related outcomes using a multi-state framework design. In **Study IV**, we studied the relation between socioeconomic deprivation and COVID-19 vaccination and infection. In **Study V**, we estimated incidence rates of thromboembolic events following COVID-19 infection and vaccination and compared these to background population rates. Finally, the Thesis provides a general discussion of the study results, including a summary of the main findings and contributions to the literature, methodological considerations, recommendations for future research, and implications for public health.

# Acknowledgments

Empecé este camino tal y como empiezan todos los caminos, en gran parte debido a la casualidad. Siendo residente de Medicina Preventiva y Salud Pública, a finales del 2019, hice una rotación en el IDIAPJGol, donde tuve la suerte de que Talita me propusiera hacer la tesis en su grupo. Decidimos pedir una beca de investigación, a pesar de que no estaba muy convencida de querer hacer el doctorado. Después, llegó la pandemia. El final de mi residencia estuvo marcado por el retorno al hospital y horas de estrés en una carrera contrarreloj que habíamos perdido de antemano. Acabé en mayo del 2020, exhausta, y con ganas de no saber nada del COVID-19 en mucho tiempo. Una semana después, empezaba esta tesis. En esos momentos había muchas oportunidades laborales para salubristas, pero todas estaban relacionadas con el COVID-19. Incluso la Tesis, si quería. En una reunión con Talita, me dio la opción de cambiar de tema y hacerla sobre el COVID-19 y algo acabó de convencerme. Me pareció que en el contexto de la pandemia la Tesis era la mejor oportunidad para seguir formándome como epidemióloga y adquirir herramientas para un (hipotético) futuro dentro de la Salud Pública. Y también me picó el gusanillo por esos "*real-world data*", que parecían el sueño de cualquier epidemiólogo. Ha sido un camino (muy) corto y también (muy) largo. En estos dos años y medio he aprendido y disfrutado mucho a nivel profesional y personal, y me he sentido crecer en muchos aspectos. Empecé este camino con dudas, pero lo termino sin ellas, sabiendo que elegí bien y que he tenido suerte. La suerte de hacer el doctorado en el momento y lugar adecuados, y, sobre todo, con las personas adecuadas. A todas ellas, les quiero dar las gracias.

Muchas gracias Talita por darme esta oportunidad. Ha sido un privilegio tenerte como directora y aprender de ti como epidemióloga y como supervisora. Haces que las cosas sean fáciles, y eso es particularmente difícil. Gracias por tu honestidad y transparencia en todo momento, sobre todo por compartir con

todos nosotros las dificultades del grupo y por decir abiertamente (y luchar por) las cosas que consideras injustas. Gracias también por tu flexibilidad, por darme la libertad para tomar mis propias decisiones como investigadora y para desarrollar proyectos más afines con mis intereses. Siempre he sentido que tenía mucha autonomía, y que las oportunidades para aprender y llevar a cabo proyectos interesantes eran infinitas dentro del grupo, ojalá hubiese podido hacer más (¡cuántas cosas quedan!). Durante tu ausencia, fui aún más consciente del increíble trabajo que haces como líder del grupo, todos te echamos mucho de menos. Me alegro mucho de los nuevos caminos que se han abierto ante ti en estos dos años, tanto personales como profesionales, seguro que serán maravillosos.

Dani, moltes gràcies per acceptar ser el meu codirector i per ensenyar-me tantes coses com a investigador. Mai oblidaré els moments on, mentre discutíem resultats preliminars en reunions de grup, feies un comentari que canviava completament la perspectiva del problema i de cop tot cobrava sentit. Tampoc oblidaré les reunions on em preguntaves quins eren els meus objectius i què volia fer en el futur. Encara no sé la resposta (ni la sabré mai!), però fer-se aquestes preguntes forma part del camí. Gràcies també per l'oportunitat d'anar a Oxford, ha estat una experiència apassionant. És un luxe poder treballar amb un grup realment multi-disciplinar (i multi-cultural!) com el teu, he après moltíssim i sé que encara hauria après molt més si hagués pogut estar-hi més temps.

Gràcies també a tothom de l'IDIAPJGol, especialment a la gent del meu grup. Martina, ¡qué suerte coincidir contigo en el IDIAP! Muchas gracias por tu ayuda, sobre todo en los primeros meses. No sabes cuanto me ayudaron a situarme esos cafés en remoto las primeras semanas. Escribir juntas el primer artículo de esta Tesis fue una experiencia única. Trabajar tan bien con otra persona es algo prácticamente imposible, aprendí muchísimo contigo y de ti. Gracias también por tus consejos y documentos

de (auto) ayuda para poder navegar por los mares de la burocracia doctoral, has abierto el camino de todos los doctorandos del grupo. Moltes gràcies també al grup de *predocs*, sobretot pels nostres "*Coffee breaks – Nothing to lose*" per pair les penes i les nostres converses per Teams plenes de gifs memorables. Hem sabut riure'ns junts de les dificultats i per mi això no té preu. Moltes gràcies, Berta per les nostres sessions de teràpia predoctoral. Totes dues hem viscut l'absència temporal de la Talita durant la tesi, i també sabem el que implica fer una tesi sobre un tema com la COVID-19, amb les seves coses bones i dolentes. Gracias Andrea por revisar mi código en más de una ocasión, no sabes cuanta tranquilidad me has dado. Gracias también por tu paciencia, como por ejemplo cuando te pedíamos constantemente que cambiaras los colores de las figuras del primer artículo de esta tesis. Gràcies Alícia pels teus consells i per ajudar a tirar el grup endavant quan la Talita no hi era. Matthew, gràcies per alegrar-nos amb els teus dolços i la teva alegria, i per ser el millor referent de com es parla el català. Carlen, gracias a ti también por tu trabajo durante la ausencia de Talita, me habría gustado vernos más en persona. Me alegré mucho de compartir tiempo juntas en Copenhague. Gracias Sergio, por tu paciencia cada vez que te preguntaba cosas de OMOP. Constanza, este camino empezó gracias a ti, cuando coincidimos en Madrid y hablamos de una posible rotación en el IDIAP. Gracias por el tiempo compartido, por tu energía vital y por tus (siempre interesantes) reflexiones durante nuestras comidas en la terraza. Moltes, moltíssimes gràcies també al Rafael Ramos, que va acceptar demanar la Rio Hortega amb mi i que ha fet que aquesta tesi sigui possible. Gràcies també Maria(s), Laura, Anna(s), Diana, Lucia, Carles, Andrea i moltes més persones amb qui he compartit una part del meu camí a l'IDIAP.

I also want to thank all the people at Oxford. Thank you, Ed, for all your help throughout these years. I started my PhD without any knowledge of R, and now at least I know how to understand and adapt your code :). Thank you also for your patience when

answering my silly questions about the database, OMOP, R packages and errors, etc., for always providing interesting (and kind) feedback to my manuscripts, and for writing with me at record speed the paper on thrombosis. This Thesis would have been very different without you. Gràcies Martí per la teva ajuda durant la meva estada a Oxford, per tenir la paciència d'ajudar-me a resoldre els meus errors en el codi, per voler convidar-me a cerveses i per aportar sempre alegria al grup (sovint des de la queixa). Thank you, Annika, for inviting me to collaborate on your Long Covid project, which I've found fascinating! It has been an amazing experience and I've really enjoyed working with you on this. Gràcies també a la resta del "team català", Marta, Núria, Albert i Kim, que m'heu fet sentir com a casa. Many thanks also to Maite, Xintong, Mhaki, Hez, Cheryl, Theresa, Frank, Mimi, Mike, Sara, Ali, Paloma, Nicola, Kristin, James, Antonella, Arani, Danielle, Hester, Francesca, Leena, Trishna and Wai; such a long list and such an amazing group! Working with you has been both a pleasure and a privilege.

Gracias a mis padres, sin los cuales nunca habría llegado hasta aquí. Gracias por apoyarme siempre, incluso en momentos en los que tomé decisiones que no compartíais. Habéis hecho todo lo posible por darme una buena educación y me habéis inculcado el afán por el conocimiento y, sobre todo, por la justicia social. Nunca os estaré lo suficientemente agradecida. Gracias a Pablo, Eva, Lúa y Teo, por ser una fuente constante de alegría en mi vida. Gracias también por cuidar y acabar queriendo a Lola durante mi estancia en Oxford. Gràcies Marta i Laia, per tot el que hem compartit, per la confiança, per acceptar-nos com som i per riure i plorar juntes en més d'una ocasió. Gracias a Lola, la compañera gatuna más bonita, arisca y cariñosa. Gracias a Mari, Antonio, Javier y Esther, por acogerme en una familia que es todo bondad. Gracias también a Laura, Andrea, Andreas, Laura, Blanca, Mario, Aaron, Jessica, Sara, Nico, Carla, Hugo, Marta, "Jimmy" (y Otto), Will, Katie, Raquel, Jessie (y Héctor), Joan y Mireia, por tantos pedacitos de tiempo compartidos.

Gracias, José Luis. Por apoyarme, ayudarme y estar a mi lado siempre y de tantas maneras diferentes. Haciendo el desayuno cada mañana, acompañándome a Oxford, aguantándome en los días malos y soñando conmigo en los días buenos. Construyendo un futuro juntos. Este camino se termina, pero nuestro viaje sigue.

Termino con un reconocimiento y un deseo. He podido escribir esta Tesis porque he tenido el privilegio de crecer como mujer en un lugar y un momento concretos: a finales del siglo XX, en un país democrático y (relativamente) igualitario, en el que existe una educación pública y una sanidad (casi) universal. Ojalá algún día nada de esto sea un privilegio, si no la norma.

Barcelona, 18 de enero del 2023

# Abstract

Three years after its emergence, the coronavirus disease 2019 (COVID-19) continues to be a major cause of morbidity and mortality worldwide. Throughout the pandemic, real-world data (RWD) have been extensively used to fill evidence gaps in the field of COVID-19. The aim of this Thesis was to characterise individuals with underlying conditions and COVID-19 infection and to investigate factors associated with COVID-19 infection and severity as well as with COVID-19 vaccine uptake and safety using RWD.

First, we provided a multi-national description of the baseline characteristics and outcomes of interest of individuals with COVID-19 infection living with obesity and with cancer history. Then, we investigated the associations between cancer and socioeconomic deprivation with COVID-19 infection and severity, as well as the associations between socioeconomic deprivation and COVID-19 vaccine uptake in Catalonia, Spain. Finally, we analysed the risks of thromboembolic events following COVID-19 infection and vaccination.

Consistent with prior research, we found that individuals living with obesity and with cancer were more susceptible to COVID-19 infection and to develop severe COVID-10. Patients with a recent cancer diagnosis and with haematological cancers were particularly vulnerable to poor COVID-19 outcomes. People living in socioeconomically deprived urban areas were also more likely to be infected and hospitalised with COVID-19 prior to the advent of COVID-19 vaccines in Catalonia. However, despite socioeconomic inequalities in COVID-19 vaccine uptake, inequalities in COVID-19 infection decreased six months after the vaccine rollout. Lastly, we found that the risks of thromboembolic events were much higher following COVID-19 infection than following COVID-19 vaccination.

The findings of this Thesis underscore the value of RWD to conduct epidemiological research and inform public health policies, particularly during emergencies. Our findings also highlight the need to address non communicable diseases and socioeconomic inequalities to reduce the burden of COVID-19 and improve the population's health.

# Resum

Tres anys després de la seva aparició, la malaltia del coronavirus 2019 (COVID-19) continua sent una causa destacada de morbiditat i mortalitat en el món. Durant la pandèmia, les dades del món real (en angles, *real-world data*) han estat extensament utilitzades per omplir mancances de coneixement en el camp de la COVID-19. L'objectiu d'aquesta tesi va ser caracteritzar les persones amb malalties preexistents i COVID-19 i investigar factors associats amb la infecció per COVID-19 i les seves complicacions, així com amb la cobertura vacunal i la seguretat de les vacunes contra la COVID-19 fent servir dades del món real.

En primer lloc, vam descriure les característiques de les persones amb COVID-19 que pateixen obesitat o que tenen un antecedent de càncer. Després, vam investigar la relació entre el càncer i la privació socioeconòmica amb el risc d'infecció i complicacions per COVID-19, així com la relació entre la privació socioeconòmica i la cobertura vacunal a Catalunya. Finalment, vam investigar els riscos d'esdeveniments tromboembòlics després de la infecció i la vacunació contra la COVID-19.

En línia amb altres estudis, hem observat que les persones amb obesitat i càncer són més susceptibles a patir la COVID-19 i les seves complicacions. Els pacients amb un diagnòstic recent de càncer i amb càncers hematològics van ser especialment vulnerables a la COVID-19. Abans de l'inici de la vacunació massiva a Catalunya, les persones que viuen en zones urbanes socialment desfavorides tenien també més riscos d'infecció i hospitalització per COVID-19. No obstant això, malgrat desigualtats socioeconòmiques en la cobertura vacunal, les desigualtats en la infecció per COVID-19 van disminuir sis mesos després de l'inici de la vacunació. Finalment, els riscos de patir esdeveniments tromboembòlics després de la infecció van ser molt més alts que després de la vacunació contra la COVID-19.

Els resultats d'aquesta tesi subratllen el valor de les dades del món real en l'àmbit de l'epidemiologia així com el seu paper per recolzar la presa de decisions en salut pública, especialment en moments d'emergència. Els nostres resultats també destaquen la importància d'abordar les malalties no transmissibles i les desigualtats socioeconòmiques per reduir la càrrega de la COVID-19 i millorar la salut de la població.

# Resumen

Tres años después de su aparición, la enfermedad por coronavirus 2019 (COVID-19) sigue siendo una causa destacada de morbilidad y mortalidad en el mundo. A lo largo de la pandemia, los datos del mundo real (en inglés, *real-world data* han sido una importante fuente de información para generar evidencia sobre la COVID-19. El objetivo de esta tesis fue caracterizar a las personas con COVID-19 con enfermedades subyacentes e investigar factores asociados con padecer COVID-19 y sus complicaciones, así como con la cobertura vacunal y la seguridad de las vacunas contra la COVID-19 usando datos del mundo real.

En primer lugar, describimos las características de personas con COVID-19 con antecedente de obesidad, así como las características de las personas con COVID-19 y antecedente de cáncer. Después, investigamos la relación entre el cáncer y la privación socioeconómica con el riesgo de infección y complicaciones por COVID-19, así como la relación entre la privación socioeconómica y la cobertura vacunal en Cataluña. Finalmente, analizamos los riesgos de eventos tromboembólicos después de la infección y la vacunación contra la COVID-19.

En línea con otros estudios, observamos que las personas con obesidad y cáncer fueron más susceptibles a la infección y a desarrollar complicaciones. Los pacientes con un diagnóstico reciente de cáncer y con cánceres hematológicos fueron especialmente vulnerables a la COVID-19. Antes del inicio de la vacunación masiva en Cataluña, las personas que vivían en zonas urbanas socialmente desfavorecidas tuvieron también más riesgo de infección y hospitalización por COVID-19. Sin embargo, a pesar de desigualdades socioeconómicas en la cobertura vacunal, las desigualdades en la infección por COVID-19 disminuyeron seis meses después del inicio de la vacunación. Por último, los riesgos de sufrir acontecimientos tromboembólicos

después de la infección fueron mucho mayores que después de la vacunación.

Los hallazgos de esta tesis subrayan el valor de los datos del mundo real en el campo de la epidemiología, así como su relevancia para apoyar la toma de decisiones en salud pública, especialmente en momentos de emergencia. Nuestros resultados también destacan la necesidad de abordar las enfermedades no transmisibles, así como las desigualdades sociales para reducir la carga de la COVID-19 y mejorar la salud de la población.

# Scientific work

## About the author

Elena Roel graduated in Medicine at the University of Barcelona (2008-2014) and specialised in Preventive Medicine & Public Health at Hospital Clínic of Barcelona (2016-2020). She holds a master's degree in Public Health (Pompeu Fabra University & Autonomous University of Barcelona, 2016-2018) and a postgraduate degree in Statistics in Health Sciences (Autonomous University of Barcelona, 2016-2019). In June 2020, she joined as a predoctoral fellow 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 (IDIAPJGol) in Barcelona, Spain. As part of her PhD training, in 2022 she did a research stay at the Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, under the supervision of Dr Prieto-Alhambra. A summary of the scientific work conducted by the author during her Thesis is provided below.

## Peer-reviewed publications

Burn E, Tebé C, Fernandez-Bertolin S, […] **Roel E**, […], Duarte-Salles T. The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nature Communications. 2021;12(1):777. doi:10.1038/s41467-021-21100-y

Recalde M**,* Roel E,*** Pistillo A,[…], 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. International Journal of Obesity. 2021;45(11):2347-2357. doi:10.1038/s41366-021-00893-4

Duarte-Salles T, Vizcaya D, Pistillo A, […], **Roel E**, […], Prieto-Alhambra. Thirty-day outcomes of children and adolescents with

COVID-19: An international experience. Pediatrics. 2021;148(3):e2020042929. doi:10.1542/peds.2020-042929

Prats-Uribe A, Sena AG, Lai LYH, […], **Roel E**, […], Prieto-Alhambra D. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. BMJ. 2021;373:n1038. doi:10.1136/bmj.n1038

**Roel E**, Pistillo A, Recalde M, […], Duarte-Salles T. Characteristics and outcomes of over 300,000 COVID-19 individuals with history of cancer in the United States and Spain. Cancer Epidemiology Biomarkers & Prevention. Published online 2021. doi:10.1158/1055-9965.epi-21-0266

Recalde M, Pistillo A, Fernandez-Bertolin S, **Roel E**, […], Duarte-Salles T. Body Mass Index and Risk of COVID-19 Diagnosis, Hospitalization, and Death: A Cohort Study of 2 524 926 Catalans. Journal of Clinical Endocrinology and Metabolism. 2021;106(12):E5030-E5042. doi:10.1210/clinem/dgab546

Reyes C, Pistillo A, Fernández-Bertolín S, […], **Roel E**, […]. Duarte-Salles T. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. BMJ Open. 2021;11(12):e057632. doi:10.1136/bmjopen-2021-057632

**Roel E**, Pistillo A, Recalde M, […], 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. International Journal of Cancer. 2022;150(5):782-794. doi:10.1002/ijc.33846

Kostka K, Duarte-Salles T, Prats-Uribe A, […], **Roel E**, […], Prieto-Alhambra D. Unraveling COVID-19: A Large-Scale Characterization of 4.5 Million COVID-19 Cases Using CHARYBDIS. Clinical Epidemiology. 2022;14:369-384. doi:10.2147/CLEP.S323292

Jacques-Aviñó C, **Roel E**, Medina-Perucha L, […]. Berenguera A. Are we leaving someone behind? A critical discourse analysis on

the understanding of public participation among people with experiences of participatory research. Villamayor-Tomas S, ed. *PLoS One*. 2022;17(9):e0273727. doi:10.1371/journal.pone.0273727

Li X,* Raventós B,* **Roel E**, […], Duarte-Salles T. Association between covid-19 vaccination, SARS-CoV-2 infection, and risk of immune mediated neurological events: population based cohort and self-controlled case series analysis. BMJ. 2022;376. doi:10.1136/bmj-2021-068373

**Roel E**, Raventós B, Burn E, […], Duarte-Salles T. Socioeconomic Inequalities in COVID-19 Vaccination and Infection in Adults, Catalonia, Spain. Emerg Infect Dis. 2022;28(11):2243-2252. doi:10.3201/eid2811.220614

Burn E,* **Roel E**,* Pistillo A, […], Duarte-Salles T. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain. Nat Commun. 2022;13(1):7169. doi:10.1038/s41467-022-34669-9

## Preprints

Lai LYH, Golozar A, Sena A, […], **Roel E**, […], Prieto-Alhambra D. Clinical characteristics, symptoms, management and health outcomes in 8,598 pregnant women diagnosed with COVID-19 compared to 27,510 with seasonal influenza in France, Spain and the US: a network cohort analysis. medRxiv. Published online January 1, 2020:2020.10.13.20211821. doi:10.1101/2020.10.13.20211821

Burn E, Sena AG, Prats-Uribe A, […], **Roel E**, […], Duarte-Salles T. Use of dialysis, tracheostomy, and extracorporeal membrane oxygenation among 240,392 patients hospitalized with COVID-19 in the United States. medRxiv. Published online January 1, 2020:2020.11.25.20229088. doi:10.1101/2020.11.25.20229088

Golozar A, Lai LYH, Sena AG, […], **Roel E**, […], Prieto-Alhambra D. Baseline phenotype and 30-day outcomes of people tested for COVID-19: an international network cohort including >3.32 million people tested with real-time PCR and >219,000 tested positive for SARS-CoV-2 in South Korea, Spain and the United States. medRxiv. Published online January 1, 2020:2020.10.25.20218875. doi:10.1101/2020.10.25.20218875

Burn E, Fernández-Bertolín S, Voss EA, […], **Roel E**, […], Duarte-Salles T. Establishing and characterising large COVID-19 cohorts after mapping the Information System for Research in Primary Care in Catalonia to the OMOP Common Data Model. *medRxiv*. Published online 2021:2021.11.23.21266734. doi:10.1101/2021.11.23.21266734

## Contributions at national and international conferences

Recalde M, * **Roel E, *** Pistillo A, et al. Characteristics and outcomes of COVID-19 patients with obesity: preliminary results of an international network study, Global Symposium Collaborator Showcase, 2020 (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, I Virtual Congress of the Spanish Society of Epidemiology (SEE), 2020 (video without defence)

**Roel E**, Pistillo A, Recalde M, et al. Cancer and the risk of Covid-19 diagnosis, hospitalisation and death: a cohort study including 4 million adults, XXXIX Congress of the Spanish Society of Epidemiology (SEE), 2021 (oral communication)

**Roel E**, Raventos B, Burn D, et al. Socioeconomic Inequalities in COVID-19 Vaccination and Infection Among Adults Living in Urban Areas of Catalonia, Spain: A Population-based Cohort Study, International Conference of Pharmacoepidemiology (ICPE), 2022 (oral communication)

**Roel E**, Raventos B, Burn D, et al. Socioeconomic Inequalities in COVID-19 Vaccination and Infection Among Adults Living in Urban Areas of Catalonia, Spain: A Population-based Cohort Study, LX Congress of the Spanish Society of Epidemiology, 2022 (oral communication)

## Grants and awards

- Rio Hortega Fellowship, Instituto de Salud Carlos III (CM20/00174), 2020.
- Scholarship to cover the fees of the course *Writing for biomedical sciences* (Open University of Catalonia), issued by Open University of Catalonia & Fundació Dr Antoni Esteve, 2020
- Scholarship to attend the 33rd Residential Summer Course of the European Educational Program in Epidemiology, issued by the Spanish Society of Epidemiology, 2021
- Scholarship to attend the International Conference for Pharmacoepidemiology (ICPE) in Copenhagen, 2022
- Travel grant to do a doctoral stay at the Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, issued by Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
- Erasmus + Traineeship grant to do a doctoral stay at the University of Oxford, 2022

## Academic training during the PhD programme

- Harvard University-edX
  - Statistics and R, 2020
  - Causal Diagrams: Draw Your Assumptions Before Your Conclusions, 2020

- Duke University - Coursera
  - Introduction to Probability and Data with R, 2020
  - Inferential Statistics, 2020

- Open University of Catalonia
  *Writing for biomedical sciences*, 2020-2021

- European Educational Program in Epidemiology
  33rd Residential Summer Course in Epidemiology, 2021
  - Epidemiological methods I & II
  - Statistical models in epidemiology I & II
  - Data analysis exercises
  - Advanced topics in statistics
  - How to deal with missing data and unmeasured confounding

- Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford (2022)
  Real World Evidence for Clinical Research Summer School

- International Conference for Pharmacoepidemiology (ICPE)
  Pre-conference skill courses, 2022
  - Risk Minimization and Communication: Science-based Approaches and Case Studies for Designing and Evaluating Effective Interventions
  - Practical Skills in Protocol Writing and Statistical Analytic Programming Relevant to Pharmacoepidemiology
  - Intermediate Pharmacoepidemiology: Approaches to Unmeasured Confounders
  - Applied Sensitivity Analyses in Pharmacoepidemiology Database Studies

## Journal peer review activity

- International Journal of Obesity (x4)
- EBioMedicine (x1)
- Emerging Infectious Diseases (x1)

## Education and outreach activities

September 2021-June 2022: Co-supervision (with Dr Talita Duarte-Salles) of a Public Health master's thesis (Pompeu Fabra University-Autonomous University of Barcelona, Spain) entitled *Cancer incidence rates for five solid tumor sites in Catalonia from 2008 to 2017 by sex and socioeconomic deprivation: a population-based cohort study*, conducted by F Mastacaneanu.

September 2022-ongoing: Co-supervision (with Berta Raventós) of a Public Health master's thesis Pompeu Fabra University-Autonomous University of Barcelona, Spain) entitled *Association between socioeconomic status and COVID-19 vaccine uptake, infection and hospitalization among children and adolescents in Catalonia*, conducted by I López Sánchez.

The studies included in this Thesis have been mentioned in several newspapers and other online resources from Spain, such as *elDiario.es, el Periódico, el Español, El Mundo, ABC, Europapress, Betevé*, or el *Punt Avui*. ER also gave an interview at two radio stations (Cadena Ser, Cadena COPE) and a television channel (TV3).

# Abbreviations

| | |
|---|---|
| ACE2 | Angiotensin Converting Enzyme 2 |
| ARDS | Acute Respiratory Distress Syndrome |
| ATE | Arterial Thromboembolism Event |
| BHA | Basic Health Area |
| BMI | Body Mass Index |
| CFR | Case Fatality Ratio |
| CHARYBDIS | Characterizing Health Associated Risks and Your Baseline Disease In SARS-CoV-2 |
| CI | Confidence Interval |
| COVID-19 | Coronavirus Disease 2019 |
| CPRD | Clinical Practice Research Datalink |
| CU-AMC-HDC | Colorado University Anschutz Medical Campus Health Data Compass |
| CUIMC | Columbia University Irving Medical Center |
| CVST | Cerebral Venous Sinus Thrombosis |
| CVT | Cerebral Venous Thrombosis |
| DAG | Direct Acyclic Graph |
| DARWIN-EU | Data Analysis and Real-World Interrogation Network |
| DNA | Deoxyribonucleic Acid |
| DVT | Deep Vein Thrombosis |
| EHR | Electronic Health Records |
| EMA | European Medicines Agency |
| HR | Hazard Ratio |
| ICD10-CDM | International Classification of Diseases, Tenth Revision, Clinical Modification |
| ICS | Institut Català De La Salut |
| ICU | Intensive Care Unit |
| IFR | Infection Fatality Ratio |
| IR | Incidence Rate |
| IRR | Incidence Rate Ratio |
| LOINC | Logical Observation Identifiers Names and Codes |
| MEDEA | Mortalidad En Áreas Pequeñas Españolas y Desigualdades Socioeconómicas y Ambientales |
| NCD | Non-Communicable Disease |
| NLP | Natural Language Processing |

| | |
|---|---|
| OHDSI | Observational Health Data Sciences and Informatics |
| OMOP CDM | Observational Medical Outcomes Partnership Common Data Model |
| OR | Odds Ratio |
| PCORnet | National Patient-Centred Clinical Research Network |
| PE | Pulmonary Embolism |
| PF4 | Platelet Factor 4 |
| RCT | Randomised Clinical Trial |
| RNA | Ribonucleic Acid |
| RT-PCR | Reverse Transcription Polymerase Chain Reaction |
| RWD | Real-World Data |
| SARS-CoV-2 | Severe Acute Respiratory Syndrome Coronavirus 2 |
| SCCS | Self-Controlled Case Series |
| SDOH | Social Determinants of Health |
| SES | Socioeconomic Status |
| SIDIAP | Information System for Research in Primary Care |
| SIR | Standardised Incidence Ratio |
| SMD | Standardised Mean Difference |
| SNOMED | Standard Nomenclature of Medicine |
| STARR | Stanford Medicine Research Data Repository |
| TTS | Thrombosis with Thrombocytopenia Syndrome |
| UK | United Kingdom |
| US | United States |
| VA | Veterans Affairs |
| VITT | Vaccine-Induced Immune Thrombotic Thrombocytopenia |
| VOC | Variant of Concern |
| VTE | Venous Thromboembolism Event |
| WHO | World Health Organisation |

# Table of Contents

# Table of Figures

# 1. INTRODUCTION

Emerging infectious diseases are a worldwide public health concern.[1] Emerging infectious diseases are new infections (i.e., not previously described in humans) or infectious diseases whose incidence has increased rapidly over the past year.[2] In late December 2020, a new infectious disease named coronavirus disease 2019 (COVID-19) emerged in China and soon evolved into a pandemic.[3–5] A pandemic can be defined as "an epidemic occurring over a very wide area, crossing international boundaries, and usually affecting a large number of people".[6] As of early 2023, more than 650 million COVID-19 cases and 6.7 million deaths have been reported globally.[7] To date, the COVID-19 pandemic is still ongoing, although the advent of vaccines against COVID-19 have transformed the threat posed by COVID-19. This introduction includes a summary of the epidemiology of COVID-19, including a timeline of the COVID-19 pandemic from late 2019 to early 2022 with a particular focus on Spain, since the studies of this Thesis were mostly underpinned by data from Catalonia, Spain. It also highlights the knowledge gaps that justified the aims of this Thesis and introduces the role of real-world data (RWD, also known as routinely collected health data) in epidemiological studies.

## 1.1. Epidemiology of COVID-19

### 1.1.1. Timeline of the COVID-19 pandemic: 2019-2022

#### i. The emergence of COVID-19

In late December 2019, cases of atypical pneumonia of unknown origin were detected in Wuhan, Hubei province, China.[3,8] On 9 January 2020, a novel coronavirus was isolated and sequenced in bronchoalveolar-lavage samples of some of these cases.[3,4] The virus, initially called 2019-nCoV, was finally named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).[9] On the same day, the World Health Organisation (WHO) named the disease caused by SARS-CoV-2 as COVID-19.[10] By late January

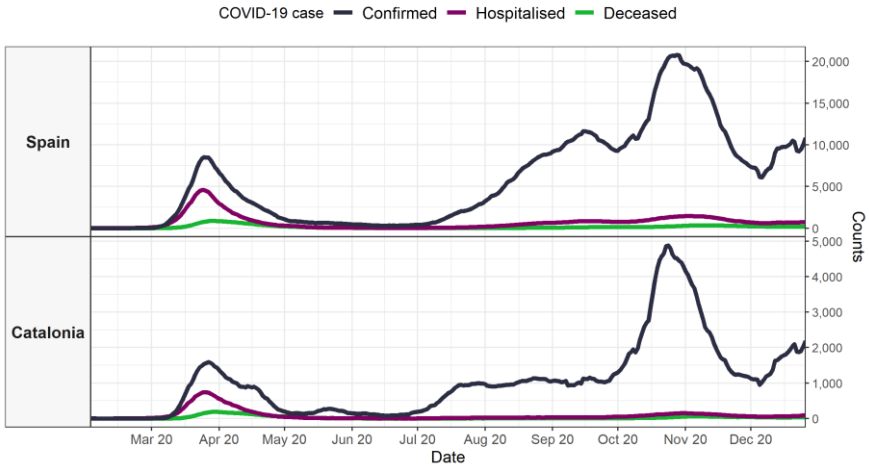
2020, there were already 7,834 confirmed cases, 98 outside China.[11] Despite the establishment of regional curfews and lockdowns in the country, the virus spread rapidly. As of 11 March 2020, with >100,000 confirmed cases in 114 countries, the WHO declared COVID-19 a pandemic.[5]

### ii. The pandemic prior to the advent of COVID-19 vaccines

In Europe, the first cases of COVID-19 were reported in January 2020,[12] although the first major COVID-19 outbreak started in Italy in late February 2020.[13] In Spain, the first COVID-19 case was reported in the Canary Islands on 31 January 2020.[14] By the end of February 2020, cases of community transmission were growing exponentially across the country. Figure 1 shows trends in the number of confirmed cases, hospitalisations, and deaths in Spain and Catalonia from late February to early December 2020.[i] On 14 March, with a total of 7,658 and 285 confirmed cases and deaths, respectively, the Spanish government declared a state of emergency (in Spanish, *estado de alarma*).[15] People's movements were restricted to commuting and groceries or pharmacy shopping, and schools were closed. This was followed by a strict nationwide lockdown on 29 March, with the suspension of all non-essential activities. The number of daily cases peaked for the first time on 20 March (10,845 cases), whereas the subsequent peak of deaths followed on 2 April (950). Following this nationwide quarantine, the growth in the number of cases declined, and restrictions were gradually lifted from May until 21 June 2020, with the end of the state of alarm.

[i] Information on the total number of SARS-CoV-2 tests performed was unavailable prior to October 2020.

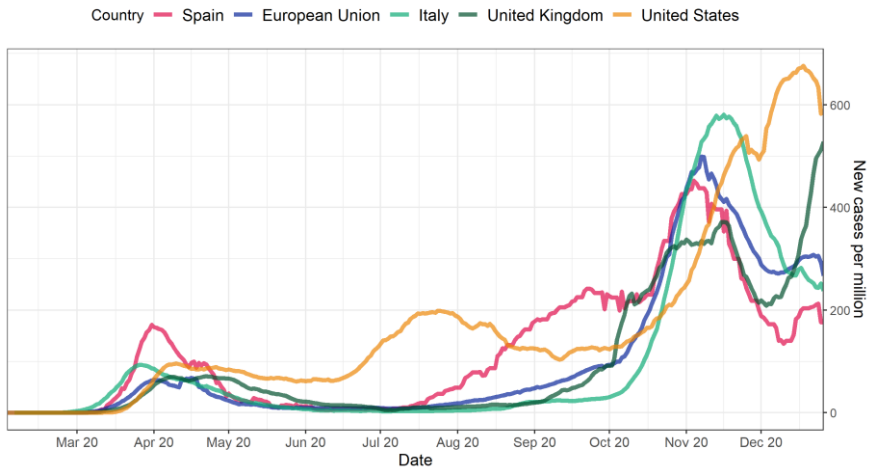
Figure 1: Daily COVID-19 cases confirmed, hospitalised, and deceased in Spain and Catalonia in 2020




Notes: 7-day rolling average.
Source: Own elaboration based on data from the Spanish National Centre of Epidemiology.[16]

The period between February and late June 2020 has been referred to as the first epidemic wave of the COVID-19 pandemic in Spain.[17] Although ill-defined, an epidemic wave denotes a pattern over time in which the incidence of disease increases, peaks, and then decreases, like a natural pattern of peaks and valleys.[18,19] Similarly, other Western countries experienced during the first semester of 2020 the first wave of the pandemic, as seen in Figure 2. In Spain, the first wave overwhelmed healthcare services in the regions where COVID-19 rates were higher, including the most densely populated cities, like Madrid or Barcelona. Elective surgeries and procedures had to be postponed, and primary care services were also disrupted.[20,21] In Madrid and Catalonia, the number of intensive care unit (ICU) beds almost tripled between March and June 2020.[22] Importantly, due to shortages in laboratory supplies, SARS-CoV-2 testing was restricted to severe COVID-19 cases and/or healthcare workers during the first wave (see section 1.1.4.,

Figure 2: Daily COVID-19 cases confirmed per million people in 2020, by country




Notes: 7-day rolling average.
Source: own elaboration based on data from Our World in Data.[23]

Diagnosis).[24,25] Thus, official figures underestimate the total number of people infected with SARS-CoV-2 during that period. As of 11 May 2020, a total of 230,000 COVID-19 cases had been confirmed, with 120,000 hospitalisations (53% of the cases) and 26,700 deaths (12% of the cases). However, a population-based seroprevalence study estimated that 5% of the Spanish population had been exposed to SARS-CoV-2 as of 11 May 2020.[26] Since Spain had 47 million inhabitants in 2020, this is equivalent to 2.3 million people infected, which is 10 times higher than the number of confirmed cases by that date.[27] Serological studies from the United Kingdom (UK), the United States (US), and Switzerland have also estimated rates of infection 10 times higher than official figures during the first wave.[28-30]

The second wave of the pandemic spanned from July to early December 2020.[17] During the summer months of 2020, SARS-CoV-2 incidence rates remained low, as seen in Figure 1. However, rates started to gradually increase in mid-September 2020, peaked in early November, and then progressively

declined until December. A similar pattern was seen in other Western countries, as seen in Figure 2. In Spain, more cases were confirmed during the second wave when compared to the first one, but the number of hospitalisations and deaths were lower. Curfews and other non-pharmacological interventions were established in various regions, including Catalonia. However, there was no rigorous lockdown during that period, and schools remained open. Unlike the first wave, SARS-CoV-2 testing was available for all individuals with suspected SARS-CoV-2 infection as well as for close contacts of cases.

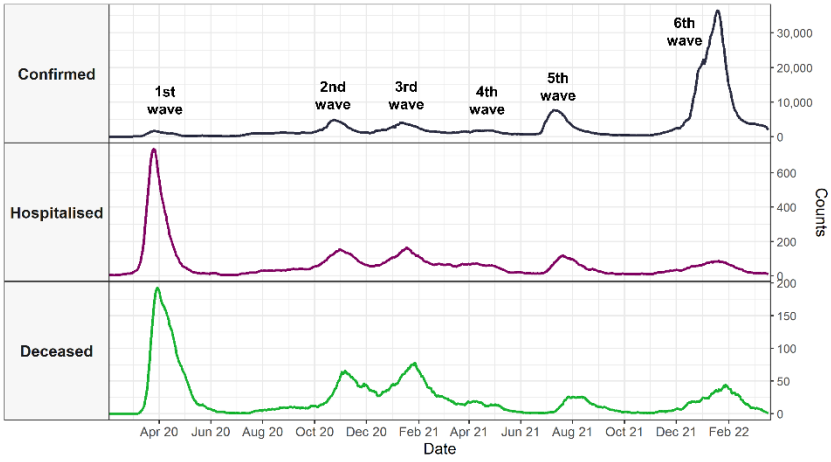
### iii. The advent of vaccines against COVID-19

Less than a year after the emergence of SARS-CoV-2, the first vaccines against COVID-19 were introduced into the market, an unprecedented milestone in the field of vaccine development.[31] In the European Union, the first COVID-19 vaccine authorised by the European Medicines Agency (EMA) was the BNT162b2 mRNA vaccine (manufactured by Pfizer BioNTech), on 21 December 2020.[32] Three other vaccines were authorised by the EMA over the following weeks: mRNA-1273 (6 January 2021, manufactured by Moderna),[33] ChAdOx1-S nCoV-19 (29 January 2021, Oxford-AstraZeneca, referred to as ChAdOx1 from now onwards),[34] and Ad.26.COV2.S (11 March 2021, Janssen).[35] A year later, on 20 December 2021, the EMA approved the NVX-CoV2373 vaccine (Novavax).[36] However, when the studies of this Thesis were conducted, only the BNT162b2, mRNA-1273, ChAdOx1, and Ad26.COV2 vaccines were available.

On 27 December 2020, Spain launched a nationwide COVID-19 vaccination campaign.[37] Spain's vaccination strategy prioritised population subgroups according to their risk of disease severity and their occupation. The first population groups eligible for vaccination were people living or working in nursing homes and healthcare workers. Gradually, other groups became eligible for vaccination, considering age (prioritising older populations), comorbidities (prioritising people with risk factors for COVID-19 severity), and occupation (prioritising people working in essential

services, such as police officers or teachers).[38] A year later after the start of the campaign, Spain had one of the highest vaccination rates worldwide, with 92% of the population aged >12 years (38 million people) vaccinated with at least one dose of a COVID-19 vaccine.[39,40] During the year 2021, four epidemic waves were observed: from late December 2020 to mid-March 2021 (3rd wave), from mid-March to June 2021 (4th wave), from July to mid-October 2021 (5th wave) and from mid-October 2021 to March 2022 (6th wave).[17] As shown in Figure 3, in Catalonia, COVID-19 hospitalisations and deaths were lower during these waves than during the first wave. Conversely, the total number of confirmed cases was higher, especially in the sixth wave. This wave coincided with the emergence of the SARS-CoV-2 variant Omicron (see section, 1.1.2. Virology).

Figure 3: Daily COVID-19 cases confirmed, hospitalised, and deceased in Catalonia from March 2020 to March 2022

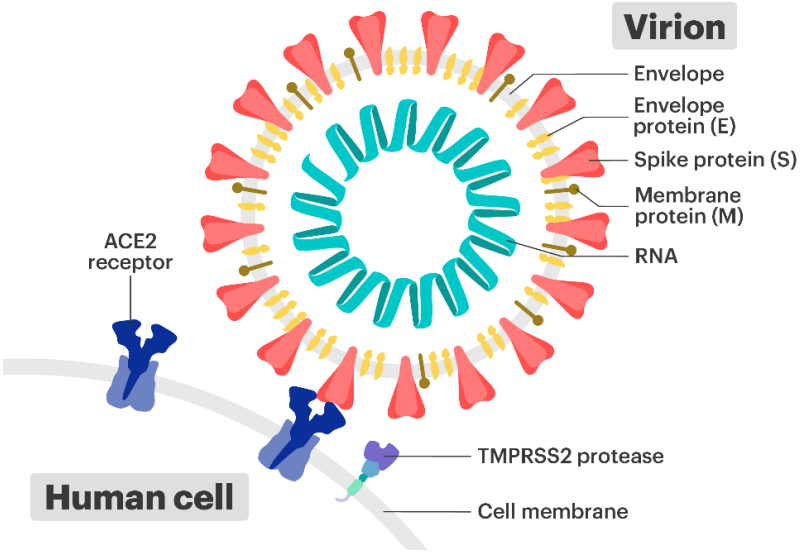



Notes: 7-day rolling average.
Source: Own elaboration based on data from the Spanish National Centre of Epidemiology.[16]

### 1.1.2. Virology

SARS-CoV-2 is a ribonucleic acid (RNA) virus including ≈30,000 nucleotides and 15 open reading frames (i.e., spans of deoxyribonucleic acid (DNA) or RNA sequence between the start and stop codons).[3,4,41] It contains four structural proteins: the spike (S), the envelope (E), the membrane (M), and the nucleocapsid (N). Briefly, the spike, envelope, and membrane constitute the viral envelope (the outer layer of the virus), whereas the nucleocapsid packages the RNA genome. SARS-CoV-2 invades host cells through the binding of a receptor-binding domain located on the S protein with a cellular receptor. The main SARS-CoV-2 receptor is angiotensin converting enzyme 2 (ACE2) (see Figure 4).

Figure 4: SARS-CoV-2 virion binding to a human cell through the angiotensin converting enzyme 2 receptor



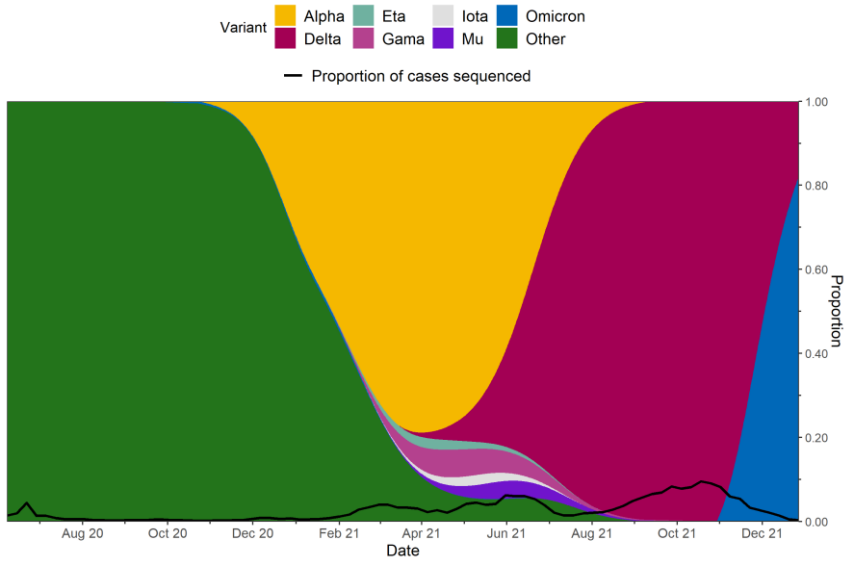

Source: Davian Ho for the Innovative Genomics Institute, available at https://innovativegenomics.org/free-covid-19-illustrations/.

ACE2 can be found in several human organs and systems, including the lungs (in type 2 alveolar epithelial cells), the heart, the kidneys, and the digestive and nervous central systems.[42] The widespread distribution of ACE2 receptors in the body might explain the diversity of symptoms associated with SARS-CoV-2 infection (see section 1.1.3. Clinical presentation). After binding, the genome of the virus is released into the host cell cytoplasm, initiating the replication of the viral genome inside the host cell.[43]

Over time SARS-CoV-2 has evolved, acquiring certain changes due to mutation and selection that have given rise to new genetic variants of the virus across the globe. Some of these have been considered variants of concern (VOC), due to their increased transmissibility, increased virulence, or decreased effectiveness of diagnostics, vaccines, or treatments.[44] Figure 5 shows the distribution of SARS-CoV-2 variants in Spain from June 2020 to December 2021.

In 2021, three SARS-CoV-2 lineages predominated in Spain: Alpha (B.1.1.7 lineage), Delta (B.1.617.2 lineage), and Omicron (B.1.1.529 lineage). Alpha was identified in September 2020 in the UK and by February 2021 became the dominant variant in Spain.[44,45] Alpha has been reported to be more transmissible and severe than wild-type SARS-CoV-2.[46,47] In October 2020, Delta was identified for the first time in India and replaced Alpha as the dominant variant in Spain by July 2021.[44,45] Delta is more transmissible and severe than Alpha, with an estimated transmissibility increase of 97%.[46] Omicron was identified in South Africa in November 2021 and rapidly became the dominant variable in Spain as of December 2021.[44,45] Compared to Delta, Omicron has been found to be more transmissible but less virulent.[48] However, Omicron has also been associated with decreased vaccine effectiveness (VE).[49]

Figure 5: Distribution of SARS-CoV-2 variants in Spain from 1 June 2020 to 31 December 2021




Source: own elaboration using data from the European Centre for Disease Prevention and Control.[50]

Regarding transmission, SARS-CoV-2 is mainly transmitted from person to person through air, via close-range contact (within 2 metres of distance).[51] Virus particles are expelled when an infected person talks, coughs, sneezes, or breathes out, and can then infect a person by inhalation (short-range airborne transmission) or when contacting its mucous membranes (droplet transmission). Viral particles can also remain suspended in the air for long time periods and spread in overcrowded or inadequately ventilated indoor environments.[51] Thus, it can be transmitted over longer distances (long-range airborne transmission).[51,52] SARS-CoV-2 has also been detected in non-respiratory specimens, such as blood or ocular secretions, as well as on surfaces.[53] However, transmission through these routes remains unclear. Studies suggest that SARS-CoV-2 can be transmitted from approximately 2 days before to 7-10 days after symptom onset, with individuals being more contagious at early

stages of acute disease.[54,55] Asymptomatic individuals can also transmit the disease.

### 1.1.3. Clinical presentation

#### i. Symptoms

Among individuals non-immune to SARS-CoV-2 (i.e., people not previously infected or people not vaccinated), SARS-CoV-2 infection clinical presentation is heterogeneous and ranges from asymptomatic to severe illness.[56–58] Approximately 30-40% of cases are asymptomatic.[59,60] Among symptomatic cases, symptoms usually start 4-5 days after exposure and last for 4-5 days.[57] The most common symptoms are cough, fatigue, and fever.[56,57] However, COVID-19 can manifest with a broad spectrum of symptoms, including sore throat, nasal congestion, shortness of breath, myalgia, headaches, loss of smell or taste, nausea, vomiting, and diarrhoea. Symptoms might also differ depending on the SARS-CoV-2 variant, with some studies suggesting that people infected with Omicron have higher odds of presenting with sore throat but less odds of loss of smell when compared to those infected with Delta.[61]

#### ii. Acute complications

Approximately 15% of symptomatic patients infected with wild-type SARS-CoV-2 suffered from a severe disease course. Typically, severe illness develops a week after the onset of symptoms.[62] COVID-19 most common complication is bilateral pneumonia, which can evolve into acute respiratory failure and, ultimately, death.[57] In a Chinese report from February 2020, 15% of 44,500 confirmed cases presented with severe disease (defined as shortness of breath, hypoxia, or lung infiltrates >50%) and 5% with critical disease (septic shock, respiratory and/or multiple organ failure).[63] Reports from May 2020 from the US showed similar patterns, with 14% of 1.3 million confirmed cases hospitalised and 2% with intensive services requirements.[64] In

Spain, as of May 2020, 53% of confirmed cases had been hospitalised, which is reflective of the testing restrictions in place during the first wave.[27] Conversely, from mid-May to late December 2020, 6% of confirmed cases were hospitalised.[65] In subsequent waves, hospitalisation rates changed also due to the emergence of new variants. For instance, the Omicron variant has been reported to be associated with less disease severity when compared to Delta.[48]

Aside from respiratory complications, SARS-CoV-2 can also present with other acute complications. These include, among others, cardiovascular complications (e.g., acute myocardial infarction, arrhythmias, heart failure), thromboembolism events (e.g., deep vein thrombosis, pulmonary embolism), and neurologic complications (e.g., ataxia, encephalopathy, Guillain-Barre syndrome, seizures, stroke).[58]

### iii. Mortality

Different estimators can be used to measure the mortality of a disease, such as the Case Fatality Ratio (CFR) or the Infection Fatality Ratio (IFR).[66,67] The CFR represents the proportion of deaths among confirmed cases. It is calculated as the number of deaths divided by the number of cases diagnosed. Therefore, the CFR depends on the likelihood of identifying cases, which in the case of COVID-19 differed across countries and over time, mostly due to different testing policies. In Spain, the CFR during the first wave was 12%,[68] which is in line with estimates from European countries as of July 2020, with CFRs ranging from 14 to 19%.[69] Conversely, from mid-May to late December 2020, the CFR in Spain was 1.3%.[65]

The IFR reflects the risk of dying when contracting the disease.[66] It is calculated as the number of deaths divided by the number of people infected. However, estimating the IFR of COVID-19 is challenging, since a substantial proportion of COVID-19 cases remain asymptomatic or do not seek care when presenting mild symptoms. Studies estimating IFR rely generally on serological

studies that estimate the proportion of the population exposed to the virus. During the first wave of the pandemic in Spain, the estimated IFR was 0.8%.[70] This is in line with studies from other countries, with IFRs ranging from 0.6 to 1%.[28,30,67]

#### iv. Long-term complications

After acute infection, 10 to 30% of survivors continue to experience symptoms and signs months after recovery.[71–73] Long-lasting symptoms following acute COVID-19 infection have been referred to as long COVID or post-COVID condition,[74–76] although this condition remains poorly understood. Long COVID symptoms are diverse and generally unspecific, those most common are fatigue, dyspnoea, and cognitive dysfunction.[73] Additionally, SARS-CoV-2 infection has also been associated with increased risks of presenting a broad range of complications one year after the acute phase of the disease.[77] This include cardiovascular (e.g., arrhythmias, cerebrovascular disorders, ischemic and non-heart ischemia disease, and thromboembolisms), renal (e.g., acute kidney injury, end-stage kidney disease, major adverse kidney events), and mental health (e.g., anxiety, depressive, stress, sleep, and substance use disorders) outcomes, although evidence on this matter is still scarce.[78–80]

### 1.1.4. Diagnosis

#### i. Clinical suspicion

COVID-19 should be suspected in every person presenting with symptoms suggestive of a viral respiratory infection. However, as discussed before, COVID-19 symptoms and signs are rather unspecific, and symptoms and signs evaluation have a poor diagnostic performance.[81] Thus, the presence or absence of any of the abovementioned symptoms is not sufficient to neither establish nor exclude a diagnosis of COVID-19.

### ii. Reverse-transcription polymerase chain reaction (RT-PCR)

The gold standard test to diagnose an acute SARS-CoV-2 infection is a nucleic acid amplification test, such as the reverse transcription polymerase chain reaction (RT-PCR) test.[41] RT-PCR tests detect viral RNA through a two-step process in which RNA sequences are transcribed into DNA and then amplified using a PCR. SARS-CoV-2 RT-PCR tests can detect RNA regions encoding proteins of the nucleocapsid, the envelope, or the spike of the virus. Upper respiratory tract samples, such as nasopharyngeal, nasal and oropharyngeal swabs, as well as saliva specimens, are the most-recommended specimen samples. SARS-CoV-2 RT-PCR tests were the first tests to be developed, and thus were the only tests available to confirm a COVID-19 diagnosis at the beginning of the pandemic.[82] Although SARS-CoV-2 RT-PCR tests are highly specific (i.e., the probability of obtaining a negative test result when a true non-case is tested), concerns have been raised regarding their sensitivity (i.e., the probability of obtaining a positive test result when a true case is tested) in the clinical setting.[83,84] Test sensitivity depends on the sample used, as well as the duration of the infection at the time of testing.[85] Indeed, the probability of a positive SARS-CoV-2 RT-PCR result declines over time after symptoms onset.[86] Therefore, a positive RT-PCR test generally confirms a COVID-19 diagnosis, whereas a negative RT-PCR test excludes COVID-19 in the majority of cases. However, in the event of a high clinical suspicion repeating testing or other diagnostic tests should be considered.

### iii. Antigen tests

Antigen tests directly detect the presence of viral proteins (antigens) produced by the virus.[82] These tests were developed later than RT-PCR tests and were introduced in mid-2020.[82] Antigen tests include laboratory-based antigen tests (i.e., tests performed in laboratories, for example at the hospital level) and rapid diagnostics tests that can be self-administered outside the healthcare setting. In Spain, antigen tests have been used to

confirm a COVID-19 diagnosis since June 2020.[87] However, in 2020-2021 people that reported a positive test result after self-testing were required to test positive in an additional confirmatory test, either a RT-PCR test or a rapid antigen test performed at primary care centres.[88]

When compared to RT-PCR tests, antigen tests are faster and less expensive. However, antigen tests have lower sensitivities than RT-PCR tests, especially among asymptomatic cases. For example, in a study including 1,732 asymptomatic cases from June to August 2020, antigen test sensitivity and specificity were 61% and 100%, respectively, when compared to RT-PCR.[89] Among 307 symptomatic cases, sensitivity and specificity were 72% and 99%, respectively.

#### iv. Antibody tests

Antibody tests detect IgM or IgG antibodies targeting a specific antigen. In the advent of an infection/vaccination, IgM antibodies are produced first, and are later followed by IgG antibodies. In the case of SARS-CoV-2, IgG antibodies can be detected approximately 14 days after infection/vaccination.[90] However, antibody tests cannot be used to confirm an acute COVID-19 infection due to their lack of sensitivity and specificity.[90] These tests can be used to detect people exposed to SARS-CoV-2 or vaccinated against SARS-CoV-2 and have been frequently used in epidemiological studies.[90]

### 1.1.5. Management

Since COVID-19 manifests often as a mild disease, most patients with COVID-19 are managed in the outpatient setting. In countries with a primary care-based health system such as Spain, primary care has been fundamental in the COVID-19 response, diagnosing, triaging, and managing most patients with COVID-19 throughout the pandemic.[91] Currently, management in the outpatient setting includes counselling (on infection control

measures and warning symptoms), symptomatic therapy (e.g., analgesics, antipyretics), and, in patients with higher risk of progression to severe disease, treatment with COVID-19 antivirals (e.g., nirmatrelvir/ritonavir, which was authorised in the European Union in early 2022).[92,93]

In the inpatient setting, hospital care for COVID-19 patients has substantially evolved throughout the pandemic.[94] In the first months, care was limited to supportive treatment with adjuvant therapies (e.g., corticosteroids, anticoagulants).[94] Repurposed drugs (i.e., drugs authorised for another indication) were also widely used, despite limited evidence on their effectiveness.[94] One of these repurposed drugs was hydroxychloroquine, an antimalarial drug that was initially administered on the basis of in vitro activity against SARS-CoV-2 but that was later seen to increase mortality among COVID-19 patients.[95] Nowadays, clinical guidelines recommend the use of corticosteroids in combination with COVID-19 antivirals and/or monoclonal antibodies in patients with severe disease that require supplemental oxygen.[96,97]

### 1.1.6. Factors associated with COVID-19 infection and severity

This section provides an overview of factors associated with COVID-19 infection and severity before the advent of COVID-19 vaccines, including demographics, underlying conditions (with a particular focus on obesity and cancer, the exposures of interest in **Study I** and **Studies II** and **III**, respectively), and socioeconomic status (the exposure of interest in **Study V).**

#### i. Demographics: age and sex

People of all ages can be infected by SARS-CoV-2. However, in early stages of the pandemic COVID-19 was mostly diagnosed among middle (40-64 years) and old-aged people (≥65years).[57,98] In Spain, as of May 2020, the median age of confirmed cases was

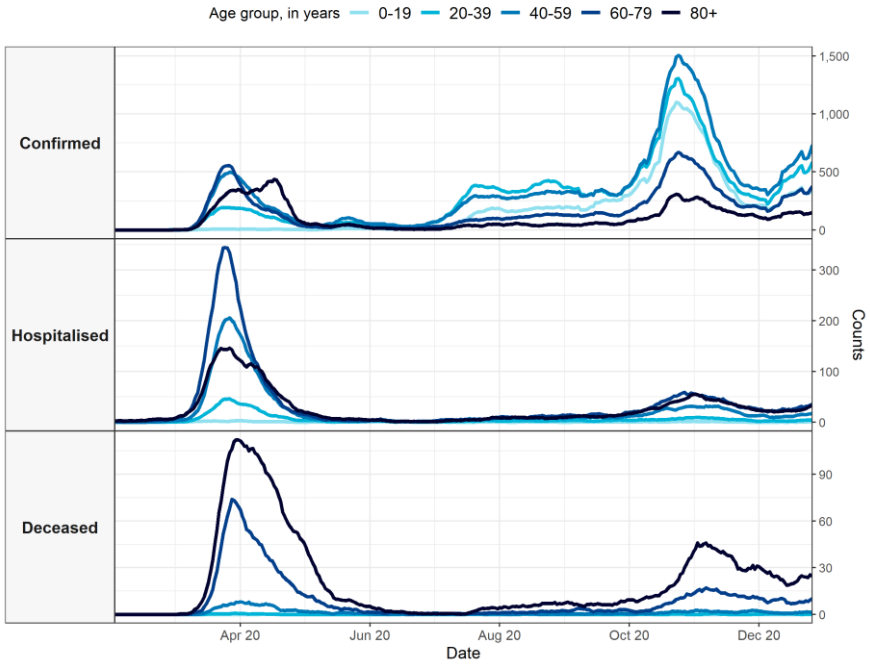
50 years, and 1 out of 4 were aged ≥80 years.[68] Over time, however, the age distribution of COVID-19 cases evolved, as seen in Figure 6, which is related to changes in testing patterns and the COVID-19 vaccine rollout. Despite this, the majority of severe cases have consistently been reported among the eldest.[63,99] During the first wave, the CFR in Spain ranged from 0.2% among children aged 0-9 years to 22% for adults aged 90 years,[68] and solid evidence shows that age is the most prominent risk factor for COVID-19 severity.[100–103] This might be related to an impaired immune system (due to ageing), a higher prevalence of comorbidities among the eldest, or medication use for these comorbidities.[104]

Males and females can be both infected by SARS-CoV-2.[105] In Spain, as of May 2020, 57% of confirmed cases were females,[68] whereas in the aforementioned serological study both females and males had a seroprevalence rate of 4.6%.[26] Population-based studies from other countries also reported similar seroprevalence rates by sex.[106] Conversely, men account for the majority of severe COVID-19 cases.[100,101,107,108] During the first wave in Spain 56% of patients hospitalised, 69% of patients admitted to the ICU, and 57% of patients deceased were males.[68] In addition, male sex has been consistently associated with increased risks of COVID 19 hospitalisation and death, even after accounting for the effect of comorbidities [100,101,109] Increased severity among males has been postulated to be related to immune differences linked to X chromosomes and to hormonal differences.[109,110]

### ii. Underlying conditions

In early reports, a substantial proportion of COVID-19 cases had underlying conditions, such as cardiovascular disease, chronic obstructive pulmonary disease, diabetes, or hypertension, among others.[57,98,101] In Spain, as of May 2020, 65% of confirmed cases had at least one underlying condition of interest. Those most common were cardiovascular diseases (29%), hypertension (20%) and diabetes (16%).[68] Additionally, the prevalence of comorbidities was higher among those with severe

Figure 6: Daily COVID-19 cases confirmed, hospitalised, and deceased in Catalonia from February to December 2020, by age group




Notes: 7-day rolling average.
Source: Own elaboration based on data from the Spanish National Centre of Epidemiology.[16]

disease.[63,68,101,102,111,112] In Spain, 81% of hospitalised cases had at least one underlying condition of interest.[68] Later studies have shown an association between several conditions and increased risks of severe COVID-19.[101] According to a systematic review from the US Centre for Disease Control, there is strong evidence that the following conditions are associated with higher risks of severe COVID-19 outcomes (i.e., hospitalisation, intensive services requirements, or death): cancer, cerebrovascular, chronic kidney, chronic lung, and chronic liver diseases, dementia, disabilities, heart conditions, human immunodeficiency virus infection, mental health disorders, obesity, physical inactivity, pregnancy, primary

immunodeficiencies, tuberculosis, type 1 and type 2 diabetes, smoking (current/former), solid organ or haematopoietic cell transplantation, and use of immunosuppressive drugs.[99]

### iii. Obesity

Obesity is an abnormal or excessive fat accumulation that presents a risk to health.[113] BMI, which is calculated by dividing weight (in kilograms) by height (in metres squared), is a method frequently used to assess obesity.[114] According to the WHO, a BMI ≥30 kg/m$^2$ is considered obese.[113] Obesity is a well-known risk factor for several noncommunicable diseases (NCDs), such as cardiovascular diseases or type 2 diabetes, but it has also been associated with increased risks of communicable diseases, such as respiratory viral infections.[115] Obesity has also been associated with increased risks of severe COVID-19,[115–117] which might be related to the metabolic alterations associated with obesity, such as insulin resistance, increased glucose levels in blood, alterations in adipose-tissue-derived hormones (i.e., adipokines, such as adiponectin or leptin), and chronic low-grade inflammation.[115] These alterations impair the immune response. Furthermore, obesity is associated with a restrictive respiratory pattern, characterised by a reduced compliance of the lungs, as well as with other respiratory diseases such as obstructive sleep apnoea syndrome or obesity hypoventilation syndrome.[49] Lastly, obesity is also strongly associated with conditions that also increased the risks of severe COVID-19 (e.g., cardiovascular diseases or type 2 diabetes).[115]

### iv. Cancer

Cancer is characterised by an abnormal and uncontrollable growth of cells that can invade nearby tissues or other parts of the body.[118] Although there are several types of cancer, cancers can be classified in solid cancers (cancers that originate in solid organs of the body, such as breast or prostate cancer) and haematological cancers (cancers that originate in blood-forming tissues, such as leukaemia, lymphoma, or multiple myeloma).

Several studies have shown an association between cancer and increased risks of COVID-19 infection and severe disease.[119–122] This might be explained by the fact that patients with cancer are particularly susceptible to infections, due to the characteristics of the cancer itself, treatment-related immunosuppression, as well as increased exposure to infections due to higher interactions with the healthcare system and increased use of invasive devices, such as catheters.[123] Infections are particularly frequent among patients with haematological malignancies, since these cancers often infiltrate the bone marrow (impairing the production of immune cells, such as neutrophils, lymphocytes, or macrophages) and cancer treatments target the bone marrow itself.[124]

### v. Socioeconomic factors

Early reports describing the characteristics of individuals with COVID-19 lacked information on socioeconomic factors, such as ethnicity, educational level, occupation, or socioeconomic status (SES).[125] This in part because such information is not generally recorded in health data sources.[125] For example, in Spain, official reports issued by the Spanish National Centre for Epidemiology describing the characteristics of COVID-19 cases included only information on demographics (age, sex), baseline conditions, symptoms, and outcomes.[65,68] Reports from the Catalan Health Department also lacked information on socioeconomic variables.[126] However, later studies showed that people from disadvantaged populations, such as people living in socioeconomically deprived areas, ethnic minorities, or migrants, displayed higher incidence rates of SARS-CoV-2 infection in early stages of the pandemic.[28,112,127–132]

The disproportionate burden of SARS-CoV-2 infection among these populations is likely reflective of increased exposure related to occupation and housing conditions.[127,133] On the one hand, people from disadvantaged populations have more frequently jobs that cannot be undertaken from home.[127] Therefore, they have a higher risk of contracting the infection due

to increased interactions with others, in the workplace or while commuting. Risks are higher for people that have a great deal of social interaction, such as people working in contact with the public or people working in high-population-density work environments. On the other hand, people with low income or migrants are more likely to share a house and to live in overcrowded households.[134] The implementation of preventive measures, such as quarantines or isolations, is challenging in overcrowded households. A person with SARS-CoV-2 infection might not be able to isolate in a separate bedroom, and facilities such as the bathroom or the kitchen are often shared by many household members, thus increasing the risks of infection for other residents.

People living in socio-economically deprived areas and from ethnic minorities are also at increased risks of severe COVID-19.[127,128,131,135,136] In Barcelona, hospitalisation and mortality rates were 1.5 times and 1.2 times higher among people living in the most deprived areas when compared to those in the least deprived areas during the first wave.[112] In the UK, a study reported that people from Black and South Asian ethnicities had 5 times and 2 times higher risks of COVID-19 death, respectively, when compared to White from March 2020 to January 2021.[136] Increased rates of hospitalisation and COVID-19 deaths among disadvantaged populations are also reflective of poorer baseline health status.[137] For instance, as discussed before, conditions such as cardiovascular diseases, obesity, or type 2 diabetes have been associated with increased risks of severe COVID-19. In Western countries, including Spain, these conditions display a socioeconomic gradient, with generally higher incidence rates among people from disadvantaged population groups, such as people with low SES or from ethnic minorities.[138–142]

### 1.1.7. Vaccines

#### i. Types

When the studies of this Thesis were conducted, there were two types of vaccines against COVID-19: mRNA-based vaccines and adenovirus-based vaccines. BNT162b2 and mRNA-1273 are mRNA-based vaccines.[32,33] These vaccines are based on sequences of synthetic messenger RNA (mRNA) that encode SARS-CoV-2 viral spike antigens. mRNA sequences enter the host cell, which then produces these antigens, activating the immune response. These vaccines are the first vaccines in history authorised for human use based on mRNA.[143]

ChAdOx1 and Ad.26.COV2.S are non-replicating adenovirus-based vaccines.[34,35] These vaccines use a viral vector to introduce genetic material that codes for a spike antigen into the host cell. In the case of SARS-CoV-2 vaccines, this vector is a modified adenovirus. Before the COVID-19 pandemic, other adenovirus-based vaccines had been tested in clinical trials to prevent diseases such as Ebola, Malaria, or tuberculosis, among others. Thus, evidence regarding potential side-effects associated with this type of vaccine was scarce prior to the pandemic.[144]

These vaccines also have different administration regimens. Three are 2-dose regimen vaccines: BNT162b2 (recommended interval between doses: 21 days), mRNA-1273 (28 days), and ChAdOx1 (4 to 12 weeks), whereas Ad.26.COV2.S is a 1-dose regimen vaccine. Additional doses for people with weakened immune systems and booster doses can also be administered after primary vaccination, although the administration of extra doses outside the primary vaccination scheme was not contemplated initially (see the next section, ii. Effectiveness).

### ii. Effectiveness

COVID-19 vaccines showed high effectiveness against symptomatic SARS-CoV-2 infection in randomised clinical trials (RCT), such as 95% for BNT162b2,[145] 94% for mRNA-1273,[146] 70% for ChAdOx1,[147] and 67% for Ad.26.COV2.S.[148] Studies conducted in the real-world setting shortly after the start of the vaccine rollout confirmed high effectiveness of these vaccines. In an early study from Israel including 590,000 BNT162b2 vaccine recipients, VE against symptomatic infection, hospitalisation, and death was 94%, 87%, and 92%, respectively, 7 days after second-dose vaccination.[149] This is in line with findings for other COVID-19 vaccines and from other countries, including Spain,[150–153] although later studies including longer follow-ups raised concerns regarding protection duration.[154]

Concerns were also raised over immune evasion of emergent SARS-CoV-2 variants, Omicron in particular.[155,156] A study from the US showed that during the Delta variant predominance mRNA VE decreased from 86% during the first six months following second dose administration to 76% more than six months after, and from 52% to 38% during the Omicron variant predominance.[157] Due to these concerns, the EMA recommended in December 2021 the administration of booster doses to people aged ≥5 years who had previously completed a primary COVID-19 vaccination scheme.[158] Later, in April 2022 and July 2022, the EMA also recommended second booster doses for individuals aged ≥80 years and people aged 60-79 years, respectively.[159] Booster doses have showed increased VE for all SARS-CoV-2 variants, including Omicron.[157,160]

### iii. Safety

In RCT settings, COVID-19 vaccines showed a good safety profile.[145–148] The most frequent adverse effects reported were mild and short-term, and included injection-site reaction, headache, fatigue, and myalgia. However, RCT include relatively small samples of people (usually 500-3,000 patients,[161] although

COVID-19 vaccine trials included 20,000-40,000 participants)[145–148] and are, therefore, not able to capture rare adverse events or subgroup populations with increased risks of such events. In March 2021, concerns were raised in Europe and the US regarding the safety of adenovirus-based vaccines due to spontaneous reports of thromboembolic events in unusual sites (e.g., cerebral venous sinus thrombosis (CVST)) associated with low platelets counts within three weeks following vaccine administration.[162–166]

These events were described as a new disease called vaccine-induced immune thrombotic thrombocytopenia (VITT),[167] an immune thrombotic thrombocytopenia syndrome mediated by antibodies against platelet factor 4 (PF4). VITT is also often referred to as thrombosis with thrombocytopenia syndrome (TTS).[163,164] Although both terms are often used interchangeably, some authors consider TTS as a broader term that encompasses thrombotic thrombocytopenia of any cause following COVID-19 vaccination, regardless of the documented presence of antibodies against PF4 (which are not routinely measured and can be difficult to capture in real-world studies).[162]

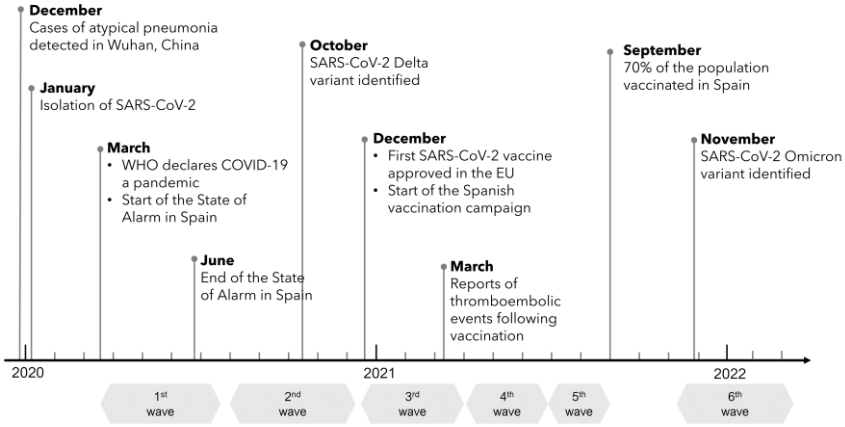
In early reports, cases of VITT/TTS were reported mostly among middle-aged females 5 to 16 days after first dose ChAdOx1 administration. Although it was unclear whether there was a causal relationship between vaccination and these events, several countries in Europe suspended vaccination with ChAdOx1, or restricted the use of this vaccine to some age subgroups (of note, Ad.26.COV2.S was not yet available in Europe at that time).[168] In Spain, ChAdOx1 was initially administered to essential workers aged ≤55 years and later restricted to people aged 60-69 years.[169]

Although this Thesis focuses on the risks of TTS, COVID-19 vaccines have also been associated with increased risks of other adverse events, including myocarditis and pericarditis following vaccination with mRNA-based vaccines (mostly among males aged <40 years).[170,171]

### 1.1.8. Key points

- In late December 2020, COVID-19 emerged in China and evolved into a pandemic by March 2020. A year later, Western countries launched nationwide COVID-19 vaccination campaigns.
- Since its emergence, SARS-CoV-2 has mutated and variants with various levels of transmissibility, virulence, and immunity to vaccines have appeared.
- SARS-CoV-2 clinical presentation ranges from asymptomatic to severe illness (15% of cases in early stages of the pandemic). After acute infection, 10 to 30% of patients continue to experience symptoms, a condition that has been labelled as post-COVID-19 condition or long COVID-19.
- SARS-CoV-2 RT-PCR and antigen tests can confirm a diagnosis of acute COVID-19.
- Being older, male sex, baseline comorbidities (including obesity and cancer), and low socioeconomic status are associated with increased risks of severe acute COVID-19.
- Concerns about the safety of COVID-19 vaccines were raised in March 2020 following spontaneous reports of rare thromboembolic events following vaccination with adenovirus-based vaccines.

Figure 7: Timeline of the COVID-19 pandemic




Source: own elaboration

## 1.2. Limitations of evidence and knowledge gaps

Despite the fast pace of COVID-19-related publications during the first months of the pandemic,[172] several uncertainties regarding the characteristics of COVID-19 cases and factors associated with COVID-19 infection and severity remained at the start of this Thesis. **Studies I, II, III** and **IV** aimed to fill some of these evidence-gaps. In addition, **Studies IV** and **V** aimed to provide timely answers to concerns over vaccine uptake and safety that emerged in the months following the COVID-19 vaccine rollout, which started in Europe (and in Spain) six months after the beginning of this Thesis.

### 1.2.1. Characteristics of COVID-19 cases

As discussed by Fox et al, descriptive studies are fundamental in public health.[173] Briefly, descriptive studies aim to characterise how exposures, conditions, or diseases are distributed in well-defined populations and in specific contexts (e.g., in one or more

geographical settings, or time points) and provide the basis for causal or predictive studies. In the event of a new disease, such as COVID-19, timely and well-designed descriptive studies are essential to inform public health strategies and guide future epidemiological studies.

Unfortunately, COVID-19 descriptive studies were scarce during the first months of the pandemic, and information on the characteristics of COVID-19 cases was mostly published by public health authorities.[173] These studies and reports used different criteria to identify COVID-19 cases, underlying conditions, or COVID-19 complications, which limited the comparability of their findings. More importantly, most of these studies/reports included only patients hospitalised or with confirmed infection. This was a major limitation, since they included a biased subsample of COVID-19 cases, thus incurring in selection bias.[133] Therefore, their results were not generalisable to the overall COVID-19 population, and the risks of complications were overestimated since participants were mostly severe cases. Additionally, most studies/reports described only outcomes such as hospitalisation, ICU admission, death, or respiratory outcomes (e.g., pneumonia, acute respiratory distress syndrome (ARDS)). Only small case series studies reported the incidence rates of other adverse outcomes such as cardiovascular events (e.g., myocardial infarction) or thromboembolic events (e.g., pulmonary embolism). Thus, the incidence rates of these outcomes during the acute phase of COVID-19 infection remained unknown. Lastly, descriptive studies targeting specific subgroup populations were scarce. For example, studies describing the characteristics and outcomes among people with COVID-19 and with underlying conditions, such as obesity or cancer, were lacking. Therefore, representative, and large population-based studies providing a detailed characterisation of COVID-19 patients among specific subgroup populations were needed.

### 1.2.2. Factors associated with COVID-19 infection and severity

While descriptive studies were scarce in the early stages of the pandemic, several studies attempted to identify factors associated with COVID-19 infection and severity. Most of these studies analysed the associations between several exposures (e.g., age, hypertension, diabetes) and COVID-19-related outcomes using models adjusted by different covariates, without considering the underlying causal relationships between variables. This approach led to controversial findings, often misinterpreted as causal associations.[173,174]

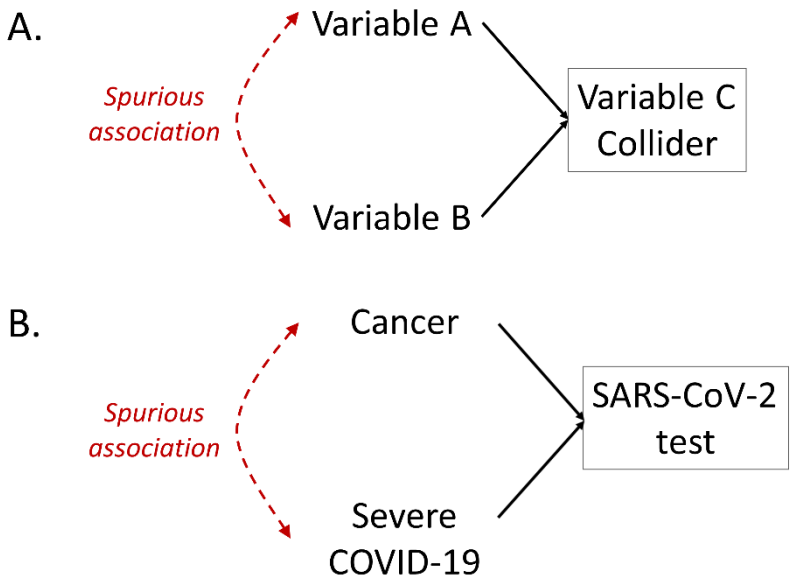
An early study underpinned by the OpenSAFELY platform is an illustrative example of this.[173–175] This study including 17 million people from the UK analysed the associations between several factors and the risks of COVID-19-related deaths through multivariable modelling.[176] Briefly, the study provided estimates for a single fully adjusted model that included all the exposures of interest. This practice, which has been referred to as the Table 2 Fallacy, can lead to erroneous interpretations.[177] For example, the study found a negative association between smoking and COVID-19 death, thus suggesting a protective effect of tobacco (although the authors highlighted that their model likely included mediators of the association between smoking and COVID-19, such as chronic respiratory disease). While Williamson et al argued that their study had a descriptive purpose (and thus, results should not be interpreted as causal),[178] they used terms such as "increased risk", "reduction", or "attributable" when describing the associations between exposures and COVID-19 deaths, which might be interpreted as causal. Later studies addressing the causal relationship between tobacco and COVID-19 outcomes found the opposite, that is, that smoking is associated with increased risks of poor COVID-19 outcomes.[179]

Another major limitation of early studies aiming to identify factors associated with COVID-19 outcomes is selection bias, and in

particular, collider bias.[180,181] Collider bias occurs when controlling or restricting (by design or analysis) on a variable that is influenced by the exposure and the outcome of interest. Using directed acyclic graphs (DAGs),[182] collider bias can be illustrated as a variable that has 2 arrows colliding, as shown in Figure 8.A. When adjusting or controlling for this collider variable (variable C), a spurious association is found between the variables that originate those arrows (variables A and B).

For example, when SARS-CoV-2 testing was not widely available, the likelihood of being tested for COVID-19 might have been influenced by having cancer (exposure) and by the severity of the disease (outcome). Thus, in analyses restricted to people tested for COVID-19, an association between cancer and risk of severe disease will be observed, regardless of the causal nature or not of the observed relationship (see Figure 8.B).

Figure 8: Illustration of collider bias using a Direct Acyclic Graph




Source: own elaboration

In summary, evidence regarding factors associated with COVID-19 infection and severity was limited due to confounding and collider bias. Furthermore, while some factors were consistently associated with increased risks of poor COVID-19 outcomes (e.g., age), other factors showed conflicting results or were rarely addressed. For example, at the beginning of the pandemic, the relationship between cancer and COVID-19 was unclear. While some studies found an association between cancer and COVID-19 infection and complications,[119,122,183] others found null associations.[111,184,185] These studies, however, used different cancer definitions, and were mostly small and prone to collider bias, since they included only a biased subsample of COVID-19 cases (e.g., confirmed or hospitalised cases during the first wave of the pandemic).[186]

Moreover, the first studies reporting factors associated with COVID-19 outcomes did not include information on socioeconomic characteristics.[187] Concerns over socioeconomic COVID-19 inequalities emerged in late April 2020, with early reports from the UK and the US showing higher SARS-CoV-2 infection and mortality rates among people from ethnic minorities, such as African American or Hispanic in the US or Black and South Asian in the UK.[188–190] Later studies analysing the associations between ethnicity, socioeconomic deprivation, and occupation found increased risks of COVID-19 infection and severity among people from vulnerable population groups.[125] However, these studies were mostly conducted in the US or the UK, and evidence from south-eastern European countries, such as Spain, was scarce.

### 1.2.3. Vaccine uptake

In the months following the start of the COVID-19 vaccine rollout, studies from the UK and the US showed inequalities in COVID-19 vaccine uptake. In a study conducted in the UK three months after the start of nationwide vaccination, Black minorities had

lower vaccine coverages than White people (68% vs 97%) among people aged ≥80 years. Further, coverage ranged from 91% for those living in the most deprived areas to 97% for those living in the least deprived areas.[191] In the US, as of May 2021, vaccine coverage ranged from 49% to 59% in the most and least deprived counties, respectively.[192]

In Catalonia, one year after the start of the vaccine rollout, coverage reached 90% for people aged ≥12 years.[39] Coverage differed by age category, with coverages ranging from 79% among those aged 30-39 years (the 10-band age group with the lowest vaccination coverage) to 100% for people aged ≥80 years. However, official reports from Catalonia and Spain regarding COVID-19 vaccination coverage only provided estimates of vaccine coverage by age group, sex, and county.[193] Thus, evidence regarding potential differences in COVID-19 vaccine uptake by socioeconomic status was missing. Furthermore, population-based studies exploring the relationship between socioeconomic factors and vaccine uptake in Spain were lacking. Similarly, although several studies from the US and the UK had analysed the associations between socioeconomic factors and COVID-19 infection and hospitalisation before the advent of COVID-19 vaccines, studies analysing the impact of the vaccine rollout on infection socioeconomic inequalities were lacking six months after the start of the COVID-19 vaccine rollout.

### 1.2.4. Vaccine safety

Despite concerns regarding risks of TTS following COVID-19 vaccination, as of June 2021, large and representative observational studies addressing this matter were scarce and limited to first doses.[194,195] A study from Denmark and Norway published in May 2021 including first dose ChAdOx1 vaccine recipients aged 18-65 years showed that individuals vaccinated had increased risks of venous thromboembolic events when compared to a historical background cohort.[194] Another study

from Scotland found suggestive evidence that ChAdOx1 was associated with increased risks of idiopathic thrombocytopenic purpura, whereas BNT161b2 was not found to be associated with increased risks of thromboembolic events.[195] In addition, these studies did not assess the risks of thromboembolic events following COVID-19 infection, and, therefore, further research was needed to understand the risks of TTS following vaccination with different COVID-19 vaccines and doses as well as following COVID-19 infection as a benchmark.

### 1.2.5. Key points

1. Descriptive studies were scarce in early stages of the pandemic, and little was known about the characteristics of patients with COVID-19 among people with underlying conditions, such as people living with obesity or with cancer.
2. Early studies addressing factors associated with COVID-19 infection and severity were mostly small and prone to selection bias. In the field of cancer, prior studies found conflicting results regarding the associations between cancer and COVID-19 outcomes.
3. In the first months after the start of the COVID-19 vaccine rollout, little was known about inequalities in COVID-19 vaccine uptake by socioeconomic status. Furthermore, evidence regarding the impact of COVID-19 vaccination on socioeconomic inequalities in COVID-19 infections and hospitalisations was lacking.
4. As of June 2021, only a couple of studies had analysed the relationship between first dose COVID-19 vaccines and thromboembolic events, with inconsistent results.

## 1.3. Real-world data

### 1.3.1. Definition

Real-world data (RWD) refers to data related to the health status and/or the delivery of health care that is routinely collected.[196] Examples of RWD include electronic health records (EHR), administrative claims data, disease registries (e.g., cancer registries), or data gathered through personal devices (e.g., mobile health applications). This type of data, although not primarily designed for research purposes, is being increasingly used in epidemiology and clinical research, especially in the field of pharmacoepidemiology.[197,198] For instance, RWD are particularly useful to conduct post-marketing effectiveness and safety studies, since they allow the inclusion of large and inclusive populations and are reflective of real practice conditions.[197,199] For example, Suchard et al recently analysed the comparative effectiveness and safety of first-line antihypertensive therapies using RWD from 9 different databases.[200] RWD have also been used to describe the natural history of diseases and drug utilisation patterns, to analyse the associations between exposures and outcomes of interest, to develop patient-level prediction models, and to provide insights into the quality of health systems, among others.[102,200-203] For example, Hripcsack et al described treatment pathways for type 2 diabetes, hypertension, and depression using RWD from four countries.[201] Recalde et al investigated the associations between body mass index (BMI) and different cancer types using EHR from Catalonia.[202] Ross et al developed a predictive model of COVID-19 outcomes using RWD from six countries.[203]

### 1.3.2. Strengths and limitations

The major strengths of RWD are their large sample size, allowing the study of rare outcomes, as well as their representativeness of the general population, which increases the generalisability and

external validity of study results.[197] They also include long follow-ups and are therefore suited to study long-term outcomes. Additionally, when compared to RCT or to other observational data sources (such as large cohorts) the use of RWD allows to conduct studies in a timely and inexpensive manner.[199] However, RWD has limitations. First, concerns are often raised regarding data quality since researchers do not control data collection. Furthermore, data collection might differ depending on clinical practice standards, which vary across settings and time periods. This might lead to heterogeneous results that are challenging to interpret, although they are reflective of the particularities of different settings. RWD might also lack information on some variables of interest. For example, EHR often lack information on lifestyle variables such as diet, physical activity, or working conditions, among others. Lastly, as in any observational study, appropriate methods must be used to minimise the risks of selection, misclassification, and confounding bias when using RWD.

Ideally, a RWD database would be updated regularly and include large and representative populations, as well as high quality information on multiple items (e.g., lifestyle factors, medical diagnoses, laboratory tests, drug prescriptions) from multiple healthcare settings (e.g., outpatient, hospital, emergency care) while preserving patient confidentiality.[161] In practice, RWD are heterogeneous in terms of size, data quality, information included, and settings. Thus, when using RWD, researchers must assess the fitness for purpose of their data, i.e., the appropriateness of the data to provide answers for a specific research question.[204]

### 1.3.3. The use of common data models

Routinely collected databases store their information using a wide variety of structures, formats, and terminologies (e.g., databases might have a different number of tables, or record

clinical information using different coding nomenclatures).[205] This limits the comparability of results across data sources, as well as hinders the conduction of multi-database studies.[205] Multi-database studies can provide insights into different contexts (e.g., understand patterns of medication in different countries), and have increased external validity and statistical power.[206] Different approaches have been used to conduct multi-database studies, such as running studies locally using analytic codes specific to each database (with methods developed in agreement across study partners) or sharing raw data to a central partner, which runs the analysis.[206] These approaches can be particularly time-consuming and sharing raw data is often challenging due to confidentiality issues.

Another approach is the use of common data models. This approach consists of transforming data stored in disparate data sources into a common and standard format, with harmonised terminologies, vocabularies, and codes.[197,207] When using common data models, researchers can apply a common analytical code to each database locally (rather than adapting the code to each database).[206] Then, aggregated results are shared across data partners. This approach expedites the results-obtention process while maintaining patient-level data locally in a secure setting. To date, several common data models have been developed, such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), the US FDA Sentinel (used mostly for post-marketing drug safety studies), the National Patient-Centered Clinical Research Network (PCORnet), or the ConcepTION CDM.[207–210]

### 1.3.4. Key points

- Real-world data (RWD) are data routinely collected related to the health status and/or the delivery of health care, such as electronic health records or claims data.
- The major strengths of RWD are their large sample size, long follow-ups, and representativeness of the general population. However, RWD might have limited data quality as well as limited information on some variables, such as lifestyle or socioeconomic factors.
- The use of common data models can facilitate the conduction of network studies, increasing generalisability and statistical power while protecting patient's confidentiality.

# 2. RATIONALE

The COVID-19 pandemic has caused an unprecedented health, economic, and social disruption worldwide. The emergence of this new disease raised several research questions that needed to be urgently addressed to inform the COVID-19 response. However, although the literature in the COVID-19 field evolved rapidly in early stages of the pandemic, several research gaps remained to be addressed at the start of this Thesis. Over the last years, RWD have been increasingly used in the field of epidemiology. Due to their large size, representativeness, high statistical power, and, most importantly, due to their availability, RWD can be an important asset in the field of COVID-19 research.

For instance, RWD could provide insights into the characteristics and occurrence of outcomes among subgroup populations of interest with COVID-19, such as individuals living with obesity or with a history of cancer, which have been poorly characterised in prior studies. RWD could also be used to analyse the associations between COVID-19-related outcomes and exposures of interest, such as cancer or socioeconomic deprivation, which remain unclear. The advent of vaccines against COVID-19 has also raised questions regarding inequalities in vaccine uptake as well as regarding the impact of vaccination campaigns on COVID-19 infection inequalities. Such questions might be answered in a timely manner using RWD. Lastly, concerns have been raised over COVID-19 vaccines safety. The use of RWD could help understand the relationship between COVID-19 vaccines and the occurrence of rare adverse events, such as thromboembolic events.

# 3. OBJECTIVES

The main aim of this Thesis was to describe the baseline characteristics and the occurrence of outcomes among people with COVID 19 infection and underlying conditions and to investigate factors associated with COVID-19 infection and severity as well as with COVID-19 vaccine uptake and safety using real-world data.

This thesis had five specific objectives:

1. To describe the baseline demographic and clinical characteristics, as well as the occurrence of outcomes among patients diagnosed and hospitalised with COVID-19 in population subgroups with underlying conditions of interest, such as obesity (**Study I** in the Results section) or cancer (**Study II** in the Results section).

2. To investigate the associations between cancer and the risk of COVID-19 diagnosis, hospitalisation, and death, overall and by years since cancer diagnosis and cancer subtype (**Study III** in the Results section).

3. To investigate the associations between socioeconomic deprivation and the risks of COVID-19 diagnosis, hospitalisation, and death before and after the COVID-19 vaccine rollout (**Study IV** in the Results section).

4. To characterise the COVID-19 vaccine rollout and to investigate the association between socioeconomic deprivation and COVID-19 vaccine uptake (**Study IV** in the Results section).

5. To describe the incidence rates of thromboembolic events and thrombocytopenia following COVID-19 infection and vaccination, and to compare these with incident rates among the general population before the COVID-19 pandemic (**Study V** in the Results section).

# 4. METHODS

Herein, we provide an overview of the study designs, settings, data sources, study populations, exposures, outcomes, and statistical analyses that were used in the studies included in this Thesis. More details about the methods of each study can be found in Section 5 Results, in the Methods section of each article.

## 4.1. Study designs

The five studies of this Thesis were cohort studies underpinned by RWD. All the databases contributing to the studies of this Thesis have been standardised to the OMOP CDM.[211] **Studies I** and **II** were international network studies including several databases from Spain, the UK, and the US. **Studies III, IV** and **V** were conducted using the Information System for Research in Primary Care (SIDIAP) database, from Catalonia, Spain.[212]

## 4.2. Settings and data sources

### 4.2.1. The CHARYBDIS project

**Studies I** and **II** were conducted using healthcare databases contributing to the CHARYBDIS (Characterizing Health Associated Risks and Your Baseline Disease In SARS-COV-2) project.[213] This project, designed by the Observational Health Data Sciences and Informatics (OHDSI) community,[214] aimed to characterise patients with COVID-19 using RWD from different countries and settings.

**Study I** included data from January to June 2020 from Spain (one database), the UK (one database), and the US (four databases). Data from Spain came from the primary care database SIDIAP, which is explained in detail further on when presenting the data source for **Studies III, IV** and **V**. Data from the UK came from the primary care database Clinical Practice Research Datalink (CPRD). Data from the US included EHR from the hospital setting and

claims data. Hospital data came from the Columbia University Irving Medical Center (CUIMC, New York) database, covering New York-Presbyterian Hospital and its affiliated physician practices; the Stanford Medicine Research Data Repository (STARR- OMOP, California), with data from Stanford Health Care, and the Department of Veterans Affairs (VA-OMOP, national), covering the national Department of Veterans Affairs health care system, which includes more than 9 million enrolled Veterans. Claims data came from IQVIA Open Claims (national), which are pre-adjudicated claims covering over 300 million people (~80% of the US population).

**Study II** included data from patients diagnosed with COVID-19 between January to October 2020, as well as data from patients diagnosed with influenza in 2017-2018, from Spain and the US. Data from Spain came from the SIDIAP database, whereas data from the US included hospital and claims data. Hospital data included the Colorado University Anschutz Medical Campus Health Data Compass (CU-AMC-HDC; Colorado), CUIMC (New York), Optum-EHR (national), STARR-OMOP (California), and VA-OMOP (national). Claims data included HealthVerity and IQVIA-OpenClaims (both national).

### 4.2.2. The SIDIAP database

**Studies III, IV** and **V** were conducted using the SIDIAP database (www.sidiap.org). **Study III** included data from 1 March 2020 to 6 May 2020, **Study IV** from 1 September 2020 to 30 June 2021, and **Study V** from 1 January 2017 to 30 June 2021. SIDIAP is a primary care database from Catalonia, an autonomous community in the Northeast of Spain. Spain has a public healthcare system in which primary care is free of charge and has a gatekeeping role. SIDIAP data is routinely collected by healthcare workers from 328 primary care centres that belong to *Institut Català de la Salut* (ICS), the main healthcare provider in Catalonia.[212] SIDIAP includes pseudo-anonymised EHR since

2006 of people registered in primary care centres from ICS, approximately 5.8 million people (75% of the population living in Catalonia) and is representative of the general population in terms of age, sex, and geographic distribution.[212] SIDIAP data comprises data on anthropometric measurements (e.g., weight), demographics (e.g., sex, age, nationality), disease diagnoses (coded using the International Classification of Diseases, Tenth Revision, Clinical Modification [ICD10-CM]), drug prescriptions and dispensations, laboratory test results, lifestyle factors (e.g., smoking), and socioeconomic status, among others. SIDIAP can also be linked to hospital discharge records from public and private hospitals of Catalonia (*Conjunt Mínim Bàsic de Dades d'Alta Hospitalària*), as well as to the Catalan public health vaccine registry. These linkages were used for **Studies IV** and **V.** The studies included in this Thesis used SIDIAP data standardised to the OMOP CDM.[215]

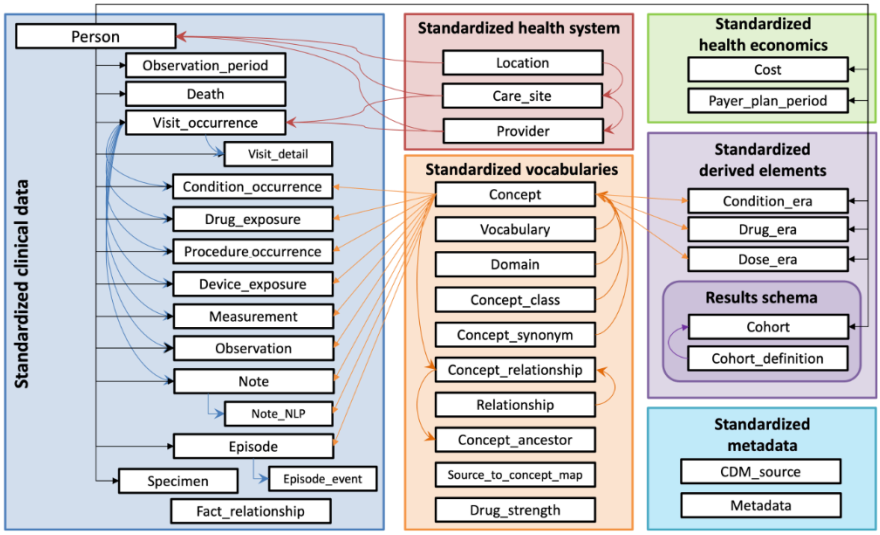
### 4.2.3. The OMOP CDM

The OMOP CDM is maintained by OHDSI, an open-science international network of researchers and observational health databases.[216] To date, more than 100 de-identified healthcare databases including records from over one billion individuals from 20 countries across all the continents have been mapped or are in progress of mapping to the OMOP CDM.[216] The OMOP CDM is a data standard that includes Clinical, Health Systems and Health economics data tables.[211] An overview of the structure of the current OMOP CDM is shown in Figure 9.[211] The OMOP CDM is person-centred: all the clinical tables are linked to the person table through a unique personal identifier and information is encoded using standardised vocabularies.[207]

In the OMOP CDM, conditions in the Condition table are recorded using the Standard Nomenclature of Medicine (SNOMED), whereas drugs are recorded using RxNorm (extended, to include drugs approved in the US and in

Europe),[207,217] and measurements using Logical Observation Identifiers Names and Codes (LOINC). Thus, the process of mapping data to the OMOP CDM includes transforming data stored in a specific nomenclature to standard vocabularies. For example, in SIDIAP a diagnosis of melanoma would be recorded as C43, *Malignant melanoma of skin,* in the ICD, 10th nomenclature. In OMOP CDM, this would be recorded in the Condition table, using the SNOMED Concept ID 141232, *Malignant melanoma of skin*. Importantly, although the information is transformed into standard vocabularies, the OMOP CDM also stores the original source concepts, therefore no information is lost when mapping the data.

Figure 9: Structure of the OMOP CDM version 5.4




Source: The Book of OHDSI.[216]

## 4.3. Study populations

For all the studies included in this Thesis, participants were required to have at least one year of medical history available prior to study start to comprehensively capture baseline characteristics. Figure 10 includes an overview of the study periods of each study.

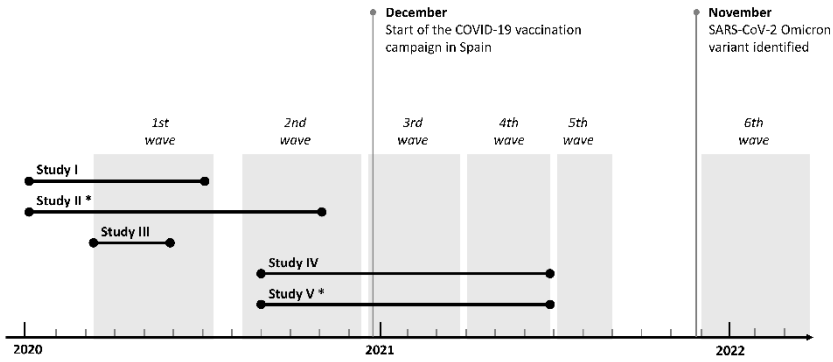
**Study I** included all individuals diagnosed with COVID-19 and hospitalised with COVID-19 between January and June 2020. **Study II** included all individuals with a history of cancer diagnosed with COVID-19 and hospitalised with COVID between January and June-October 2020 (depending on the contributing database) as well as hospitalised with seasonal influenza in 2017-2018. The index date for those diagnosed and hospitalised was the day of diagnosis and hospitalisation, respectively. Participants were followed until the end of follow-up (30 days), exit from the database, or end of data collection, whichever occurred first.

**Study III** included adults (18 years and older) without a COVID-19 diagnosis prior to study start (1 March 2020). Participants were followed until the occurrence of a state of interest (see sections 4.5. Statistical analyses and 5.3. Study III) or end of data collection (6 May 2020), whichever occurred first.

**Study IV** included people aged 40 years and older living in urban areas with information on socioeconomic deprivation and without a COVID-19 diagnosis prior to study start (1 September 2020). To analyse inequalities in vaccine uptake, only individuals with a complete follow-up as of 30 June 2021 were included. To analyse inequalities in COVID-19 infections and hospitalisations, participants were followed until the occurrence of an outcome of interest, death, exit from the database, or end of study period (prior to vaccine rollout: 26 December 2020, after vaccine rollout: 30 June 2021), whichever occurred first.

**Study V** included people aged 20 years and older vaccinated with BNT162b2 or ChAdOx1 between 27 December 2020 and 23

Figure 10: Overview of the study periods of the studies included in this Thesis




Source: Own elaboration
Notes: For each study, the period of data collection is depicted through a black line delimited with dots. To provide some context, the figure illustrates the waves of the pandemic in Spain as grey areas, as well as highlight some important milestones. * Studies II and V also include data from 2017-2018 and from 2017-2019, respectively.

June 2021 or infected with SARS-CoV-2 between 1 September 2020 and 23 June 2021. An historical comparator cohort was also generated, including people aged 20 years and older registered in the SIDIAP database between 1 January 2017 and 31 December 2018. All cohorts were followed until the occurrence of the outcome of interest, end of follow-up (21 days for those vaccinated, 90 days for those diagnosed with COVID-19, and 31 December 2019 for the background general population cohort), exit from the database, or end of data collection (30 June 2021), whichever occurred first.

## 4.4. Variables

### 4.4.1. COVID-19 infection

**In Studies I, II,** and **V**, COVID-19 infection was an exposure, whereas in **Studies III** and **IV**, COVID-19 infection was an outcome. Additionally, these studies used different criteria to identify people infected with COVID-19

**Studies I, II** and **III** included data from the first months of the pandemic (mostly from January until June 2020), when SARS-CoV-2 testing was restricted to severe cases. Therefore, in those studies, we used test results as well as clinical diagnoses of COVID-19 to identify COVID-19 cases. In **Studies I** and **II**, the concept sets used to identify clinical diagnoses were: Suspected disease caused by severe acute respiratory coronavirus 2 (SNOMED Concept ID: 37311060), Disease due to Coronaviridae (4100065), Disease caused by severe acute respiratory syndrome coronavirus 2 (37311061), and Coronavirus infection (439676). In **Study III**, COVID-19 diagnoses were identified based on a record of a clinical code for COVID-19 (37311061).

Conversely, **Studies IV** and **V** included data from the second wave of the pandemic onwards, when SARS-CoV-2 testing was widely available. Therefore, in those studies, we used positive test results to identify COVID-19 cases. These included SARS-CoV-2 RT-PCR and antigen tests.

### 4.4.2. COVID-19 hospitalisation

**In Studies I** and **II**, COVID-19 hospitalisation was an exposure, whereas in **Studies III** and **IV**, COVID-19 hospitalisation was an outcome. Additionally, these studies used different criteria to identify people hospitalised with COVID-19.

In **Studies I** and **II**, COVID-19 hospitalisation was defined as a hospitalisation episode along with a clinical diagnosis or positive

SARS-CoV-2 within a time window from 21 days prior to admission up to the end of the hospitalisation episode to give sufficient time for patients to be hospitalised as well as sufficient time to be tested. In **Study III**, COVID-19 hospitalisation was defined as a hospitalisation episode along with a clinical diagnosis or positive RT-PCR SARS-CoV-2 test result between 21 days before and 3 days after admission (to minimise the risk of including hospital-acquired infections). In **Study IV**, COVID-19 hospitalisation was defined as a hospitalisation episode along with a positive SARS-CoV-2 test (RT-PCR or antigen) between 21 days before and 3 days after admission.

### 4.4.3. COVID-19 vaccination

Vaccination against COVID-19 was an outcome in **Study IV** and an exposure in **Study V**.

In **Study IV**, we identified individuals vaccinated with a first dose of any of COVID-19 vaccine, namely BNT162b2, mRNA-1273, ChAdOx1, and Ad.26.COV2.S.

In **Study V**, we identified individuals vaccinated with BNT162b2 (first and second doses) and ChAdOx1 (first doses). ChAdOx1 second doses as well as mRNA-1273 and Ad.26.COV2.S vaccines were not included due to small sample sizes.

### 4.4.4. Additional exposures of interest

In **Study I**, we described the characteristics of patients diagnosed and hospitalised with COVID-19 stratified by a previous record of a diagnosis of obesity. Obesity was defined as any previous record of a clinical diagnosis of obesity, or a body mass index (BMI) measurement between 30 and 60 $kg/m^2$, or a body weight measurement between 120 and 200 kg. We used SNOMED codes to identify obesity diagnoses, including *Obesity* (concept

ID: 433736), *Body mass index 30+ - obesity* (4060985), *Pulmonary hypertension with extreme obesity* (4081038), *Obesity associated disorder* (4176962), *Body mass index 40+ - severely obese* (4256640) and *Lymphedema associated with obesity* (45766204).

In **Study II**, we included only individuals with a history of cancer, which we defined as having a record of a malignant neoplasm excluding non-melanoma skin cancer prior to the date of study start. We used SNOMED codes to identify individuals with a history of cancer, including *Malignant neoplastic disease* (concept Id: 443392) and H*/O: malignant neoplasm* (4144289) and we excluded those with a record of a *squamous cell carcinoma of skin* (4111921), *History of malignant neoplasm of skin* (4179242) and *Basal cell carcinoma of skin* (4112752).

In **Study III**, history of cancer was the exposure of interest, which we defined as having a record of a primary invasive solid or haematological cancer excluding non-melanoma skin cancer prior to the date of study start. Since this study was underpinned exclusively by the SIDIAP database, we used source codes to identify cancer diagnoses, including the ICD-10-CM codes C00 to C96 aside from C44 (*non-melanoma skin cancer*) and C77-C79 (secondary cancers).

In **Study IV**, socioeconomic deprivation was the exposure of interest. We measured socioeconomic deprivation using a validated composite score, the *Mortalidad en áreas pequeñas españolas y desigualdades socioeconómicas y ambientales* (MEDEA) index.[218] The MEDEA deprivation index was calculated for census tract urban areas using information from the Spanish national census of 2001. It is based on 5 indicators; three are related to working conditions (proportion of unemployment, manual workers, and eventual workers) and two to education level (proportion of people unable to read and write, overall and among young people). The MEDEA deprivation index is linked to each individual's residential address (using the most recent address registered in SIDIAP) and categorised it into quintiles of

deprivation, with the first quintile (Q1) representing the least deprived area and the fifth quintile (Q5) representing the most deprived area.

### 4.4.5. Additional outcomes of interest

In **Studies I** and **II** we assessed the following 30-day outcomes: ARDS, acute kidney injury, cardiovascular disease events, deep vein thrombosis (DVT), pulmonary embolism (PE), sepsis, requirement of intensive services (identified by a recorded mechanical ventilation and/or a tracheostomy and/or extra-corporeal membrane oxygenation procedure), and death (from all causes). Specific definitions using SNOMED codes were created for each of these outcomes, which can be consulted in the Appendix of the corresponding published studies.

In **Studies III** and **IV,** we also assessed COVID-19-related deaths. Since we lacked information on cause of death, we defined COVID-19- related deaths as deaths of any cause occurring: i) after a COVID-19 event (diagnosis or hospitalisation) in **Study III,**; and ii) within 28 days following a COVID-19 diagnosis in **Study IV**. The latter approach is in line with definitions used by other researchers, as well as by countries such as the UK.[176,219]

In **Study V**, the outcomes of interest were thromboembolic events, thrombocytopenia, and TTS (a proxy of VITT). These included venous thromboembolism events (VTE, including DVT and PE) and arterial thromboembolism events (ATE, including myocardial infarction and stroke). TTS was defined as an occurrence of a thromboembolic event in which thrombocytopenia was seen between 10 days before and 10 days after the thromboembolic event. Thrombocytopenia was identified using diagnostic codes or a measurement of a platelet count between 10,000 and 150,000 platelets/microliter. We used definitions previously established, [220,221] developed in agreement with the EMA.

A detailed description of the concept codes included to identify each outcome can be found at https://livedataoxford.shinyapps.io/CovCoagOutcomesCohorts/, under the tab "Included (Source) Concepts".

## 4.5. Statistical analyses

In **Studies I** and **II**, we conducted our analyses through a federated network approach. We developed a common analytical code that was run locally at each participating database site. Only aggregated results were shared across data partners. We reported demographics, comorbidities, and outcomes of interest by cohort and database as proportions (calculated by the number of persons within a given category, divided by the total number of persons), with corresponding 95% confidence intervals (CI). In **Study II**, we also reported the ranking and prevalence of the 10 most common cancer types by frequency. To compare characteristics between study cohorts, we calculated standardised mean differences (SMD), with an |SMD|≥0.1 indicating a meaningful difference in the prevalence of a given condition.[222]

In **Study III,** we investigated the associations between history of cancer and the risks of COVID-19 diagnosis, hospitalisation, and death using a multi-state model approach. This approach can be used to describe processes in which individuals transition from one health status to another.[223] Our model included four different health statuses: general population (or baseline), diagnosed with COVID-19, hospitalised with COVID-19, and death. We fitted multivariable-adjusted Cox proportional hazard models and estimated hazard ratios (HR)s with corresponding 95% CI for five different transitions: from general population to diagnosed with COVID-19 (1), hospitalised with COVID-19 (2), or death (3), and from diagnosed with COVID-19 to hospitalised with COVID-19 (4), and to death (5). We stratified our results by years since cancer diagnosis (<1 year, 1-5 years and ≥5 years), sex, age (<70

years and ≥70 years), and cancer type (solid, haematological, as well as by 5 solid cancer types: breast, prostate, colorectal, lung, and bladder). We established the age cut-off at 70 years because this was the median age observed among patients with cancer. All models were relative to cancer-free patients and adjusted by age, sex, smoking, socioeconomic deprivation, and baseline comorbidities; missing data were handled as an additional category. We used a direct acyclic graph (DAG) to guide our modelling strategy. In sensitivity analyses, we re-estimated our models stratifying by calendar time, restricting participants to never smokers, and after performing multiple imputation of the variables with missing data (smoking and socioeconomic deprivation).

In **Study IV**, we analysed the association between socioeconomic deprivation and non-vaccination by age group (working-age: 40-64 years, retirement-age: ≥65 years) using logistic regression models and estimated Odds Ratios (OR) with 95% CI. To analyse the associations between socioeconomic deprivation and COVID-19 infection, hospitalisation, and death before and the start of the COVID-19 vaccine rollout, we performed multivariable-adjusted Cox proportional hazard models and calculated HR with 95% CI: by age group and period (3 months before and 1 to 6 months after the start of the COVID-19 vaccine rollout). Models were relative to the least deprived quintile and adjusted by age, sex, and nationality. Our modelling strategy was guided by a DAG. In sensitivity analyses, we estimated our models for vaccination coverage excluding people with a COVID-19 infection during follow-up and analysed the associations between socioeconomic deprivation and COVID-19 outcomes restricting participants to Spanish individuals and using a different time period after the start of the vaccine rollout (3 to 6 months after).

In **Study V**, we used the historical rate comparison method to compare incidence rates (IR) of thromboembolic events among vaccines recipients and among COVID-19 cases to IR among a

historical background population (2017-2018). We first calculated IR of thromboembolic events in the 21 days following first dose COVID-19 vaccination and in the 90 days following COVID-19 infection by dividing the total number of events by the person-time at risk per 100,000 person-years with 95% CI. We then calculated crude incidence rate ratios (IRR) with 95% CI: to compare IR in the vaccinated and infected COVID-19 cohorts compared to the background population cohort. We also estimated the number of events expected among the vaccinated and diagnosed with COVID-19 cohorts using indirect standardisation (10-year age bands), with the general population cohort as the standard population. We then calculated standardised incidence ratios (SIRs) with 95% CI dividing the number of events observed by the number of events expected.

In all the studies included in this Thesis, we blinded results with less than 5 individuals for confidentiality purposes. Analyses were conducted using R.

# 5. RESULTS

## 5.1. Study I: Characteristics and outcomes of 627 044 COVID-19 patients living with and without obesity in the United States, Spain, and the United Kingdom

Recalde M,* Roel E,* Pistillo A, Sena AG, Prats-Uribe A, Ahmed WU, Alghoul H, Alshammari TM, Alser O, Areia C, Burn E, Casajust P, Dawoud D, DuVall SL, Falconer T, Fernández-Bertolín S, Golozar A, Gong M, Lai LYH, Lane JCE, Lynch KE, Matheny ME, Mehta PP, Morales DR, Natarjan K, Nyberg F, Posada JD, Reich CG, Rijnbeek PR, Schilling LM, Shah K, Shah NH, Subbian V, Zhang L, Zhu H, Ryan P, Prieto-Alhambra D, Kostka K, Duarte-Salles T.

*International Journal of Obesity*. 2021;45(11):2347-2357
Doi: 10.1038/s41366-021-00893-4.

International Journal of Obesity www.nature.com/ijo

ARTICLE OPEN

Check for updates

Epidemiology and Population Health

# Characteristics and outcomes of 627 044 COVID-19 patients living with and without obesity in the United States, Spain, and the United Kingdom

Martina Recalde[1,2,36], Elena Roel[1,2,36], Andrea Pistillo[1], Anthony G. Sena[3,4], Albert Prats-Uribe[5], Waheed-Ul-Rahman Ahmed[6,7], Heba Alghoul[8], Thamir M. Alshammari[9], Osaid Alser[10], Carlos Areia[11], Edward Burn[1,5], Paula Casajust[12], Dalia Dawoud[13], Scott L. DuVall[14,15], Thomas Falconer[16], Sergio Fernández-Bertolín[1], Asieh Golozar[17,18], Mengchun Gong[19], Lana Yin Hui Lai[20], Jennifer C. E. Lane[6], Kristine E. Lynch[14,15], Michael E. Matheny[21,22], Paras P. Mehta[23], Daniel R. Morales[24], Karthik Natarjan[16,25], Fredrik Nyberg[26], Jose D. Posada[27], Christian G. Reich[28], Peter R. Rijnbeek[4], Lisa M. Schilling[29], Karishma Shah[6], Nigam H. Shah[27], Vignesh Subbian[30], Lin Zhang[31,32], Hong Zhu[33,34], Patrick Ryan[3,16], Daniel Prieto-Alhambra[4,5], Kristin Kostka[28,35,37] and Talita Duarte-Salles[1,37]✉



**BACKGROUND:** A detailed characterization of patients with COVID-19 living with obesity has not yet been undertaken. We aimed to describe and compare the demographics, medical conditions, and outcomes of COVID-19 patients living with obesity (PLWO) to those of patients living without obesity.

**METHODS:** We conducted a cohort study based on outpatient/inpatient care and claims data from January to June 2020 from Spain, the UK, and the US. We used six databases standardized to the OMOP common data model. We defined two non-mutually exclusive cohorts of patients *diagnosed* and/or *hospitalized* with COVID-19; patients were followed from index date to 30 days or death. We report the frequency of demographics, prior medical conditions, and 30-days outcomes (hospitalization, events, and death) by obesity status.

**RESULTS:** We included 627 044 (Spain: 122 058, UK: 2336, and US: 502 650) *diagnosed* and 160 013 (Spain: 18 197, US: 141 816) *hospitalized* patients with COVID-19. The prevalence of obesity was higher among patients *hospitalized* (39.9%, 95%CI: 39.8−40.0) than among those *diagnosed* with COVID-19 (33.1%; 95%CI: 33.0−33.2). In both cohorts, PLWO were more often female. Hospitalized PLWO were younger than patients without obesity. Overall, COVID-19 PLWO were more likely to have prior medical conditions, present with cardiovascular and respiratory events during hospitalization, or require intensive services compared to COVID-19 patients without obesity.

**CONCLUSION:** We show that PLWO differ from patients without obesity in a wide range of medical conditions and present with more severe forms of COVID-19, with higher hospitalization rates and intensive services requirements. These findings can help guiding preventive strategies of COVID-19 infection and complications and generating hypotheses for causal inference studies.

*International Journal of Obesity*; https://doi.org/10.1038/s41366-021-00893-4

## INTRODUCTION

Obesity is associated with increased mortality and is a well-known risk factor for chronic conditions, such as diabetes, hypertension, cardiovascular disease, and cancer [1, 2]. Due to its proinflammatory state that impairs the immune response, obesity has also been related to an increased risk of viral infections [3]. The novel coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in December 2019 in Wuhan, China, and rapidly spread around the world [4]. This new virus causes a respiratory tract infection with clinical manifestations ranging from asymptomatic/mild symptoms to severe illness requiring intensive services. Partly due to its similarities with other viral infections such as seasonal influenza or H1N1, people with obesity were soon labeled as "at-risk" individuals [5]. Since obesity is a worldwide public health priority, granular information on patients with COVID-19 and obesity is needed to guide preventive strategies as well as to generate hypotheses for etiological studies [6].

A review and meta-analysis of 75 studies reported that obesity is a risk factor for testing positive for SARS-CoV-2, for severe COVID-19 and for COVID-19 related mortality [7]. While undoubtedly relevant to the field, these studies mainly focused on exploring multiple risk factors related to COVID-19 and thus did not offer a detailed characterization of patients with COVID-19

A full list of author affiliations appears at the end of the paper.

Received: 23 November 2020 Revised: 7 June 2021 Accepted: 24 June 2021
Published online: 15 July 2021

living with obesity. For instance, an exhaustive description of the medical conditions and COVID-19 related outcomes, such as thromboembolic events, among these patients is lacking. Other current limitations include the susceptibility to collider bias of studies reporting "risk factors" of COVID-19 infection and progression due to sampling mechanisms (e.g., subsamples of tested or hospitalized populations) [8]. A large characterization study focussing exclusively on patients with COVID-19 living with obesity using real-world data from different health settings and countries could address the limitations of the previous evidence.

In this study, we aimed to describe and compare the demographics, medical conditions, and outcomes of COVID-19 patients living with obesity (PLWO) to those of COVID-19 patients living without obesity, in inpatient or outpatient settings.

## METHODS

### Study design, setting, and data sources

We conducted a multinational cohort study using routinely collected healthcare data from January to June 2020 from Spain, the United Kingdom (UK), and the United States (US). This study was part of the "Characterizing Health Associated Risks, and Your Baseline Disease In SARS-COV-2 (CHARYBDIS)" study (protocol available for download at https://www.ohdsi.org/wp-content/uploads/2020/07/Protocol_COVID-19-Charybdis-Characterisation_V5.docx) designed by the Observational Health Data Sciences and Informatics (OHDSI) community. All data were standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [9]. The OHDSI network maintains the OMOP-CDM, along with a wide range of tools developed by its members to facilitate analyses of mapped data [10]. Data results for this study were extracted on July, 16th, 2020.

We included primary, outpatient and inpatient care data from electronic health records (EHRs) and health insurance claims data from six databases. Data from Spain included the Information System for Research in Primary Care (SIDIAP), which includes primary linked to inpatient care data covering approximately 80% of the population in Catalonia, Spain [11]. The UK data covered the Clinical Practice Research Datalink (CPRD), with patients from over 600 general practices in the UK [12]. Data from the US included: Columbia University Irving Medical Center (CUIMC), covering New York-Presbyterian Hospital and its affiliated physician practices; IQVIA Open Claims, which are pre-adjudicated claims collected from office-based physicians and specialists covering over 300 million lives (~80% of the US population); the Stanford Medicine Research Data Repository (STARR-OMOP), with data from Stanford Health Care [13], and the United States Department of Veterans Affairs (VA-OMOP), covering the national Department of Veterans Affairs health care system which serves more than 9 million enrolled Veterans (of whom 93% are male). A more detailed description of the included data sources is available in Supplementary Appendix 1.

### Study participants

We included two non-mutually exclusive cohorts of patients: (1) all patients *diagnosed* with COVID-19 (clinical diagnosis and/or positive test for SARS-CoV-2), and (2) all patients *hospitalized* with a COVID-19 diagnosis. We considered clinical diagnoses in the definition of COVID-19 cases due to testing restrictions during the first months of the pandemic (e.g., in Spain) [14]. The diagnostic codes used are described in Supplementary Appendix 2. Patients *hospitalized* with COVID-19 were identified as those having a hospitalization episode along with a clinical diagnosis or positive SARS-CoV-2 test within a time window from 21 days prior to admission up to the end of their hospitalization. We chose this time window to include patients with a diagnosis prior to hospitalization and to allow for a record delay in test results or diagnoses [15]. We included individuals with at least one year of observation time prior to the index date to capture observed baseline characteristics. In the *diagnosed* cohort, the index date was defined as the date of the COVID-19 clinical diagnosis or the earliest test day registered within seven days of a first positive test, whichever occurred first. In the *hospitalized* cohort, the index date was the day of hospitalization. Patients were followed from the index date to the earliest of death, end of the observation period, or 30 days [16].

Both the *diagnosed* and *hospitalized* COVID-19 cohorts were stratified by obesity status: PLWO vs patients living without obesity (from now on, referred to as patients without obesity). Obesity was defined as having an ever-recorded obesity diagnosis (Supplementary Appendix 3) and/or a body mass index (BMI) measurement between 30 and 60 kg/m$^2$ and/or a bodyweight measurement between 120 and 200 kg prior or at index date. We included upper cut-off thresholds to discard implausible observations. Patients without obesity were those who did not fulfill the obesity definition.

### Baseline characteristics and outcomes of interest

Demographics (sex and age) were obtained at the index date. More than 15 000 medical conditions from up to one year prior to the index date were identified based on the Systematized Nomenclature of Medicine (SNOMED) hierarchy, with all descendant codes included [15]. Specific definitions for comorbidities of particular interest were created; the detailed definitions of these variables can be consulted in Supplementary Appendix 3. We reported here a list of key comorbidities based on their prevalence in the cohorts of the participating sites, as well as on their clinical relevance to obesity and the COVID-19 research field [17].

Our 30-day outcomes of interest for the *diagnosed* cohort were hospitalization and fatality. For the *hospitalized* cohort, the 30-days outcomes were a requirement of intensive services (IS) (identified by a recorded mechanical ventilation and/or a tracheostomy and/or extra-corporeal membrane oxygenation procedure), respiratory, cardiovascular, thromboembolic, and other events and fatality.

### Data analysis

We described the number of patients included and the prevalence of obesity in each database as well as the demographics, comorbidities, and outcomes as proportions (calculated by the number of persons within a given category, divided by the total number of persons) with their respective 95% confidence intervals (CIs) for each database, by obesity status. To calculate these proportions in each database, we established a minimum count required (of five individuals), to minimize the risk of re-identification of patients. To compare medical conditions across groups, we calculated standardized mean differences (SMDs) [18], which we summarized in Manhattan-style plots. The SMD can be used to compare the prevalence of a dichotomous variable between two groups and is independent of sample size [19]. A $|SMD| > 0.1$ indicates a meaningful difference in the prevalence of a given condition; in the context of this study, a $SMD > 0.1$ indicates a higher prevalence in PLWO, whereas a $SMD < -0.1$ indicates a higher prevalence among patients without obesity. This study was descriptive by nature and, therefore, statistical modeling was out of scope. Differences across the groups compared should not be interpreted as causal effects.

To ensure data privacy at all times, we employed a federated analysis approach [16]. Following a pre-specified analysis plan, a common analytical code for the whole CHARYBDIS study was developed for the OHDSI Methods library, available at https://github.com/ohdsi-studies/Covid19CharacterizationCharybdis, and was run locally in each database. Individual-level data remained within host institutions and only aggregate results from each database were provided to the research team and publicly shared. All the results reported in this paper and additional data are available for consultation at a dynamic and interactive website, which changes over time as new databases are added and/or results are updated to CHARYBDIS (https://data.ohdsi.org/Covid19CharacterizationCharybdis/).

We used R version 3.6 for data visualization. All the data partners obtained Institutional Review Board (IRB) approval or exemption to conduct this descriptive study.

## RESULTS

### Prevalence of obesity

We included 627 044 *diagnosed* and 160 013 *hospitalized* patients with COVID-19 (Table 1). The *diagnosed* cohort consisted of 122 058 patients from Spain (SIDIAP), 2336 from the UK (CPRD), and 502 650 from the US (CUIMC: 8519; IQVIA-OpenClaims: 466 191; STARR-OMOP: 3328; VA-OMOP: 24 612). The *hospitalized* cohort included 18 197 patients from Spain (SIDIAP) and 141 816 from the US (CUIMC: 2600; IQVIA-OpenClaims: 133 091; STARR-OMOP: 615; VA-OMOP: 5510). Among *diagnosed* and *hospitalized* patients, 207 859 (33.1%; 95%CI: 33.0−33.2) and 63 866 (39.9%, 95%CI: 39.8−40.0) had obesity, respectively. In all databases, the prevalence

**Table 1.** Demographic characteristics of patients diagnosed and hospitalized with COVID-19 in each database, stratified by obesity status.

| | SIDIAP (Spain) | | CPRD (UK) | | CUIMC (US) | | IQVIA-Open Claims (US) | | STARR-OMOP (US) | | VA-OMOP (US) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | With obesity | Without obesity | With obesity | Without obesity | With obesity | Without obesity | With obesity | Without obesity | With obesity | Without obesity | With obesity | Without obesity |
| Diagnosed with COVID-19 | | | | | | | | | | | | |
| All, *n* | 36 409 | 85 649 | 976 | 1360 | 3446 | 5073 | 154 325 | 311 866 | 1157 | 2171 | 11 546 | 13 066 |
| All, %* (95%CI) | 29.8 (29.5–30.1) | 70.2 (69.9–70.5) | 41.8 (39.8–43.8) | 58.2 (56.2–60.2) | 40.5 (39.5–41.5) | 59.5 (58.5–60.5) | 33.1 (33.0–33.2) | 66.9 (66.8-67) | 34.8 (33.2–36.4) | 65.2 (63.6–66.8) | 46.9 (46.3–47.5) | 53.1 (52.5–53.7) |
| Laboratory confirmed, % (95%CI) | 39.2 (38.9–39.5) | 27.2 (27–27.4) | 51.5 (49.5–53.5) | 50.3 (48.3–52.3) | 66.7 (65.7–67.7) | 65.5 (64.5–66.5) | – | – | 24.4 (22.9–25.9) | 28.6 (27.1–30.1) | – | – |
| Female sex, % (95%CI) | 62.5 (62.2–62.8) | 55.9 (55.6–56.2) | 56.2 (54.2–58.2) | 57.2 (55.2–59.2) | 60.9 (59.9–61.9) | 56.4 (55.3–57.5) | 61.0 (60.9–61.1) | 51.7 (51.6-51.8) | 52.9 (51.2–54.6) | 52.8 (51.1–54.5) | 13.2 (12.8–13.6) | 19.0 (18.5–19.5) |
| *Age, % (95%CI)* | | | | | | | | | | | | |
| <18 years | 1.2 (1.1–1.3) | 4.7 (4.6–4.8) | | | 1.2 (1–1.4) | 3.5 (3.1–3.9) | 1.0 (0.97–1.03) | 3.9 (3.8–4) | | | | |
| 18–64 years | 58.9 (58.6–59.2) | 75.6 (75.4–75.8) | 55.4 (53.4–57.4) | 52.3 (50.3–54.3) | 65.2 (64.2–66.2) | 65.8 (64.8–66.8) | 66.7 (66.6–66.8) | 58.0 (57.9–58.1) | 66.2 (64.6–67.8) | 69.6 (68–71.2) | 53.7 (53.1–54.3) | 58.4 (57.8–59) |
| >65 years | 40.0 (39.7–40.3) | 19.6 (19.4–19.8) | 44.7 (42.7–46.7) | 47.7 (45.7–49.7) | 33.6 (32.6–34.6) | 30.6 (29.6–31.6) | 32.3 (32.2–32.4) | 38.2 (38.1–38.3) | 33.8 (32.2–35.4) | 30.4 (28.8–32) | 46.3 (45.7–46.9) | 41.6 (41–42.2) |
| Hospitalized with COVID-19 | | | | | | | | | | | | |
| All, *n* | 8403 | 9794 | – | – | 1408 | 1192 | 50863 | 82228 | 274 | 341 | 2918 | 2592 |
| All, % (95%CI) | 46.2 (45.5–46.9) | 53.8 (53.1–54.5) | – | – | 54.2 (52.3–56.1) | 45.8 (43.9–47.7) | 38.2 (37.9–38.5) | 61.8 (61.5–62.1) | 44.6 (40.7–48.5) | 55.4 (51.5–59.3) | 53.0 (51.7–54.3) | 47.0 (45.7–48.3) |
| Laboratory confirmed, % (95%CI) | 76.0 (75.4–76.6) | 72.9 (72.3–73.5) | – | – | 88.5 (87.3–89.7) | 91.6 (90.5–92.7) | | | 10.6 (8.2–13) | 18.3 (15.2–21.4) | | |
| Female sex, % (95%CI) | 51.2 (50.5–51.9) | 39.9 (39.2–40.6) | – | – | 54.7 (52.8–56.6) | 39.8 (37.9–41.7) | 55.0 (54.7–55.3) | 43.5 (43.2–43.8) | 51.5 (47.6–55.4) | 49.0 (45.0–53.0) | 6.6 (5.9–7.3) | 4.4 (3.9–4.9) |
| *Age, % (95%CI)* | | | | | | | | | | | | |
| <18 years | – | – | – | – | – | – | 0.5 (0.5–0.5) | 2.0 (1.9–2.1) | – | – | – | – |
| 18–64 years | 37.9 (37.2–38.6) | 50.6 (49.9–51.3) | – | – | 53.3 (51.4–55.2) | 34.5 (32.7–36.3) | 51.7 (51.4–52) | 38.5 (38.2–38.8) | 64.2 (60.4–68) | 57.2 (53.3–61.1) | 37.1 (35.8–38.4) | 27.4 (26.2-28.6) |
| >65 years | 62.1 (61.4–62.8) | 49.4 (48.7–50.1) | – | – | 46.7 (44.8–48.6) | 65.6 (63.8–67.4) | 47.8 (47.5–48.1) | 59.5 (59.2–59.8) | 35.8 (32.0–39.6) | 42.8 (38.9–46.7) | 62.9 (61.6–64.2) | 72.6 (71.4–73.8) |

*Notes:* * Proportion of patients with and without obesity among all patients (row percentage); - data not available or below the minimum cell count required (five individuals)

Abbreviations: *CI* confidence interval, *COVID-19* coronavirus disease 2019, *CPRD* Clinical Practice Research Datalink, *CUIMC* Columbia University Irving Medical Center, *SIDIAP* Information System for Research in Primary Care, *STARR-OMOP* Stanford Medicine Research Data Repository, *UK* United Kingdom, *US* United States, *VA-OMOP* United States Department of Veterans Affairs.

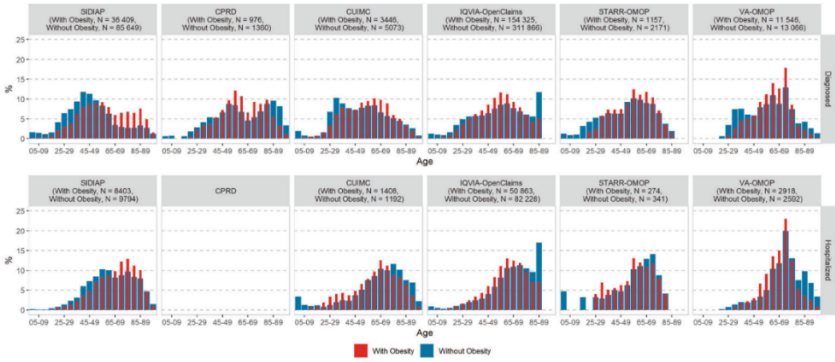



**Fig. 1 Distribution of age among patients living with and without obesity in each database, stratified by COVID-19 cohort type (diagnosed and hospitalized).** CPRD Clinical Practice Research Datalink, COVID-19 coronavirus disease 2019, CUIMC Columbia University Irving Medical Center, SIDIAP Information System for Research in Primary Care, STARR-OMOP Stanford Medicine Research Data Repository, VA-OMOP United States Department of Veterans Affairs.

of obesity was lower among *diagnosed* patients than among those *hospitalized*, with differences ranging from 5 (IQVIA-OpenClaims) to 16% (SIDIAP).

### Baseline demographics

The sex distribution (proportions and 95% CIs) of the patients are reported in Table 1. Aside from VA-OMOP, in the *diagnosed* cohort, patients with and without obesity were mostly female. The proportion of females was higher among PLWO compared to patients without obesity in SIDIAP (63% vs 56%), CUIMC (61% vs 56%), and IQVIA-OpenClaims (61% vs 52%), while in VA-OMOP the opposite was observed (13% vs 19%). No differences were observed in CPRD and STARR-OMOP. In the *hospitalized* cohort, patients without obesity were predominantly male (female ranged from 40 to 49%, VA-OMOP: 4%) but PLWO still were more commonly female in all databases aside from VA-OMOP (range: 51−55%, VA-OMOP: 7%). Differences in the proportion of females between PLWO and patients without obesity ranged from 3 (VA-OMOP) to 15% (CUIMC).

The age distribution in each database is summarized in Table 1 with proportions and their respective 95% CIs and in Fig. 1 with histograms. In the *diagnosed* cohort, PLWO were slightly older than those without obesity (i.e., the age distribution for PLWO was slightly skewed to the left compared to patients without obesity). This was particularly marked in SIDIAP, where 40% of the PLWO were aged above 65 years and only 20% were so without obesity. *Hospitalized* patients were older than those *diagnosed*. In the *hospitalized* cohorts, PLWO were fairly consistently younger than those without obesity (except for SIDIAP). The proportion of patients aged above 65 ranged from 36 to 63% for PLWO and from 43 to 73% for those without obesity.

### Baseline medical conditions

We compared baseline medical conditions of PLWO to those of patients without obesity in the *diagnosed* and *hospitalized* cohorts using SMDs, which are summarized in Fig. 2. We depicted the SMDs of 485 (CPRD) to 5050 (VA-OMOP) medical conditions in the *diagnosed* cohort, and 529 (STARR-OMOP) to 5240 (IQVIA-OpenClaims) in the *hospitalized* cohort. In both cohorts, medical conditions were largely more frequent among PLWO than patients without obesity.

The distribution of the selected key comorbidities is shown in Fig. 3, and the proportions with their respective 95% CIs and SMDs between PLWO and patients without obesity are available in Supplementary Appendices 4 and 5. In the *diagnosed* cohorts, PLWO consistently had a higher prevalence of comorbidities compared to those without obesity; these differences were meaningful (i.e., with a SMD > 0.1, which indicates a meaningfully higher prevalence among PLWO) for the majority of comorbidities across databases. For example, while the prevalence of hypertension for PLWO ranged from 30 to 32% in Europe (SIDIAP and CPRD) and from 55 to 81% in the US, in those without obesity it ranged from 12 to 16% and from 26 to 53%, respectively. The SMD for hypertension was above 0.1 in all databases. As in the *diagnosed* cohort, PLWO *hospitalized* with COVID-19 had a higher prevalence of comorbidities than those without obesity, and these differences were meaningful for the majority of comorbidities. However, the differences between groups were less obvious. For example, heart disease differed by 20% among those *diagnosed* in VA-OMOP (PLWO: 60%, without obesity: 40%) and by 9% among those *hospitalized* (PLWO: 74%, without obesity: 65%); although the SMD was still above 0.1 in all databases.

### 30-day outcomes of interest

The distribution of 30-days outcomes is shown in Fig. 4, the proportions with their respective 95% CI and SMDs between PLWO and patients without obesity are available in Table 2. In the *diagnosed* cohorts, hospitalization rates were higher among PLWO than among those without obesity in all databases. For example, in SIDIAP the proportion of patients hospitalized was 20% for PLWO and 10% for patients without obesity. However, these differences were meaningful (SMD > 0.1) only in three databases: SIDIAP, CUIMC, and STARR-OMOP. In PLWO, fatality ranged from 5 to 12% and was higher than in patients without obesity in SIDIAP and CUIMC (7% vs 3% and 8% vs 5%, respectively), while in CPRD and VA-OMOP it was similar in both groups. SIDIAP was the only database with a meaningful difference in the proportion of fatality.

Overall, in the *hospitalized* cohort, PLWO more frequently had adverse events occurring in the 30 days after the index date than patients without obesity. For example, PLWO required IS and presented with ARDS more frequently than patients without obesity in the largest databases: IQVIA-OpenClaims (IS: 13% vs 10%; ARDS: 35% vs 31%) and VA-OMOP (IS: 22% vs 15%; 46% vs 41%), whereas in CUIMC and STARR-OMOP percentages were similar. VA-OMOP was the only database with a meaningful

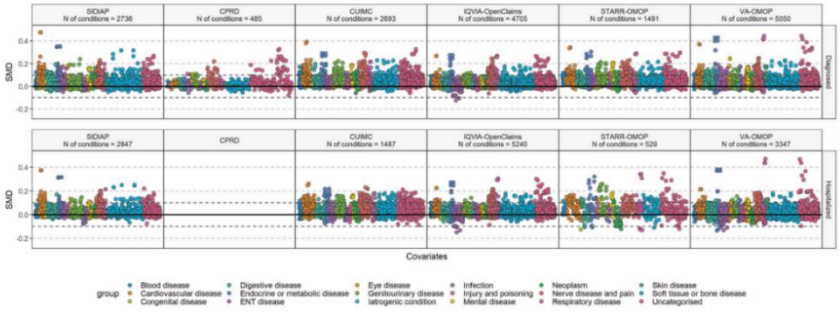



**Fig. 2 Standardized mean differences in conditions among patients living with obesity compared to patients living without obesity in each database, stratified by COVID-19 cohort type (diagnosed and hospitalized).** SMD < 0 means the prevalence was greater in COVID-19 patients living without obesity, SMD > 0 means the prevalence was greater in COVID-19 patients living with obesity. COVID-19 coronavirus disease 2019, CPRD Clinical Practice Research Datalink, CUIMC Columbia University Irving Medical Center, SIDIAP Information System for Research in Primary Care, SMD standardized mean difference, STARR-OMOP Stanford Medicine Research Data Repository, VA-OMOP United States Department of Veterans Affairs.

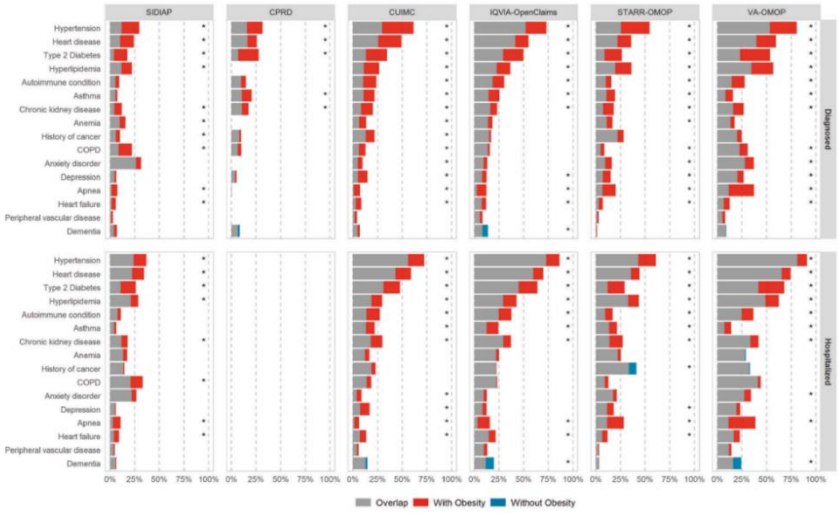



**Fig. 3 Comorbidities at baseline among patients living with obesity compared to patients living without obesity in each database, stratified by COVID-19 cohort type (diagnosed and hospitalized).** Prevalence of comorbidities for COVID-19 patients living with obesity (red) and without obesity (blue) are depicted in overlapped horizontal bars. The gray color is the overlap between groups. E.g., in CPRD, 32% of COVID-19 patients living with obesity and 16% living without obesity have hypertension. Comorbidities with a meaningful difference (|SMD| > 0.1) between patients living with and without obesity are marked with an asterisk (*). COPD chronic obstructive pulmonary disease, COVID-19 coronavirus disease 2019, CPRD Clinical Practice Research Datalink, CUIMC Columbia University Irving Medical Center, SIDIAP Information System for Research in Primary Care, SMD standardized mean difference, STARR-OMOP Stanford Medicine Research Data Repository, VA-OMOP United States Department of Veterans Affairs.

difference in the proportion of IS. Similarly, heart failure was also more frequent among PLWO than among patients without obesity in CUIMC: 7% vs 3%, IQVIA-OpenClaims: 7% vs 5%, STARR-OMOP: 16% vs 9%, and VA-OMOP: 23% vs 17%), these differences were meaningful in CUIMC and STARR-OMOP. Sepsis, cardiac arrhythmia, and cardiovascular disease events were slightly more frequent among PLWO, although SMDs were below 0.1 in all databases. Acute kidney injury was the only outcome that was more frequent among patients without obesity; however, this difference was not meaningful in any database.

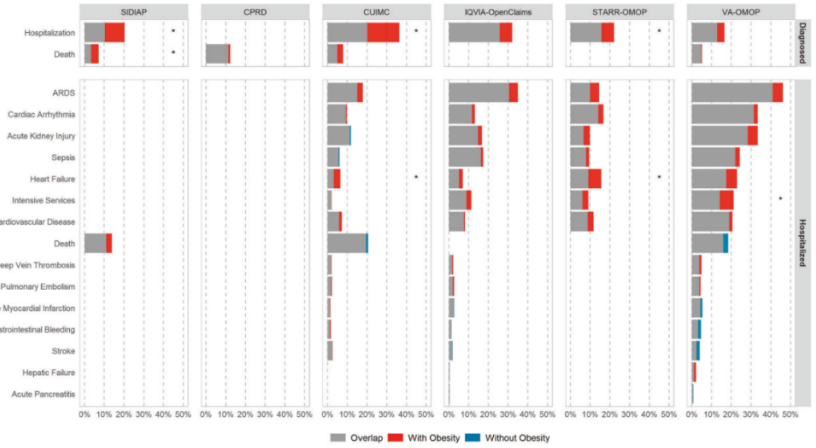



**Fig. 4 A comparison of 30-day events among patients living with and without obesity in each database, by COVID-19 cohort type (diagnosed and hospitalized).** Proportion of outcomes for COVID-19 patients living with obesity (red) and without obesity (blue) are depicted in overlapped horizontal bars. The gray color is the overlap between groups. E.g., in the diagnosed cohort, 20 and 10% of patients living with and without obesity in SIDIAP, respectively, were hospitalized. Outcomes with a meaningful difference (|SMD| > 0.1) between patients living with obesity and patients without obesity are marked with an asterisk (*). ARDS acute respiratory distress syndrome, COVID-19 coronavirus disease 2019, CPRD Clinical Practice Research Datalink, CUIMC Columbia University Irving Medical Center, SIDIAP Information System for Research in Primary Care, SMD standardized mean difference, STARR-OMOP Stanford Medicine Research Data Repository, VA-OMOP United States Department of Veterans Affairs.

As for fatality, there were no consistent nor meaningful differences between PLWO and patients without obesity in the *hospitalized cohort:* while it was higher for PLWO in SIDIAP (14% vs 11%), there were no differences in CUIMC (20% vs 21%) nor in VA-OMOP (16% vs 18%).

## DISCUSSION

In this large cohort study including 627 044 COVID-19 patients from Spain, the UK, and the US, we found that the prevalence of obesity was higher among COVID-19 patients *hospitalized* (40%) compared to those *diagnosed* (31%). PLWO *diagnosed* and *hospitalized* with COVID-19 were more commonly female, and those hospitalized were younger than patients without obesity. The extraction of more than 15 000 medical conditions revealed PLWO were not only more prone to have obesity-related comorbidities, such as hypertension, heart disease, and type 2 diabetes but also to more than a thousand different health conditions. After 30-days of follow-up, PLWO presented with higher hospitalization rates and intensive services requirements, although these differences were only meaningful in some databases.

Our study has several strengths, such as its large amount of data. By bringing together harmonized data using a federated approach, we have conducted a large-scale study while respecting the confidentiality of patient records. The international approach of this study is a strong asset given that we are investigating the intersection of two major global threats, namely the obesity epidemic and the COVID-19 pandemic. The former, together with the diverse healthcare settings and populations described in this study, increase the generalizability of our findings. Further, we provide a wide overview of the characteristics and outcomes of patients with and without obesity, using data visualization tools to summarize large amounts of medical data. This exhaustive characterization goes far beyond prior studies reporting few comorbidities and supports the generation of new hypotheses that can be tested in future studies. In addition, for the sake of transparency and reproducibility, we have made methods, tools, and all results publicly available. As CHARYBDIS is an ongoing study, results (included longer follow-up time) will be updated and new studies focussing on obesity could be conducted. All of the above has been accomplished through the coordinated efforts of the OHDSI community to provide a rapid response to the COVID-19 pandemic.

Our study also has limitations. First, we cannot exclude a selection bias of COVID-19 cases due to underreporting in the context of testing restrictions and asymptomatic or paucisymptomatic cases that usually do not seek medical care. Additionally, testing policies have varied across countries and time depending on the course of the pandemic. Nevertheless, the inclusion of patients clinically diagnosed (not tested) in different settings likely provided consistency to our data, although it might have incurred in false positives. Second, we did not have information on BMI as a continuous variable, which prevented us from investigating the impact of different categories of obesity in COVID-19 outcomes. This might explain the higher proportion of comorbidities and outcomes observed in the US databases, as PLWO from the US might have higher BMIs than those from Europe [20]. In addition, our definition of obesity included diagnoses and measurements recorded at any time prior to or at the index date, and therefore some individuals might have been misclassified due to changes in BMI since the most recently recorded status. However, previous evidence shows that BMI trajectories in adults are relatively stable, with a tendency to increase with age [21]. Therefore adults with obesity are likely to still have obesity over time. Finally, this study was underpinned by routinely collected data which can raise concerns about the quality of the data. Some databases are prone to oversampling certain groups of people as a result of how these

**Table 2.** Occurrence of 30-day events, in % (95%CI), among patients living with and without obesity in each database, by COVID-19 cohort type (diagnosed and hospitalized).

| | SIDIAP (Spain) | | | CPRD (UK) | | | CUIMC (US) | | | IQVIA-Open Claims (US) | | | STARR-OMOP (US) | | | VA-OMOP (US) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD |
| *Patients diagnosed, n* | 36 409 | 854 649 | NA | 976 | 1360 | NA | 3446 | 5073 | NA | 154 325 | 311 866 | NA | 1157 | 2171 | NA | 11 546 | 13 066 | NA |
| Hospitalization | 20.3 (19.9 –20.7) | 10.3 (10.2 –10.4) | 0.20 | – | – | – | 36.4 (34.8 –38.0) | 20.2 (19.1 –21.3) | 0.26 | 32.1 (31.9 –32.3) | 25.8 (25.6 –26.0) | 0.10 | 22.1 (19.7 –24.5) | 15.8 (14.3 –17.3) | 0.11 | 16.6 (15.9 –17.3) | 12.8 (12.2 –13.4) | 0.08 |
| Death | 7.1 (6.8 –7.4) | 3.3 (3.3 –3.3) | 0.12 | 12.3 (10.2 –14.4) | 11.3 (9.6 –13.0) | 0.02 | 8.0 (7.1 –8.9) | 5.0 (4.4 –5.6) | 0.09 | – | – | – | – | – | – | 5.1 (4.7 –5.5) | 4.8 (4.4 –5.2) | 0.01 |
| *Patients hospitalized, n* | 8403 | 9794 | NA | – | – | NA | 1408 | 1192 | NA | 50 863 | 8228 | NA | 274 | 341 | NA | 2918 | 2592 | NA |
| Intensive service requirement | – | – | – | – | – | – | 2.3 (1.5 –3.1) | 2.0 (1.2 –2.8) | 0.01 | 13.0 (12.7 –13.3) | 9.7 (9.1 –10.3) | 0.06 | 9.1 (5.7 –12.5) | 6.2 (3.6 –8.8) | 0.08 | 21.9 (20.4 –23.4) | 14.5 (13.1 –15.9) | 0.13 |
| Death | 13.9 (13.2 –14.6) | 10.9 (10.3 –11.5) | 0.06 | – | – | – | 19.5 (17.4 –21.6) | 20.6 (18.3 –22.9) | –0.02 | – | – | – | – | – | – | 15.9 (14.6 –17.2) | 18.4 (16.9 –19.9) | –0.05 |
| *Cardiovascular events* | | | | | | | | | | | | | | | | | | |
| Acute myocardial infarction | – | – | – | – | – | – | 1.3 (0.7 –1.9) | 1.1 (0.5 –1.7) | 0.01 | 2.1 (2.0 –2.2) | 2.2 (1.9 –2.5) | –0.01 | – | – | – | 5.5 (4.7 –6.3) | 6.9 (5.9 –7.9) | –0.03 |
| Cardiac arrhythmia* | – | – | – | – | – | – | 9.8 (8.2 –11.4) | 9.3 (7.7 –10.9) | 0.01 | 13.1 (12.8 –13.4) | 11.7 (11.0 –12.4) | 0.03 | 16.8 (12.4 –21.2) | 14.1 (10.4 –17.8) | 0.05 | 33.4 (31.7 –35.1) | 31.4 (29.6 –33.2) | 0.03 |
| Heart failure* | – | – | – | – | – | – | 6.5 (5.2 –7.8) | 3.1 (2.1 –4.1) | 0.11 | 7.0 (6.8 –7.2) | 5.1 (4.6 –5.6) | 0.06 | 15.7 (11.4 –20.0) | 9.1 (6.0– | 0.14 | 22.9 (21.4 –24.4) | 17.4 (15.9 –18.9) | 0.10 |
| Stroke | – | – | – | – | – | – | 2.6 (1.8 –3.4) | 2.5 (1.6 –3.4) | 0.00 | 1.7 (1.6 –1.8) | 1.9 (1.6 –2.2) | –0.01 | – | – | – | 2.4 (1.8 –3.0) | 4.0 (3.2 –4.8) | –0.06 |
| Total cardiovascular disease events | – | – | – | – | – | – | 7.3 (5.9 –8.7) | 5.9 (4.6 –7.2) | 0.04 | 8.2(8.0 –8.4) | 7.6 (7.0 –8.2) | 0.02 | 11.7 (7.9 –15.5) | 8.8 (5.8 –11.8) | 0.07 | 20.6 (19.1 –22.1) | 19.0 (17.5 –20.5) | 0.03 |
| *Thromboembolic events* | | | | | | | | | | | | | | | | | | |
| Deep vein thrombosis | – | – | – | – | – | – | 2.1 (1.4 –2.8) | 1.8 (1.0 –2.6) | 0.02 | 2.2 (2.1 –2.3) | 1.7 (1.4 –2.0) | 0.03 | – | – | – | 4.9 (4.1 –5.7) | 3.8 (3.1 –4.5) | 0.04 |
| Pulmonary embolism | – | – | – | – | – | – | 2.2 (1.4 –3.0) | 1.8 (1.0 –2.6) | 0.02 | 1.8 (1.7 –1.9) | 1.5 (1.2 –1.8) | 0.03 | – | – | – | 3.8 (3.1 –4.5) | 3.3 (2.6 –4.0) | 0.02 |
| *Other events* | | | | | | | | | | | | | | | | | | |
| Acute kidney injury* | – | – | – | – | – | – | 18.1 (16.1 –20.1) | 25.4 (22.9 –27.9) | –0.01 | 11.1 (10.8 –11.4) | 10.4 (9.7 –11.1) | 0.04 | 8.4 (5.1 –11.7) | 9.1 (6.0 –12.2) | 0.08 | 17.1 (15.7 –18.5) | 18.6 (17.1 –20.1) | 0.08 |
| Acute pancreatitis | – | – | – | – | – | – | – | – | – | 0.2 (0.2 –0.2) | 0.2 (0.1 –0.3) | 0.00 | – | – | – | 0.4 (0.2 –0.6) | 0.7 (0.4 –1.0) | –0.02 |
| ARDS* | – | – | – | – | – | | 17.9 (15.9 –19.9) | 15.2 (13.2 –17.2) | 0.05 | 34.9 (34.5 –35.3) | 30.5 (29.5 –31.5) | 0.07 | 14.6 (10.4 –18.8) | 10.0 (6.8 –13.2) | 0.10 | 46.2 (44.4 –48.0) | 41.0 (39.1 –42.9) | 0.07 |
| | – | – | – | – | – | – | | | 0.02 | | | 0.00 | – | – | – | | | –0.06 |

**Table 2** continued

| | SIDIAP (Spain) | | | CPRD (UK) | | | CUIMC (US) | | | IQVIA-Open Claims (US) | | | STARR-OMOP (US) | | | VA-OMOP (US) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD | With obesity | Without obesity | SMD |
| Gastrointestinal bleeding | | | | | | | 1.7 (1.0 −2.4) | 1.0 (0.4 −1.6) | | 1.3 (1.2- −1.4) | 1.3 (1.1 −1.5) | | | | | 3.1 (2.5 −3.7) | 4.7 (3.9 −5.5) | |
| Hepatic failure | – | – | – | – | – | – | 0.8 (0.3 −1.3) | – | – | 0.2 (0.2 −0.2) | 0.2 (0.1 −0.3) | 0.00 | – | – | – | 1.6 (1.1 −2.1) | 0.8 (0.5 −1.1) | 0.06 |
| Sepsis* | – | – | – | – | – | – | 5.6 (4.4 −6.8) | 6.0 (4.7 −7.3) | −0.02 | 17.4 (17.1 −17.7) | 16.2 (15.4 −17.0) | 0.02 | 9.5 (6.0 −13.0) | 7.9 (5.0 −10.8) | 0.04 | 24.3 (22.7 −25.9) | 22.0 (20.4 −23.6) | 0.04 |

*Notes.* - data not available or below the minimum cell count required (five individuals); events marked with an * were recorded only during hospitalization. SMD < 0 means the prevalence was greater in COVID-19 patients living without obesity, SMD > 0 means the prevalence was greater in COVID-19 patients living with obesity; A SMD > |0.1| indicates a meaningful difference in the prevalence of a given condition
Abbreviations: *ARDS* acute respiratory distress syndrome, *CI* confidence interval, *COVID-19* coronavirus disease 2019, *CPRD* Clinical Practice Research Datalink, *CUIMC* Columbia University Irving Medical Center, *NA* not applicable, *SIDIAP* Information System for Research in Primary Care, *SMD* standardized mean difference, *STARR-OMOP* Stanford Medicine Research Data Repository, *UK* United Kingdom, *US* United States, *VA-OMOP* United States Department of Veterans Affairs.

data are captured (e.g., the Veterans Affairs system historically serves more men than women, routine claims data may only reflect health outcomes in commercially insured populations, etc.). Obesity, comorbidities, and outcomes were assessed based on having a record of a condition/measurement, therefore they may be underestimated. In addition, outcomes such as hospitalization or intensive services requirements are also influenced by factors external to the patient's condition (i.e., bed availability, criteria for admission), which might differ across databases. Even still, the consistency of our findings across databases that differ by setting and country lends credence to the generalizability of our findings.

Given the prevalence of obesity in Spain (24%), the UK (27%), and the US (37%), a high proportion of PLWO among COVID-19 cases was expected [20]. However, the prevalence of obesity among diagnosed COVID-19 patients was higher than the general population in four databases: SIDIAP (Spain): 30%; CPRD (UK): 42%; CUIMC and VA-OMOP (US): 41 and 47%, respectively, which is suggestive of an increased risk of diagnosis in PLWO. In addition, the prevalence of obesity was higher in hospitalized COVID-19 patients, with an overall prevalence of obesity of 40%, which is in line with three cohort studies from the US that reported that 40, 42, and 48% of inpatients were living with obesity [17, 22, 23]. A large meta-analysis of observational studies reported that obesity is associated with a higher risk of testing positive for SARS-CoV-2 or being diagnosed with COVID-19 as well as of being hospitalized with COVID-19 [7]. While this could be due to an increased vulnerability to SARS-CoV-2 in PLWO, other hypotheses should be considered in future studies. On the one hand, individuals with obesity could be more likely to seek care and be tested for SARS-CoV-2 since they are (presumably) a high-risk population, have multiple comorbidities, and are more prone to respiratory symptoms due to their compromised pulmonary function [2, 7]. On the other hand, given the fact that obesity disproportionately affects disadvantaged populations, potential differential exposures across subpopulation groups should also be explored (e.g., differential occupational risks) [2].

Women predominated among hospitalized patients with obesity, even though obesity rates are similar in both sexes in the three countries [20]. Although male sex is a well-established risk factor for COVID-19-related hospitalization and death, little is known about the role of obesity on COVID-19 outcomes stratified by sex [14, 23–25]. Recent studies addressing this issue in secondary analyses have reported inconsistent results. A study conducted among UK Biobank participants found that the impact of BMI in COVID-19-related death was higher among females compared to men, while others have found a higher effect among males, opposite effects of sex in different age strata or null differences [26–29]. Thus, the intersection between sex/gender and obesity in relation to COVID-19-outcomes warrants further investigation. Because sex-stratification was beyond the pre-specified analysis plan of our study, we were unable to report our results by sex, which could have provided valuable insights on the matter. We intend, however, to address this issue in upcoming studies from the CHARYBDIS project.

We also found that *hospitalized* PLWO were younger than those without obesity. Although younger individuals have less risk of infections and complications than older people due to having fewer comorbidities and a stronger immune system, this is not the case for those with obesity [2, 7, 30–32]. Some authors have postulated that PLWO younger than 60 years could have a greater risk of severe COVID-19 outcomes [33]. PLWO also had many more comorbidities than patients without obesity. Unsurprisingly, the highest differences were observed in obesity-related conditions, such as hypertension, diabetes, and heart disease, which have been identified as risk factors for severe COVID-19 outcomes [14, 17, 25, 34, 35]. However, as our findings revealed, PLWO with COVID-19 differ from patients without obesity in a wider range of medical conditions than previously described. Future etiological

studies aiming to disentangle the effect of obesity in COVID-19 outcomes should have this information present and consider data-driven techniques to account for confounding, such as propensity score estimation and its adjustment methods [18].

Finally, PLWO experienced adverse events more frequently than those without obesity, particularly hospitalization and the requirement of intensive services. Certainly, our results must be interpreted carefully considering the differences in demographics and comorbidities between these groups. Interestingly, in patients hospitalized, we did not observe clear differences in fatality between patients with and without obesity. While two meta-analyses reported that obesity is associated with a higher risk of COVID-19 related mortality; other large observational studies from the US and the UK using finer categories of BMI only found an association with mortality for morbid obesity (BMIs ≥ 35 $kg/m_2$ or ≥40 $kg/m_2$) [7, 25, 28, 29, 34]. Given the scarcity of evidence regarding the frequency of specific adverse events during hospitalization among PLWO, our findings are of special interest to the field and should be addressed in upcoming etiological studies.

In this large international cohort, we showed that among COVID-19 cases, PLWO were more likely to be female, have more comorbidities, and worse outcomes than patients without obesity. The prevalence of obesity was higher among hospitalized patients with COVID-19 compared to patients diagnosed with COVID-19. Our results may be useful in guiding clinical practice and aid future preventative strategies for patients living with obesity, as well as providing useful data to support subsequent etiological studies focussed on obesity and COVID-19.

## REFERENCES

1. Di Angelantonio E, Bhupathiraju SN, Wormser D, Gao P, Kaptoge S, de Gonzalez AB, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet. 2016;388:776–86.
2. Hu FB. Obesity Epidemiology. New York: Oxford University Press, 2009.
3. Moser JAS, Galindo-Fraga A, Ortiz-Hernández AA, Gu W, Hunsberger S, Galán-Herrera JF. et al.Underweight, overweight, and obesity as independent risk factors for hospitalization in adults and children from influenza and other respiratory viruses.Influenza Other Respir Viruses.2019;13:3–9.
4. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020. WHO. 2020. https://www.who.int/directorgeneral/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-oncovid-19—11-march-2020.
5. Louie JK, Acosta M, Samuel MC, Schechter R, Vugia DJ, Harriman K, et al. A novel risk factor for a novel virus: obesity and 2009 pandemic influenza a (H1N1). Clin Infect Dis. 2011;52:301–12.
6. GOV.UK. New obesity strategy unveiled as country urged to lose weight to beat coronavirus (COVID-19) and protect the NHS. GOV.UK. 2020. https://www.gov.uk/government/news/new-obesity-strategy-unveiled-as-country-urged-to-lose-weight-to-beat-coronavirus-covid-19-and-protect-the-nhs.
7. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. Obes Rev. 2020;21:e13128.
8. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun. 2020;11:1–12.
9. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Informatics Assoc. 2015;22:553–64.
10. 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.
11. 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. Med Clin. 2012;138:617–21.
12. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa Tvan, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44:827–36.
13. Datta S, Posada J, Olson G, Li W, O'Reilly C, Balraj D, et al. A new paradigm for accelerating clinical data science at Stanford Medicine. arXiv:2003.10534v1 [Preprint]. 2020 [cited 2020 Aug 20]. Available from: https://arxiv.org/abs/2003.10534v1
14. Burn E, Tebé C, Fernandez-Bertolin S, Aragon M, Recalde M, Roel E, et al. The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nat Commun. 2021;12:1–12.
15. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahão MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun. 2020;11:5009.
16. Observational Health Data Sciences and Informatics. The Book of OHDSI (Independently published). 2019. https://ohdsi.github.io/TheBookOfOhdsi/.
17. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020;323:2052–9.
18. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46:399–424.
19. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput. 2009;38:1228–34.
20. World Health Organization. WHO Overweight and obesity. World Health Organization. 2020. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors.
21. Kelly SP, Lennon H, Sperrin M, Matthews C, Freedman ND, Albanes D, et al. Body mass index trajectories across adulthood and smoking in relation to prostate cancer risks: the NIH-AARP diet and health study. Int J Epidemiol. 2019;48:464–73.
22. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 states, March 1–30, 2020. MMWR Morb Mortal Wkly Rep. 2020;69:458–64.
23. Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ. 2020;369:m1966.
24. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ. 2020;369:m1985.
25. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, et al. Factors associated with death in critically Ill patients with coronavirus disease 2019 in the US. JAMA Intern Med. 2020;180:1436–47.
26. Anderson MR, Geleris J, Anderson DR, Zucker J, Nobel YR, Freedberg D, et al. Body mass index and risk for intubation or death in SARS-CoV-2 infection: a retrospective cohort study. Ann Intern Med. 2020;173:782–90.
27. Sattar N, Ho FK, Gill JM, Ghouri N, Gray SR, Celis-Morales CA, et al. BMI and future risk for COVID-19 infection and death across sex, age and ethnicity: preliminary findings from UK biobank. Diabetes Metab Syndr Clin Res Rev. 2020;14:1149–51.
28. Tartof SY, Qian L, Hong V, Wei R, Nadjafi RF, Fischer H, et al. Obesity and mortality among patients diagnosed with COVID-19: results from an integrated health care organization. Ann Intern Med. 2020;173:773–81.
29. Du Y, Lv Y, Zha W, Zhou N, Hong X. Association of body mass index (BMI) with critical COVID-19 and in-hospital mortality: a dose-response meta-analysis. Metabolism. 2021;117:154373.
30. Divo MJ, Martinez CH, Mannino DM. Ageing and the epidemiology of multimorbidity. Eur Respir J. 2014;44:1055–68.
31. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proc R Soc B Biol Sci. 2015;282:20143085.
32. Muscogiuri G, Pugliese G, Laudisio D, Castellucci B, Barrea L, Savastano S, et al. The impact of obesity on immune response to infection: plausible mechanisms and outcomes. Obes Rev. 2021;22:e13216.
33. Lighter J, Phillips M, Hochman S, Sterling S, Johnson D, Francois F, et al. Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission. Clin Infect Dis. 2020;71:896–7.
34. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584:430–6.
35. Reilev M, Kristensen KB, Pottegård A, Lund LC, Hallas J, Ernst MT, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. Int J Epidemiol. 2020;49:1468–81.

## ACKNOWLEDGEMENTS

We would like to acknowledge the patients who suffered from or died of this devastating disease, and their families and carers. We would also like to thank the

healthcare professionals involved in the management of COVID-19 during these challenging times, from primary care to intensive care units.

## AUTHOR CONTRIBUTIONS

MR, ER, APU, PR, DPA, KK, and TDS conceived and designed the study. SLD, TF, KEL, MEM, KN, JDP, CGR, NHS, PR, KK, and TDS coordinated data contributions at their respective sites. AP, AGS, TF, SFB, JDP, KK, and TDS analyzed the data; MR, ER, and AP produced the figures and tables. MR, ER, EB, DRM, FN, PR, LMS, DPA, KK, and TDS interpreted the data. MR, ER, and TDS searched the literature and wrote the first draft with insightful contributions from EB, LYHL, JCEL, DRM, FN, PR, LMS, DPA, and KK. All authors contributed to the revision of the first draft, reviewed and approved the final version of the paper.

## FUNDING

The European Health Data & Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. This research received partial support from the National Institute for Health Research (NIHR) Oxford Biomedical Research Center (BRC), US National Institutes of Health, US Department of Veterans Affairs, Janssen Research & Development, and IQVIA. The University of Oxford received funding related to this work from the Bill & Melinda Gates Foundation (Investment ID INV-016201 and INV-019257). APU has received funding from the Medical Research Council (MRC) [MR/K501256/1, MR/N013468/1] and Fundación Alfonso Martin Escudero (FAME) (APU). VINCI [VA HSR RES 13-457] (SLD, MEM, KEL). JCEL has received funding from the Medical Research Council (MR/K501256/1) and Versus Arthritis (21605). MR is funded by Wereld Kanker Onderzoek Fonds (WKOF), as part of the World Cancer Research Fund International grant program [grant number: 2017/1630]. No funders had a direct role in this study. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Clinician Scientist Award program, NIHR, Department of Veterans Affairs or the United States Government, NHS, or the Department of Health, England.

## COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Mr. Sena reports personal fees from Janssen Research & Development, outside the submitted work; Dr. DuVall reports grants from Anolinx, LLC, grants from Astellas Pharma, Inc, grants from AstraZeneca Pharmaceuticals LP, grants from Boehringer Ingelheim International GmbH, grants from Celgene Corporation, grants from Eli Lilly and Company, grants from Genentech Inc., grants from Genomic Health, Inc., grants from Gilead Sciences Inc., grants from GlaxoSmithKline PLC, grants from Innocrin Pharmaceuticals Inc., grants from Janssen Pharmaceuticals, Inc., grants from Kantar Health, grants from Myriad Genetic Laboratories, Inc., grants from Novartis International AG, grants from Parexel International Corporation through the University of Utah or Western Institute for Biomedical Research, outside the submitted work; Mr Ahmed reports funding from the NIHR Oxford Biomedical Research Center (BRC), Aziz Foundation, Wolfson Foundation, and the Royal College Surgeons of England; Dr. Golozar reports personal fees from Regeneron Pharmaceuticals, outside the submitted work. She is a full-time employee at Regeneron Pharmaceuticals. This work was not conducted at Regeneron Pharmaceuticals. Miss Lane is supported by a Medical Research Council Doctoral Research Fellowship (MR/K501256/1) and a Versus Arthritis Clinical Research Fellowship (21605). Dr. Morales is supported by a Wellcome Trust Clinical Research Development Fellowship (Grant 214588/Z/18/Z) and reports grants from Chief Scientist Office (CSO), grants from Health Data Research UK (HDR-UK), grants from National Institute of Health Research (NIHR), outside the submitted work; Dr. Nyberg reports other from AstraZeneca, outside the submitted work; Dr. Subbian reports grants from National Science Foundation, grants from State of Arizona; Arizona Board of Regents, grants from Agency for Healthcare Research and Quality, outside the submitted work; Dr Prieto-Alhambra reports grants and other from AMGEN; grants, non-financial support and other from UCB Biopharma; grants from Les Laboratoires Servier, outside the submitted work; and Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programs organized by DPA's department and open for external participants. Ms. Kostka and Dr. Reich report being employees of IQVIA Inc at the time the analysis was conducted. Dr. Rijnbeek reports grants from Innovative Medicines Initiative, grants from Janssen Research and Development, during the conduct of the study. Dr. Ryan is an employee of Janssen Research and Development and shareholder of Johnson & Johnson. The views expressed are those of the authors and do not necessarily represent the views or policy of the Department of Veterans Affairs or the United States Government. No other relationships or activities that could appear to have influenced the submitted work.

## ETHICAL APPROVAL

All the data partners received Institutional Review Board (IRB) approval or exemption. STARR-OMOP had approval from IRB Panel #8 (RB-53248) registered to Leland Stanford Junior University under the Stanford Human Research Protection Program (HRPP). The use of VA data was reviewed by the Department of Veterans Affairs Central IRB, was determined to meet the criteria for exemption under Exemption Category 4 (3), and approved for Waiver of HIPAA Authorization. The research was approved by the Columbia University Institutional Review Board as an OHDSI network study. The use of SIDIAP was approved by the Clinical Research Ethics Committee of the IDIAPJGol (project code: 20/070-PCV). The use of CPRD was approved by the Independent Scientific Advisory Committee (ISAC) (protocol number 20_059RA2). The use of IQVIA-OpenClaims was exempted from IRB approval.

## TRANSPARENCY DECLARATION

Lead authors affirm that the paper is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

## DATA SHARING STATEMENT

Analyses were performed locally in compliance with all applicable data privacy laws. Although the underlying data is not readily available to be shared, authors contributing to this paper have direct access to the data sources used in this study. All results (e.g., aggregate statistics, not presented at a patient-level with redactions for minimum cell count) are available for public inquiry. These results are inclusive of site-identifiers by contributing data sources to enable interrogation of each contributing site. All analytic code and result sets are made available at: https://github.com/ohdsi-studies/Covid19CharacterizationCharybdis.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41366-021-00893-4.

**Correspondence** and requests for materials should be addressed to T.D-S.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[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. [3]Janssen Research & Development, Titusville, NJ, USA. [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands. [5]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, UK. [6]Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK. [7]College of Medicine and Health, University of Exeter, St Luke's Campus, Exeter, UK. [8]Faculty of Medicine, Islamic University of Gaza, Gaza, Palestine. [9]College of Pharmacy, Riyadh Elm University, Riyadh, Saudi Arabia. [10]Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [11]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [12]Real-World Evidence, Trial Form Support, Barcelona, Spain. [13]Cairo University, Faculty of Pharmacy, Cairo, Egypt. [14]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA. [15]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA. [16]Department of Biomedical Informatics, Columbia University, New York, NY, USA. [17]Department of Epidemiology, Johns Hopkins School of Public, Baltimore, MD, USA. [18]Pharmacoepidemiology, Regeneron Pharmaceuticals, Tarrytown, NY, USA. [19]DHC Technologies co, Ltd, Beijing, China. [20]Division of Cancer Sciences, School of Medical Sciences, University of Manchester, Manchester, UK. [21]Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN, USA. [22]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [23]College of Medicine, The University

of Arizona, Tucson, AZ, USA. [24]Division of Population Health and Genomics, University of Dundee, Dundee, UK. [25]New York-Presbyterian Hospital, New York, NY, USA. [26]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [27]Department of Medicine, Stanford University, Palo Alto, CA, USA. [28]Real World Solutions, IQVIA, Cambridge, MA, USA. [29]Data Science to Patient Value Program, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [30]College of Engineering, The University of Arizona, Tucson, AZ, USA. [31]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. [32]Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia. [33]Institute of Health Management, Southern Medical University, Guangzhou, China. [34]Nanfang Hospital, Southern Medical University, Guangzhou, China. [35]The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA. [37]These authors contributed equally: Martina Recalde, Elena Roel. [37]These authors jointly supervised this work: Kristin Kostka, Talita Duarte-Salles. ✉email: tduarte@idiapjgol.org

## 5.2. Study II: Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain

Roel E, Pistillo A, Recalde M, Sena AG, Fernández-Bertolín S, Aragón M, Puente D, Ahmed WU, Alghoul H, Alser O, Alshammari TM, Areia C, Blacketer C, Carter W, Casajust P, Culhane AC, Dawoud D, DeFalco F, DuVall SL, Falconer T, Golozar A, Gong M, Hester L, Hripcsak G, Tan EH, Jeon H, Jonnagaddala J, Lai LYH, Lynch KE, Matheny ME, Morales DR, Natarajan K, Nyberg F, Ostropolets A, Posada JD, Prats-Uribe A, Reich CG, Rivera DR, Schilling LM, Soerjomataram I, Shah K, Shah NH, Shen Y, Spotniz M, Subbian V, Suchard MA, Trama A, Zhang L, Zhang Y, Ryan PB, Prieto-Alhambra D, Kostka K, Duarte-Salles T.

*Cancer Epidemiology Biomarkers and Prevention*.
021;30(10):1884-1894
Doi: 10.1158/1055-9965.EPI-21-026.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266

CANCER EPIDEMIOLOGY, BIOMARKERS & PREVENTION | RESEARCH ARTICLE

# Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain

Elena Roel[1,2], Andrea Pistillo[1], Martina Recalde[1,2], Anthony G. Sena[3,4], Sergio Fernández-Bertolín[1], Maria Aragón[1], Diana Puente[1,2], Waheed-Ul-Rahman Ahmed[5,6], Heba Alghoul[7], Osaid Alser[8], Thamir M. Alshammari[9], Carlos Areia[10], Clair Blacketer[3], William Carter[11], Paula Casajust[12], Aedin C. Culhane[13,14], Dalia Dawoud[15], Frank DeFalco[3], Scott L. DuVall[16,17], Thomas Falconer[18,19], Asieh Golozar[20,21], Mengchun Gong[22], Laura Hester[23], George Hripcsak[18,19], Eng Hooi Tan[24], Hokyun Jeon[25], Jitendra Jonnagaddala[26], Lana Y.H. Lai[27], Kristine E. Lynch[16,17], Michael E. Matheny[28,29], Daniel R. Morales[30,31], Karthik Natarajan[18,19], Fredrik Nyberg[32], Anna Ostropolets[18], José D. Posada[33], Albert Prats-Uribe[24], Christian G. Reich[34], Donna R. Rivera[35], Lisa M. Schilling[11], Isabelle Soerjomataram[36], Karishma Shah[5], Nigam H. Shah[33], Yang Shen[22], Matthew Spotniz[18], Vignesh Subbian[37], Marc A. Suchard[38], Annalisa Trama[39], Lin Zhang[40,41], Ying Zhang[22], Patrick B. Ryan[3,18], Daniel Prieto-Alhambra[24], Kristin Kostka[34], and Talita Duarte-Salles[1]

## ABSTRACT

**Background:** We described the demographics, cancer subtypes, comorbidities, and outcomes of patients with a history of cancer and coronavirus disease 2019 (COVID-19). Second, we compared patients hospitalized with COVID-19 to patients diagnosed with COVID-19 and patients hospitalized with influenza.

**Methods:** We conducted a cohort study using eight routinely collected health care databases from Spain and the United States, standardized to the Observational Medical Outcome Partnership common data model. Three cohorts of patients with a history of cancer were included: (i) diagnosed with COVID-19, (ii) hospitalized with COVID-19, and (iii) hospitalized with influenza in 2017 to 2018. Patients were followed from index date to 30 days or death. We reported demographics, cancer subtypes, comorbidities, and 30-day outcomes.

**Results:** We included 366,050 and 119,597 patients diagnosed and hospitalized with COVID-19, respectively. Prostate and breast cancers were the most frequent cancers (range: 5%–18% and 1%–14% in the diagnosed cohort, respectively). Hematologic malignancies were also frequent, with non-Hodgkin's lymphoma being among the five most common cancer subtypes in the diagnosed cohort. Overall, patients were aged above 65 years and had multiple comorbidities. Occurrence of death ranged from 2% to 14% and from 6% to 26% in the diagnosed and hospitalized COVID-19 cohorts, respectively. Patients hospitalized with influenza ($n = 67,743$) had a similar distribution of cancer subtypes, sex, age, and comorbidities but lower occurrence of adverse events.

**Conclusions:** Patients with a history of cancer and COVID-19 had multiple comorbidities and a high occurrence of COVID-19-related events. Hematologic malignancies were frequent.

**Impact:** This study provides epidemiologic characteristics that can inform clinical care and etiologic studies.

[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, Spain. [3]Janssen Research and Development, Titusville, New Jersey. [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands. [5]NDORMS, University of Oxford, Botnar Research Centre, Windmill Road, Oxford, United Kingdom. [6]College of Medicine and Health, University of Exeter, St Luke's Campus, Heavitree Road, Exeter, United Kingdom. [7]Faculty of Medicine, Islamic University of Gaza, Gaza, Palestine. [8]Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts. [9]Medication Safety Research Chair, King Saud University, Riyadh, Saudi Arabia. [10]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom. [11]Data Science to Patient Value Program, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado. [12]Real-World Evidence, Trial Form Support, Barcelona, Spain. [13]Department of Data Science, Dana-Farber Cancer Institute, Boston, Massachusetts. [14]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [15]Faculty of Pharmacy, Cairo University, Cairo, Egypt. [16]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, Utah. [17]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah. [18]Department of Biomedical Informatics, Columbia University, New York, New York. [19]New York-Presbyterian Hospital, New York, New York. [20]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, Maryland. [21]Pharmacoepidemiology, Regeneron Pharmaceuticals, Westchester County, New York. [22]Digital Health China Technologies Co., Ltd., Beijing, China. [23]Janssen Research and Development, LLC, Raritan, New Jersey. [24]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, United Kingdom. [25]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea. [26]School of Public Health and Community Medicine, UNSW Sydney, Kensington, Australia. [27]School of Medical Sciences, University of Manchester, Manchester, United Kingdom. [28]Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, Tennessee. [29]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee. [30]Division of Population Health and Genomics, University of Dundee, Dundee, United Kingdom. [31]University of Southern Denmark, Odense, Denmark. [32]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [33]Department of Medicine, School of Medicine, Stanford University, Redwood City, California. [34]Real World Solutions, IQVIA,

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



## Introduction

Shortly after the emergence of the coronavirus disease 2019 (COVID-19), patients with cancer were reported to be a high-risk population for COVID-19 (1, 2). These patients have an increased susceptibility to infections as a result of their immunosuppressed state, caused by the cancer itself, certain types of chemo- or immunotherapy, or surgery and a higher exposure to healthcare-associated infections (3). In addition, patients with cancer are often older and have additional comorbidities, which might increase their risk of worse COVID-19 outcomes (4).

Prior studies assessing COVID-19-related risks in the cancer population have demonstrated conflicting results. Some studies found that patients with cancer have an increased risk of COVID-19-related hospitalization, admission to intensive care units, and mortality compared with patients without cancer (1, 2, 4, 5), whereas others did not (6, 7). These studies included a limited number of patients with cancer (mostly hospitalized) and used different definitions for cancer (e.g., active cancer, history of cancer), which limit their generalizability. Furthermore, they presented results for models adjusted by (different) arbitrary covariates, without a theoretical framework of confounding variables, which limits the interpretation for descriptive and causal inference purposes (8, 9).

Given that COVID-19 is a novel disease, large descriptive studies are needed to inform public health strategies and clinical care, as well as to provide the groundwork for etiologic studies. In addition, large studies with detailed information of medical conditions and health outcomes, such as thromboembolic events, in patients with cancer and COVID-19 are lacking to date. To fill that gap, we described the demographics, cancer subtypes, comorbidities, and outcomes of patients with a history of cancer and COVID-19. In addition, we compared patients with a history of cancer hospitalized with COVID-19 to (i) patients with a history of cancer diagnosed with COVID-19; and (ii) patients with a history of cancer hospitalized with seasonal influenza (2017–2018) as a benchmark.

## Materials and Methods

### Study design and setting

This multinational cohort study was part of the CHARYBDIS (Characterizing Health Associated Risks, and Your Baseline Disease In SARS-COV-2) project, designed by the Observational Health Data Sciences and Informatics (OHDSI) community. CHARYBDIS is a large-scale study aiming to characterize individuals with COVID-19 using routinely-collected healthcare data (protocol available at https://www.ohdsi.org/wp-content/uploads/2020/07/Protocol_COVID-19-Charybdis-Characterisation_V5.docx). Twenty-two databases standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM; ref. 10) have contributed to CHARYBDIS to date. The OHDSI network maintains the OMOP-CDM, and its members have developed a wide range of tools to facilitate analyses of such mapped data (11). Results for this substudy were extracted from the overarching result set on January 29, 2021.

We included those databases reporting on at least 140 subjects with a history of cancer diagnosed and/or hospitalized with COVID-19. This cut-off was established to estimate the prevalence of conditions affecting 10% of the study population with a confidence interval (CI) width of ±5%. The selection process of databases is depicted in Supplementary Fig. S1. Eight databases from Spain and the United States were included in this study.

Spanish data came from the Information System for Research in Primary Care (SIDIAP) database, a primary care database from Catalonia, a northeastern region in Spain (12). Data from the United States included Electronic Health Records (EHR) from the hospital setting: Colorado University Anschutz Medical Campus Health Data Compass (CU-AMC-HDC; Colorado), Columbia University Irving Medical Center (CUIMC; New York), Optum-EHR (national; ref. 13), Stanford Medicine Research Data Repository (STARR-OMOP; California), Department of Veteran Affairs (VA-OMOP; national, including mostly veterans with 93% males); and claims data: HealthVerity and IQVIA-OpenClaims (both national). A description of each database is provided in Supplementary Table S1. SIDIAP and CUIMC included patients with COVID-19 identified from March to May 2020, HealthVerity, and STARR-OMOP spanned to June 2020, CU-AMC-HDC to July 2020, VA-OMOP to September 2020, and IQVIA-OpenClaims and Optum-EHR to October 2020.

### Study participants

We included three non-mutually exclusive cohorts of patients with a history of cancer: (i) diagnosed with COVID-19, (ii) hospitalized with COVID-19, and (iii) hospitalized with seasonal influenza in 2017 to 2018.

We included all patients (regardless of age) with at least 1 year of observation time available prior to index date (i.e., date of start of the cohort). Patients with a history of cancer were defined as those having a record of any malignant neoplasm excluding non-melanoma skin cancer prior to index date. Patients diagnosed with COVID-19 were those having a clinical diagnosis and/or a positive severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) test documented in outpatient or inpatient records. Patients hospitalized with COVID-19 were those who had a hospitalization episode and a COVID-19 clinical diagnosis or positive SARS-CoV-2 test within a time window of 21 days prior to admission up to the end of their hospitalization. We chose this time window to include patients with a diagnosis prior to hospitalization and to allow for a record delay in

Cambridge, Massachusetts. [35]Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, Maryland. [36]Section of Cancer Surveillance, International Agency for Research on Cancer, Lyon, France. [37]College of Engineering, University of Arizona, Tucson, Arizona. [38]Fielding School of Public Health, University of California, Los Angeles, California. [39]Fondazione IRCSS Istituto Nazionale dei Tumori, Milan, Italy. [40]School of Population Medicine and Public Health, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China. [41]School of Population Health and Global Health, University of Melbourne, Melbourne, Victoria, Australia.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

K. Kostka and T. Duarte-Salles contributed equally as the co-senior authors of this article.

**Corresponding Author:** Talita Duarte-Salles, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, 08007, Spain. E-mail: tduarte@idiapjgol.org

Cancer Epidemiol Biomarkers Prev 2021;XX:XX–XX

**doi:** 10.1158/1055-9965.EPI-21-0266

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



diagnoses or test results. Similarly, patients hospitalized with seasonal influenza were those who had a hospitalization episode and a influenza clinical diagnosis or positive test result for influenza in 2017 to 2018 (14). The criteria to define patients with cancer history and COVID-19 and influenza cases can be found in Supplementary Table S2.

Index date for the diagnosed cohort was the date of clinical diagnosis or the earliest test day registered within seven days of a first positive test, whichever occurred first. Index date for both hospitalized cohorts (COVID-19 and influenza) was the day of hospitalization. Therefore, although time windows are slightly different, both COVID-19 cohorts largely overlap, as most individuals in the hospitalized cohort are also included in the diagnosed cohort.

All patients were followed from the index date to the earliest of either death, end of the observation period (15), or 30 days.

### Patient characteristics and outcomes

We identified over 15,000 baseline medical conditions based on the Systematized Nomenclature of Medicine (SNOMED) hierarchy, with all descendant codes included (15). In addition, we created specific definitions for comorbidities and outcomes of particular interest (available in Supplementary Table S2). To describe the frequency of cancer subtypes by topographical location (henceforth, referred to as cancer types), we selected 26 cancer types based on the most prevalent cancers in both countries (16). The codes used to identify each cancer type are available in Supplementary Table S3. Of note, although we required all subjects in our study to have at least 1 year of prior history available, all the conditions recorded at any time prior to the index date (including the day prior) were reported.

We report here sex, age, race, antineoplastic and immunomodulating treatment received the month and year prior to index date, and key comorbidities. The only information available for race was the proportion of African American patients, which was reported in four databases (CU-AMC-HDC, CUIMC, Optum-EHR, and VA-OMOP).

The 30-day outcomes of interest in the diagnosed cohort were hospitalization and death (from all causes). In the hospitalized cohorts (COVID-19 and influenza), the outcomes of interest were acute respiratory distress syndrome (ARDS), acute kidney injury (AKI), cardiovascular disease events, deep vein thrombosis, pulmonary embolism, sepsis, requirement of intensive services (identified by a recorded mechanical ventilation and/or a tracheostomy and/or extracorporeal membrane oxygenation procedure), and death (from all causes). SIDIAP only reported death and hospitalization, whereas CU-AMC-HDC did not report any outcome.

### Analysis

Analysis was performed through a federated analysis approach (15). Following a prespecified analysis plan, an analytical code for the whole CHARYBDIS study was developed and run locally in each site (code available at zenodo.org) (17). Individual-level data remained within host institutions, only aggregate results were provided to the research team. All the results are available for consultation on a regularly updated website as new databases and/or results are added (https://data.ohdsi.org/Covid19CharacterizationCharybdis/).

We report results by cohort and database. Demographics, cancer types, comorbidities, and outcomes are reported as proportions along with 95% confidence intervals (CI). To calculate these proportions, a minimum count required of 5 individuals was established to minimize the risk of re-identification of patients. We also report the ranking of the 10 most common cancer types by frequency. In addition, we summarized the prevalence of all the baseline conditions retrieved in a Manhattan-style plot (a type of scatter plot used to represent large numbers of data points).

To compare characteristics between study cohorts, we calculated standardized mean differences (SMD). SMD are independent of sample sizes and can be used to compare the prevalence of dichotomous variables between two groups. An $|SMD|>0.1$ indicates a meaningful difference in the prevalence of a given condition (18, 19). As this study was designed as a detailed descriptive study, statistical modelling was out of scope in the developed analytical packages. Therefore, differences across the groups compared should not be interpreted as causal effects.

We used R version 3.6 for data visualization. All the data partners obtained Institutional Review Board (IRB) approval or exemption to conduct this study.

## Results

### Lifetime cancer prevalence

Overall, we identified 3,067,116 patients diagnosed and 572,300 patients hospitalized with COVID-19. The lifetime cancer prevalence range across databases was 4% to 25% in patients diagnosed; and 11% to 40% in patients hospitalized (Supplementary Table S4). In addition, 274,557 patients hospitalized with seasonal influenza in 2017 to 2018 were identified (lifetime cancer prevalence range: 18%–39%).

We included 366,050 patients diagnosed (Spain: 8,854; United States: 357,196) and 119,597 patients hospitalized (Spain: 2,610; United States: 116,987) with COVID-19 and cancer history; and 67,743 patients hospitalized (all from the United States) with seasonal influenza and cancer history.

### Demographics

The distribution of demographics, comorbidities, and outcomes of both COVID-19 cohorts can be found in **Table 1** (95% CI of each condition available in Supplementary Table S5). In the diagnosed cohort, patients were more commonly female (range: 53%–59%), aside from STARR-OMOP (47%) and VA-OMOP (7%). In contrast, in the hospitalized cohort, male slightly predominated in all databases (51%–60%, VA-OMOP: 96%), aside from HealthVerity and Optum-EHR (50% in both). Patients were mainly aged above 65 years in both COVID-19 cohorts but patients hospitalized were consistently older than those diagnosed (Supplementary Fig. S2). In the few databases reporting race, the proportion of African American patients was higher in the hospitalized cohort (9%–35%) than in the diagnosed cohort (6%–29%).

### Cancer types

For both COVID-19 cohorts, the frequency of each cancer type is reported in Supplementary Table S6. The top 10 cancer types by frequency are reported in **Table 2**. In the diagnosed cohort, the most frequent cancers in four databases were breast (SIDIAP: 14.2%; CU-AMC-HDC: 7.3%; Optum-EHR: 6.7%; and STARR-OMOP: 12.3%) and prostate cancer (CUIMC: 6.1%; HealthVerity: 12.2%; IQVIA-OpenClaims: 6.4%; VA-OMOP: 18.1%). In all databases, non-Hodgkin's lymphoma (NHL) was among the five most common cancers. Bladder, colorectal, leukemia, and lung cancer were among the ten most frequent in at least seven databases.

In the hospitalized cohort, prostate cancer was the most frequent cancer in all databases (equally with NHL in CU-AMC-HDC, 6.4%); aside from Optum-EHR (second most frequent). NHL was among the three most frequent cancers in all databases aside from SIDIAP (fifth

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



**Table 1.** Demographics, comorbidities, and outcomes among patients with a history of cancer diagnosed and hospitalized with COVID-19.

| | Patients with history of cancer diagnosed with COVID-19 | | | | | | | | Patients with history of cancer hospitalized with COVID-19 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Characteristics, in % | SIDIAP $n = 8,854$ | CU-AMC-HDC $n = 806$ | CUIMC $n = 1,433$ | Health Verity $n = 4,857$ | IQVIA-Open Claims $n = 315,523$ | Optum-EHR $n = 22,996$ | STARR-OMOP $n = 821$ | VA-OMOP $n = 10,760$ | SIDIAP $n = 2,610$ | CU-AMC-HDC $n = 265$ | CUIMC $n = 561$ | Health Verity $n = 797$ | IQVIA-Open Claims $n = 105,931$ | Optum-EHR $n = 5,806$ | STARR-OMOP $n = 244$ | VA-OMOP $n = 3,383$ |
| Sex | | | | | | | | | | | | | | | | |
| Female | 53.9 | 53.0 | 54.7 | 53.7 | 55.1 | 58.6 | 47.0 | 7.1 | 39.7 | 46.8 | 46.3 | 50.1 | 48.7 | 50.3 | 41.8 | 3.6 |
| Male | 46.1 | 47.0 | 45.3 | 46.3 | 44.9 | 41.4 | 53.0 | 92.9 | 60.3 | 53.2 | 53.7 | 49.9 | 51.3 | 49.7 | 58.2 | 96.4 |
| Race, African American | – | 5.6 | 9.7 | – | – | 11.5 | – | 29.0 | – | 9.1 | 10.0 | – | – | 14.2 | – | 35.1 |
| Antineoplastic and immunomodulating agents | | | | | | | | | | | | | | | | |
| The month prior | 10.9 | 22.5 | 12.3 | 5.4 | 9.8 | 6.3 | 18.8 | 12.0 | 10.8 | 28.3 | 14.6 | 5.9 | 10.7 | 10.9 | 18.9 | 14.2 |
| The year prior | 13.6 | 35.1 | 24.6 | 20.3 | 18.0 | 16.3 | 31.9 | 21.5 | 13.7 | 37.7 | 26.0 | 18.6 | 19.9 | 20.7 | 29.9 | 24.1 |
| Comorbidities | | | | | | | | | | | | | | | | |
| Asthma | 4.7 | 15.4 | 20.5 | 9.9 | 17.2 | 20.3 | 14.3 | 10.8 | 3.9 | 12.1 | 22.8 | 10.0 | 16.3 | 15.8 | 16.4 | 9.9 |
| COPD | 33.7 | 23.7 | 19.8 | 17.3 | 26.8 | 18.4 | 11.9 | 43.3 | 41.5 | 29.4 | 28.2 | 31.4 | 34.8 | 28.6 | 11.1 | 53.2 |
| Type 2 diabetes | 17.8 | 26.9 | 38.0 | 31.7 | 47.5 | 29.9 | 22.0 | 50.3 | 22.5 | 36.2 | 55.1 | 43.0 | 56.9 | 40.5 | 23.8 | 58.7 |
| Hyperlipidemia | 22.8 | 38.7 | 30.8 | 40.1 | 38.4 | 44.7 | 38.7 | 58.8 | 24.6 | 42.3 | 38.0 | 48.4 | 41.5 | 49.1 | 44.3 | 61.2 |
| Obesity | 41.8 | 49.3 | 52.8 | 21.8 | 35.5 | 60.4 | 40.0 | 52.7 | 47.8 | 52.1 | 57.0 | 25.7 | 38.8 | 60.4 | 39.8 | 53.4 |
| Heart disease | 34.3 | 50.0 | 62.8 | 39.9 | 68.1 | 51.1 | 40.8 | 71.4 | 41.6 | 59.2 | 76.5 | 60.9 | 78.6 | 63.9 | 45.5 | 79.1 |
| Hypertension | 33.2 | 60.2 | 70.0 | 58.1 | 80.8 | 61.3 | 52.4 | 86.7 | 37.3 | 69.4 | 83.1 | 74.4 | 89.4 | 73.3 | 58.2 | 92.8 |
| Anxiety | 24.2 | 19.1 | 12.2 | 13.5 | 13.5 | 18.6 | 16.6 | 31.1 | 19.3 | 18.1 | 10.7 | 19.4 | 14.2 | 18.2 | 18.9 | 30.1 |
| Dementia | 10.7 | 5.8 | 11.0 | 7.1 | 18.3 | 6.3 | 1.7 | 14.3 | 7.2 | 7.5 | 21.2 | 15.8 | 21.5 | 10.8 | – | 23.7 |
| Depression | 7.2 | 9.2 | 14.8 | 2.7 | 12.3 | 10.9 | 10.4 | 21.1 | 5.7 | 9.1 | 20.0 | 4.9 | 12.7 | 9.6 | 12.7 | 19.1 |
| Anemia | 19.6 | 22.1 | 19.3 | 24.8 | 24.8 | 18.0 | 24.7 | 25.8 | 20.0 | 35.1 | 25.0 | 40.9 | 33.3 | 29.1 | 32.0 | 36.7 |
| Autoimmune condition | 10.4 | 19.9 | 30.8 | 13.0 | 32.0 | 18.3 | 14.6 | 30.3 | 11.2 | 20.0 | 38.1 | 17.8 | 36.2 | 20.8 | 11.9 | 35.4 |
| Chronic kidney disease | 19.4 | 23.8 | 27.4 | 18.4 | 31.7 | 27.9 | 18.0 | 34.1 | 23.6 | 34.0 | 40.8 | 36.0 | 44.1 | 39.2 | 20.1 | 44.3 |
| Chronic liver disease | 1.9 | 3.2 | 3.6 | 2.2 | 2.0 | 2.4 | 7.3 | 6.1 | 1.8 | 4.9 | 5.0 | 4.0 | 3.0 | 3.5 | 4.9 | 8.8 |
| Outcomes | | | | | | | | | | | | | | | | |
| Death | 14.4 | – | 10.4 | – | – | 1.7 | – | 7.6 | 21.4 | – | 26.2 | – | – | 5.5 | – | 18.1 |
| Hospitalization | 24.9 | – | 34.8 | 13.5 | 32.1 | 24.1 | 27.0 | 27.1 | NA | NA | NA | NA | NA | NA | NA | NA |
| Intensive services requirement | NA | NA | NA | NA | NA | NA | NA | NA | – | – | – | – | 8.7 | 13.0 | 5.7 | 16.0 |
| ARDS during hospitalization | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 15.9 | 26.5 | 33.3 | 41.8 | 8.2 | 41.2 |
| Cardiovascular disease events | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 6.8 | 10.8 | 11.2 | 16.7 | 8.2 | 20.8 |
| Deep vein thrombosis events | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 2.1 | 3.0 | 2.4 | 4.0 | – | 4.8 |
| Pulmonary embolism events | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 2.7 | 2.1 | 2.1 | 3.5 | – | 4.0 |
| Acute kidney injury during hospitalization | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 16.0 | 11.4 | 9.6 | 16.6 | 11.9 | 14.4 |
| Sepsis during hospitalization | NA | NA | NA | NA | NA | NA | NA | NA | – | – | 6.1 | 17.7 | 18.4 | 25.0 | 9.8 | 20.5 |

Notes: – indicates data not available or below the minimum cell count required (5 individuals); NA indicates not applicable.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



**Table 2.** Top 10 cancer types among patients with a history of cancer diagnosed and hospitalized with COVID-19.

| | Patients with a history of cancer diagnosed with COVID-19 | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SIDIAP | | CU-AMC-HDC | | CUIMC | | HealthVerity | | IQVIA-Open Claims | | Optum-EHR | | STARR-OMOP | | VA-OMOP | |
| Rank | *n* = 8,854 | | *n* = 806 | | *n* = 1,433 | | *n* = 4,857 | | *n* = 315,523 | | *n* = 22,996 | | *n* = 821 | | *n* = 10,760 | |
| | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) |
| 1 | Breast | 14.2 (13.5–14.9) | Breast | 7.3 (5.5–9.1) | Prostate | 6.1 (4.9–7.3) | Prostate | 12.2 (11.3–13.1) | Prostate | 6.4 (6.3–6.5) | Breast | 6.7 (6.4–7.0) | Breast | 12.3 (10.0–14.6) | Prostate | 18.1 (17.4–18.8) |
| 2 | Colorectal | 10.4 (9.8–11.0) | Prostate | 6.8 (5.1–8.5) | NHL | 5.4 (4.2–6.6) | Breast | 11.5 (10.6–12.4) | Breast | 6.2 (6.1–6.3) | Prostate | 5.0 (4.7–5.3) | Prostate | 10.2 (8.1–12.3) | Lung | 3.9 (3.5–4.3) |
| 3 | Prostate | 9.4 (8.8–10.0) | NHL | 5.0 (3.5–6.5) | Breast | 5.2 (4.0–6.4) | NHL | 4.6 (4.0–5.2) | NHL | 3.1 (3.0–3.2) | Uterus | 3.3 (3.1–3.5) | Liver | 7.7 (5.9–9.5) | NHL | 3.8 (3.4–4.2) |
| 4 | Bladder | 6.4 (5.9–6.9) | Lung | 3.8 (2.5–5.1) | Leukemia | 4.4 (3.3–5.5) | Colorectal | 3.9 (3.4–4.4) | Colorectal | 2.5 (2.4–2.6) | LOCP | 3.3 (3.1–3.5) | Lung | 6.9 (5.2–8.6) | Bladder | 3.3 (3.0–3.6) |
| 5 | NHL | 3.0 (2.6–3.4) | Leukemia | 3.3 (2.1–4.5) | Liver | 3.3 (2.4–4.2) | Thyroid | 3.5 (3.0–4.0) | Lung | 2.5 (2.4–2.6) | NHL | 2.7 (2.5–2.9) | NHL | 5.8 (4.2–7.4) | LOCP | 2.9 (2.6–3.2) |
| 6 | Melanoma | 2.9 (2.6–3.2) | Melanoma | 3.0 (1.8–4.2) | Lung | 3.1 (2.2–4.0) | Leukemia | 2.8 (2.3–3.3) | Leukemia | 2.1 (2.1–2.1) | Leukemia | 2.2 (2.0–2.4) | Thyroid | 5.2 (3.7–6.7) | Leukemia | 2.6 (2.3–2.9) |
| 7 | Leukemia | 2.7 (2.4–3.0) | Colorectal | 2.9 (1.7–4.1) | Multiple myeloma | 3.1 (2.2–4.0) | Lung | 2.7 (2.2–3.2) | Bladder | 1.5 (1.5–1.5) | Lung | 2.1 (1.9–2.3) | LOCP | 4.9 (3.4–6.4) | Colorectal | 2.6 (2.3–2.9) |
| 8 | Uterus | 2.5 (2.2–2.8) | Multiple myeloma | 2.5 (1.4–3.6) | Colorectal | 2.7 (1.9–3.5) | Bladder | 2.5 (2.1–2.9) | Liver | 1.4 (1.4–1.4 | Colorectal | 1.9 (1.7–2.1) | Leukemia | 3.8 (2.5–5.1) | Kidney | 2.5 (2.2–2.8) |
| 9 | Kidney | 2.4 (2.1–2.7) | Bladder | 2.2 (1.2–3.2) | Uterus | 2.0 (1.3–2.7) | Multiple myeloma | 2.1 (1.7–2.5) | LOCP | 1.3 (1.3–1.3) | Thyroid | 1.5 (1.3–1.7) | Colorectal | 3.7 (2.4–5.0) | Liver | 1.8 (1.5–2.1) |
| 10 | LOCP | 1.4 (1.2–1.6) | Thyroid | 2.2 (1.2–3.2) | Kidney | 1.8 (1.1–2.5) | Kidney | 2.1 (1.7–2.5) | Kidney | 1.3 (1.3–1.3) | Bladder | 1.3 (1.2–1.4) | Bladder | 3.2 (2.0–4.4) | Larynx | 1.3 (1.1–1.5) |
| | Patients with a history of cancer hospitalized with COVID-19 | | | | | | | | | | | | | | | |
| | SIDIAP | | CU-AMC-HDC | | CUIMC | | HealthVerity | | IQVIA-Open Claims | | Optum-EHR | | STARR-OMOP | | VA-OMOP | |
| Rank | *n* = 2,610 | | *n* = 265 | | *n* = 561 | | *n* = 797 | | *n* = 105,931 | | *n* = 5,806 | | *n* = 244 | | *n* = 3,383 | |
| | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) | Cancer | % (95% CI) |
| 1 | Prostate | 12.8 (11.5–14.1) | NHL | 6.4 (3.5–9.3) | Prostate | 7.7 (5.5–9.9) | Prostate | 10.8 (8.6–13.0) | Prostate | 7.5 (7.3–7.7) | Breast | 6.2 (5.6–6.8) | Prostate | 10.7 (6.8–14.6) | Prostate | 19.4 (18.1–20.7) |
| 2 | Colorectal | 11.9 (10.7–13.1) | Prostate | 6.4 (3.5–9.3) | NHL | 6.4 (4.4–8.4) | Breast | 9.2 (7.2–11.2) | Breast | 5.5 (5.4–5.6) | Prostate | 6.1 (5.5–6.7) | LOCP | 9.0 (5.4–12.6) | Lung | 5.7 (4.9–6.5) |
| 3 | Breast | 9.4 (8.3–10.5) | Lung | 6.0 (3.1–8.9) | Leukemia | 4.1 (2.5–5.7) | NHL | 7.4 (5.6–9.2) | NHL | 4.2 (4.1–4.3) | NHL | 4.2 (3.7–4.7) | Lung | 9.0 (5.4–12.6) | NHL | 4.6 (3.9–5.3) |
| 4 | Bladder | 8.5 (7.4–9.6) | Multiple myeloma | 4.2 (1.8–6.6) | Lung | 3.7 (2.1–5.3) | Colorectal | 5.0 (3.5–6.5) | Lung | 4.1 (4.0–4.2) | Lung | 3.6 (3.1–4.1) | Breast | 7.4 (4.1–10.7) | Bladder | 3.9 (3.2–4.6) |
| 5 | NHL | 4.3 (3.5–5.1) | Leukemia | 4.2 (1.8–6.6) | Liver | 3.6 (2.1–5.1) | Leukemia | 4.1 (2.7–5.5) | Colorectal | 3.2 (3.1–3.3) | Leukemia | 3.4 (2.9–3.9) | Liver | 6.6 (3.5–9.7) | Leukemia | 3.3 (2.7–3.9) |
| 6 | Leukemia | 4.2 (3.4–5.0) | Breast | 3.8 (1.5–6.1) | Colorectal | 3.4 (1.9–4.9) | Lung | 4.0 (2.6–5.4) | Leukemia | 3.0 (2.9–3.1) | LOCP | 3.0 (2.6–3.4) | Thyroid | 6.1 (3.1–9.1) | Colorectal | 3.2 (2.6–3.8) |
| 7 | Kidney | 2.8 (2.2–3.4) | Liver | 3.8 (1.5–6.1) | Multiple myeloma | 3.4 (1.9–4.9) | Multiple myeloma | 3.4 (2.1–4.7) | Liver | 2.3 (2.2–2.4) | Colorectal | 2.7 (2.3–3.1) | Pancreas | 5.3 (2.5–8.1) | Liver | 2.8 (2.2–3.4) |
| 8 | Melanoma | 2.3 (1.7–2.9) | – | | Breast | 3.2 (1.7–4.7) | Bladder | 3.0 (1.8–4.2) | Multiple myeloma | 1.9 (1.8–2.0) | Uterus | 2.6 (2.2–3.0) | Leukemia | 4.9 (2.2–7.6) | LOCP | 2.8 (2.2–3.4) |
| 9 | Uterus | 1.8 (1.3–2.3) | – | | Central nervous system | 2.0 (0.8–3.2) | Uterus | 2.8 (1.7–3.9) | Bladder | 1.9 (1.8–2.0) | Bladder | 2.4 (2.0–2.8) | Oropharynx | 4.5 (1.9–7.1) | Kidney | 2.7 (2.2–3.2) |
| 10 | Liver | 1.5 (1.0–2.0) | – | | Uterus | 1.8 (0.7–2.9) | LOCP | 2.5 (1.4–3.6) | Kidney | 1.7 (1.6–1.8) | Liver | 2.3 (1.9–2.7) | – | | Multiple myeloma | 1.8 (1.4–2.2) |

Notes: – indicates data not available. A single individual can have multiple cancer types recorded.
Abbreviation: LOCP: Lip, oral cavity, and pharynx.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



most frequent) and STARR-OMOP. Leukemia, liver and lung cancer were also within the top 10 in the majority of databases. We did not observe meaningful differences (i.e., |SMD|>0.1) when comparing cancer types between the diagnosed and the hospitalized cohorts (Supplementary Fig. S3).

### Prior comorbidities

In both COVID-19 cohorts, the most common comorbidities were cardiometabolic conditions, which were more frequent in U.S. databases (especially VA-OMOP) than in the Spanish SIDIAP database. For example, in the United States, the range of hypertension was 52%–87% (Spain: 32%) among diagnosed and 58%–93% (Spain: 33%) among hospitalized patients (**Table 1**). The prevalence of all the prior conditions summarized is shown in **Fig. 1**. Several comorbidities were more frequent among patients hospitalized compared with patients diagnosed (SMD>0.1): heart disease and chronic kidney disease (all databases except STARR-OMOP); hypertension and type 2 diabetes (all except SIDIAP and STARR-OMOP; **Fig. 2**).

### Thirty-day outcomes

In the COVID-19 diagnosed cohort, hospitalization in the U.S. databases ranged from 14% to 35% (Spain: 25%) and occurrence of death from 2% to 10% (Spain: 14%). In the COVID-19 hospitalized cohort, outcomes were heterogeneous across databases. ARDS (range 8%–42%) was higher than 30% in three out of six databases (IQVIA-OpenClaims, Optum-EHR, VA-OMOP). Sepsis (6%–25%), cardiovascular disease events (7%–21%) and AKI (10%–17%) were also common. Thromboembolic events were less frequent (deep vein thrombosis: 2%–5%; pulmonary embolism: 2%–4%). Intensive services requirement ranged from 6% to 16%, whereas occurrence of death ranged from 6% to 26% in the United States (Spain: 21%).

### Comparison of patients hospitalized with COVID-19 to those with influenza

The characteristics of patients hospitalized with seasonal influenza and the frequency of each cancer type are reported in Supplementary Tables S7 and S8, respectively. Aside from VA-OMOP (96% male), the proportion of males ranged from 45% to 53%, and the majority of patients clustered around the ages of 60 to 85 years old (Supplementary Fig. S4). The proportion of African American patients was lower in the Influenza cohort than in the hospitalized COVID-19 cohort (Optum-EHR: 10% vs. 14%; VA-OMOP: 17% vs. 35%). When comparing the frequency of cancer types between patients with COVID-19 and influenza, we did not observe consistent differences across databases (Supplementary Fig. S5). The distribution of comorbidities was similar in both groups, with few exceptions (**Fig. 3A**). For example, chronic obstructive pulmonary disease (COPD) was more common among patients with influenza in CU-AMC-HDC, Optum-EHR, and VA-OMOP (**Fig. 4A**). Aside from CUIMC, outcomes were slightly more frequent in patients with COVID-19 in all databases. ARDS and death were meaningfully more frequent in patients with COVID-19. ARDS ranged from 16% to 42% (COVID-19) versus 14%–30% (influenza), with SMD>0.2 in IQVIA-OpenClaims and Optum-EHR and SMD>0.1 in VA-OMOP. Occurrence of death was higher among patients with COVID-19 compared with patients with influenza in Optum-EHR and VA-OMOP: 6% vs. 1% and 18% vs. 6%, respectively (SMD>0.2; **Figs. 3B** and **4B**).

## Discussion

In this multinational cohort study, we described the characteristics of 366,050 patients with a history of cancer and COVID-19, including outcomes rarely reported in this population (e.g., deep vein thrombosis, pulmonary embolism, or acute kidney injury).

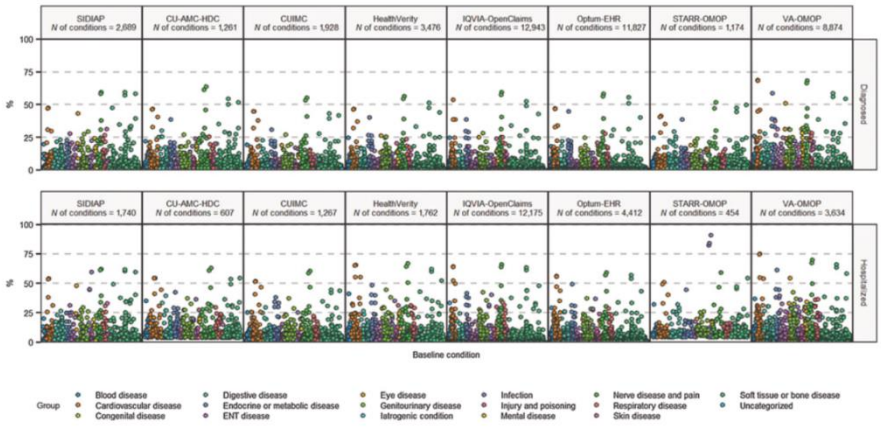



**Figure 1.**

Prevalence of baseline conditions among patients with a history of cancer diagnosed and hospitalized with COVID-19. Each dot represents the prevalence of one baseline condition, with the color indicating the type of condition (i.e., the group, for example blood disease, etc.). Conditions are represented by cohort and database along the *x*-axis, whereas the prevalence (in %) is displayed on the *y*-axis. NOTES: Only conditions meeting the minimum count requirement (5 individuals) are shown. *N* of conditions means the total number of conditions depicted (by cohort and database).

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



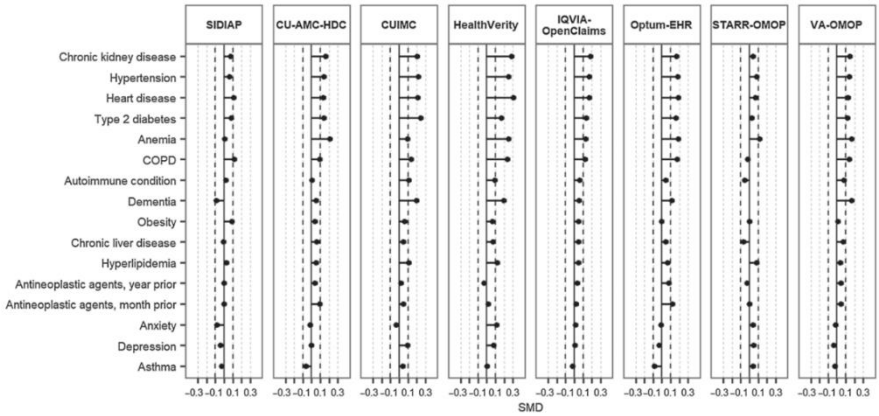



**Figure 2.**

SMDs of selected baseline comorbidities between patients with cancer diagnosed and hospitalized with COVID-19. SMD<0 indicates that the prevalence was greater in patients diagnosed, SMD>0 indicates that the prevalence was greater in patients hospitalized. NOTES: Comorbidities ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Black-dotted lines indicate an |SMD| of 0.1. SMD calculated for comorbidities meeting the minimum count required (5 individuals) in each database and cohort.

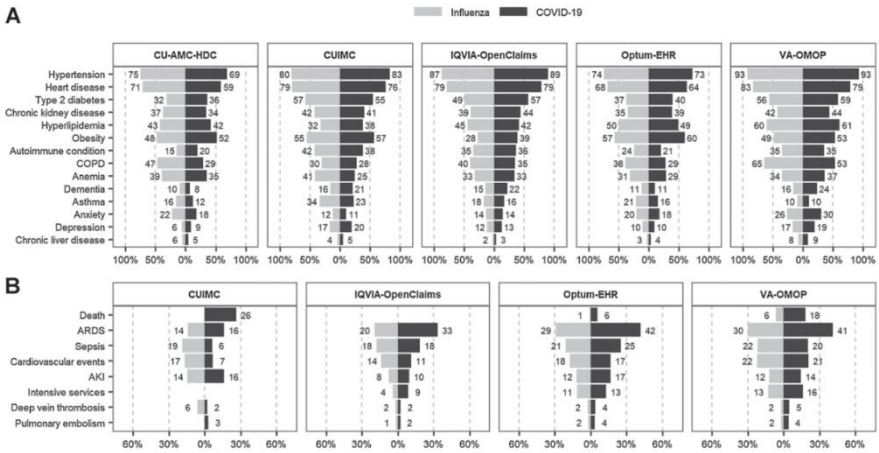



**Figure 3.**

Baseline comorbidities (**A**) and 30-day outcomes (**B**) among patients with history of cancer hospitalized with COVID-19 and with seasonal influenza. NOTES: Comorbidities and outcomes ordered according to descending values in the largest database (IQVIA-OpenClaims). Comorbidities and outcomes are shown if meeting the minimum count required (5 individuals) in each database and cohort. Outcomes not shown due to data not available: all outcomes in CU-AMC-HDC, occurence of death in CUIMC (influenza cohort) and IQVIA-OpenClaims, intensive services in CUIMC.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



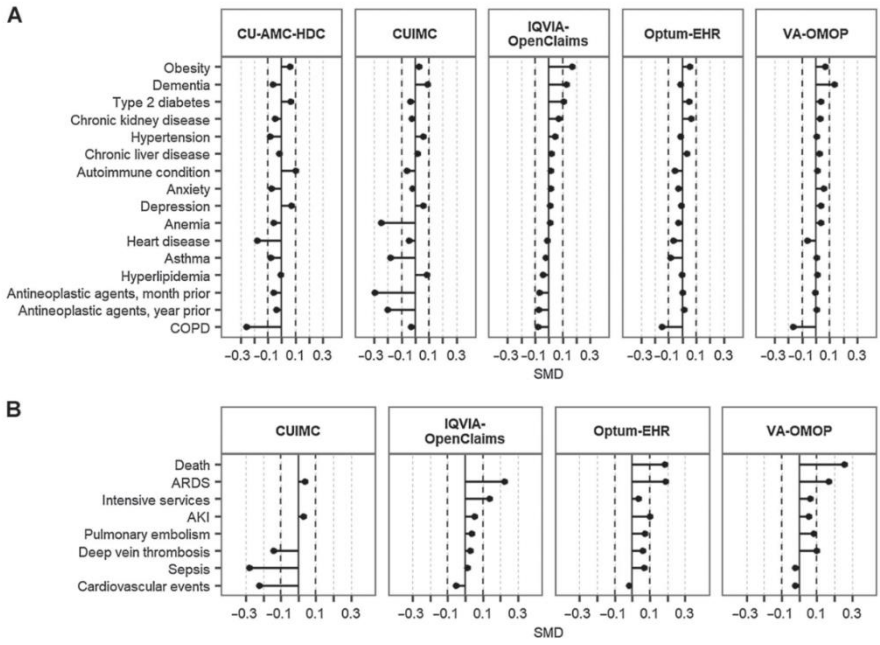



**Figure 4.**

SMDs of selected baseline comorbidities (**A**) and 30-day outcomes (**B**) between patients with a history of cancer hospitalized with COVID-19 and with seasonal influenza. SMD<0 indicates that the prevalence was greater in patients with seasonal influenza, SMD>0 indicates that the prevalence was greater in patients hospitalized. NOTES: Comorbidities and outcomes ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Black-dotted lines indicate an |SMD| of 0.1. SMD calculated for comorbidities and outcomes meeting the minimum count required (5 individuals) in each database and cohort. Outcomes not shown due to data not available: all outcomes in CU-AMC-HDC, occurence of death in CUIMC and IQVIA-OpenClaims, intensive services in CUIMC.

In both COVID-19 cohorts, the most frequent cancer types were prostate cancer and breast cancer; hematologic malignancies were also frequent. The proportion of patients that had received anticancer therapies the year or the month prior was similar in both cohorts. Comorbidities were common in both cohorts but were higher among those hospitalized. Occurrence of death ranged from 2% to 14% among those diagnosed and from 6% to 26% among those hospitalized. When compared with patients with cancer history hospitalized with seasonal influenza, patients hospitalized with COVID-19 had a similar distribution of age and comorbidities but had more severe outcomes.

In the United States, the lifetime cancer prevalence is 5% (data on the lifetime cancer prevalence in Spain is unavailable to our knowledge; ref. 20), which is lower than our findings in patients with COVID-19 (range 4%–25% in the diagnosed and 11%–40% in the hospitalized cohort). Although comparisons are limited due to different cancer definitions, these prevalences are also higher than prior reports on patients with COVID-19 at hospital settings, with cancer prevalences of 6% to 11% in studies from Europe and the United States (21–24). A Danish study, however, found a lifetime cancer prevalence among patients hospitalized with COVID-19 of 17%, in line with our results (6).

The most lifetime-prevalent cancer types in the United States are prostate and breast cancer (20). These cancer types were also those more frequent in our COVID-19 cohorts. However, hematologic malignancies were more frequent than expected in all our cohorts. For example, in the COVID-19 hospitalized cohort, NHL, leukemia, and multiple myeloma were among the third, fifth, and tenth most common cancers, respectively. However, in the U.S. cancer survivors' population, NHL is only the fifth/sixth most frequent (men and women, respectively), whereas leukemia is the ninth in men. The overrepresentation of hematologic malignancies in both COVID-19 cohorts raises questions on whether patients with these malignancies are more exposed or more vulnerable to SARS-CoV-2 infection, or both. Prior studies have reported a higher incidence of COVID-19 infection and (25, 26), more worryingly, an increased risk of COVID-19 complications in patients with hematologic malignancies compared with patients with other cancers (5, 25).

We also found that the proportions of patients that had received antineoplastic and immunomodulating agents the year or the month

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



prior to the index date were similar in both the diagnosed and the hospitalized cohorts. Although this suggests that recent cancer therapies might not be associated with increased COVID-19 severity, this finding must be interpreted with caution due to the overlap between cohorts. However, two studies including over 800 and 900 patients with cancer [from the UK Coronavirus Cancer Monitoring Project (UKCCMP) and the COVID-19 and Cancer Consortium (CCC19), respectively] found no association between cancer therapies and increased COVID-19-related mortality (4, 27).

As expected, patients with cancer history were older and had more comorbidities than overall COVID-19 cases. In a meta-analysis comprising 12,149 COVID-19 cases (mostly hospitalized), hypertension (23%), heart failure (20%), and diabetes (12%) were the most common comorbidities (28). These numbers are substantially lower than our findings. Compared with studies describing patients with cancer, we also found higher prevalences of comorbidities. For example, chronic kidney disease (range 20%–44%), diabetes (24%–59%), and obesity (26%–60%) were higher in our hospitalized cohort than in a study including COVID-19 inpatients with a history of solid cancer (16%, 22% and 10% had chronic kidney disease, diabetes and obesity, respectively; ref. 22). In addition, heart disease, chronic kidney disease, and type 2 diabetes were meaningfully higher among those hospitalized compared with those diagnosed. These conditions have been previously reported as potential risk factors for hospitalization, increased severity, and mortality among COVID-19 cases (29). Comorbidities should be taken into consideration when designing future studies assessing the effect of cancer on COVID-19-related health outcomes, as failing to adjust for some comorbidities or adjusting for others (over-adjustment) could lead to confounding and/or selection bias.

In June 2020, the case-fatality ratio among confirmed COVID-19 cases was 11% in Spain and 5% in the United States (30), which is lower than the all-cause mortality observed in both cohorts in SIDIAP, CUIMC, and VA-OMOP. Undoubtedly, increased age and underlying comorbidities play a substantial role in COVID-19-related mortality among these patients. However, mortality was remarkably lower in the database including cases as of October 2020, Optum-EHR (2% in patients diagnosed, 6% in patients hospitalized). These suggest that studies from the beginning of the pandemic, when testing was limited, might have overestimated mortality rates in patients with COVID-19, including those with cancer. For instance, a meta-analysis including studies prior to July 2020, with data over 18,000 patients with cancer with COVID-19 (mostly inpatients), reported a pooled case mortality rate of 25.6% (95% CI, 22.0%–29.5%; ref. 31), which is in line with our results in the hospitalized cohort in CUIMC (26%) and VA-OMOP (18%) but higher than results in Optum-EHR.

Finally, we compared patients with cancer history hospitalized with COVID-19 to those with seasonal influenza as a benchmark. We previously showed that patients with COVID-19 are more often male, younger and less likely to have respiratory and cardiovascular diseases than patients with influenza (14). Interestingly, patients with COVID-19 and influenza with a history of cancer had a similar sex and age distribution and were of comparable health status. Despite this similarity, patients with cancer history and COVID-19 had a higher occurrence of adverse outcomes than those with influenza.

This study has several strengths, such as its large size. We have reported in a publicly available website more than 10,000 characteristics from over 300,000 and 100,000 patients diagnosed and hospitalized with history of cancer and COVID-19, respectively, using eight different databases. The diverse healthcare settings and populations described, together with our multinational approach, increase the generalizability of our findings. Further, we expect that more databases from additional countries will provide sufficient data on the cancer population as the pandemic evolves. By including only individuals with at least 1 year of observation time available, we have comprehensively captured baseline comorbidities, which could explain the higher prevalence of comorbidities in our cohorts. In addition, we ensured confidentiality throughout the study using a federated analysis approach. Finally, for the purposes of transparency and reproducibility, our methods, tools, and results are all publicly available.

However, this study also has limitations. First, we were not able to provide detailed cancer information, such as year of cancer diagnosis, nor identify patients with active cancer treatment; although we had information on the use of antineoplastic agents during the year and month prior to the index date. Second, by including patients with a clinical COVID-19 diagnosis we might have incurred some false positives. However, we used a broad COVID-19 definition to reduce selection bias due to testing restrictions during the first months of the pandemic (32), as well as (hypothetical) differential patterns in testing between patients with cancer versus patients without cancer. In addition, we did not have information on socioeconomic status, ethnicity, nor race in most databases. We also lacked information on the cause of death and reported instead all-cause death. Third, the overlap between the diagnosed and hospitalized COVID-19 cohorts might have masked some differences in the prevalence of comorbidities between cohorts. Moreover, some patients might be included in more than one database (e.g., in a hospital-based and claims-based database from the United States). Unfortunately, we were unable to determine the degree of overlap across data sources because patient-level data was not shared for confidentiality purposes. Fourth, the differences found in the COVID-19/seasonal influenza comparison may have been influenced by temporal changes in clinical practice standards and coding. Further, the influenza vaccine likely contributed to the low frequency of adverse events among influenza patients. Fifth, the use of routinely collected data could have led to an underestimation of the lifetime cancer prevalence, cancer types, comorbidities, and outcomes due to incomplete reporting. Finally, our findings were heterogenous across data sources. Heterogeneity is a known phenomenon when using real-world data that reflects the existence of different coding practices, observation period, healthcare settings, and populations. Although the interpretation of heterogeneous results is challenging, these also provide valuable insights into the particularities of each setting. Yet, despite this heterogeneity, we found consistent patterns when comparing characteristics across cohorts, which lends credence to our results.

This in-depth characterization revealed that patients with COVID-19 with a history of cancer are mostly aged above 65 years old and have multiple comorbidities that may explain the high frequency of severe COVID-19 outcomes in this population. In addition, we found that hematological malignancies were more frequent than expected. These findings are foundational for guiding future studies and highlight the importance of protecting patients with cancer while guaranteeing cancer care continuity during the pandemic.

## Authors' Disclosures

A.G. Sena reports employment with Janssen R&D and Johnson & Johnson stock ownership. C. Blacketer reports other support from Janssen Research & Development during the conduct of the study, as well as other support from Janssen Research & Development outside the submitted work. S.L. DuVall reports grants from Genomic

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266



Health, Inc., Gilead Sciences Inc., GlaxoSmithKline PLC, Innocrin Pharmaceuticals Inc., Janssen Pharmaceuticals, Inc., Kantar Health, Myriad Genetic Laboratories, Inc., Novartis International AG, and Parexel International Corporation outside the submitted work. A. Golozar reports other support from Regeneron Pharmaceuticals outside the submitted work. L. Hester reports other support from Janssen R&D, LLC outside the submitted work. G. Hripcsak reports grants from NIH during the conduct of the study, as well as grants from Janssen Research outside the submitted work. K. Natarajan reports grants from NIH during the conduct of the study. F. Nyberg reports other support from AstraZeneca outside the submitted work. J.D. Posada reports grants from National Library of Medicine during the conduct of the study. V. Subbian reports grants from National Science Foundation, State of Arizona, Arizona Board of Regents, Agency for Healthcare Research and Quality, and NIH outside the submitted work. M.A. Suchard reports grants from U.S. Department of Veterans Affairs during the conduct of the study, as well as grants from U.S. NIH and IQVIA and personal fees from Janssen Research and Development outside the submitted work. P.B. Ryan reports employment with Janssen Research and Development and ownership of Johnson & Johnson stock. D. Prieto-Alhambra reports grants and other support from Amgen; grants, nonfinancial support, and other support from UCB Biopharma; and grants from Les Laboratoires Servier outside the submitted work; in addition, Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programs organized by DPA's department and open for external participants. K. Kostka reports other support from IQVIA and grants from NIH outside the submitted work. No disclosures were reported by the other authors.

## Authors' Contributions

**E. Roel:** Conceptualization, visualization, writing–original draft, writing–review and editing. **A. Pistillo:** Formal analysis, visualization, writing–review and editing. **M. Recalde:** Writing–review and editing. **A.G. Sena:** Data curation, formal analysis, writing–review and editing. **S. Fernández-Bertolín:** Formal analysis, writing–review and editing. **M. Aragón:** Data curation, writing–review and editing. **D. Puente:** Writing–review and editing. **W.U.R. Ahmed:** Writing–review and editing. **H. Alghoul:** Writing–review and editing. **O. Alser:** Writing–review and editing. **T.M. Alshammari:** Writing–review and editing. **C. Areia:** Writing–review and editing. **C. Blacketer:** Data curation, writing–review and editing. **W. Carter:** Data curation, writing–review and editing. **P. Casajust:** Writing–review and editing. **A.C. Culhane:** Writing-review and editing. **D. Dawoud:** Writing–review and editing. **F. DeFalco:** Data curation, writing–review and editing. **S.L. DuVall:** Data curation, writing–review and editing. **T. Falconer:** Data curation, writing–review and editing. **A. Golozar:** Writing–review and editing. **M. Gong:** Writing–review and editing. **L. Hester:** Writing–review and editing. **G. Hripcsak:** Data curation, writing–review and editing. **E.H. Tan:** Writing–review and editing. **H. Jeon:** Writing–review and editing. **J. Jonnagaddala:** Writing–review and editing. **L.Y.H. Lai:** Writing–review and editing. **K.E. Lynch:** Data curation, writing–review and editing. **M.-E. Matheny:** Data curation, writing–review and editing. **D.R. Morales:** Writing-review and editing. **K. Natarajan:** Data curation, writing–review and editing. **F. Nyberg:** Writing–review and editing. **A. Ostropolets:** Data curation, writing–review and editing. **J.D. Posada:** Data curation, formal analysis, writing–review and editing. **A. Prats-Uribe:** Conceptualization, formal analysis, writing–review and editing. **C.G. Reich:** Data curation, writing–review and editing. **D.R. Rivera:** Writing–review and editing. **L.M. Schilling:** Data curation, writing–review and editing. **I. Soerjomataram:** Writing–review and editing. **K. Shah:** Writing–review and editing. **N.H. Shah:** Data curation, writing–review and editing. **Y. Shen:** Writing–review and editing. **M. Spotniz:** Data curation, writing–review and editing. **V. Subbian:** Writing–review and editing. **M.A. Suchard:** Writing–review and editing. **A. Trama:** Writing–review and editing. **L. Zhang:** Writing–review and editing. **Y. Zhang:** Writing–review and editing. **P.B. Ryan:** Conceptualization, writing–review and editing. **D. Prieto-Alhambra:** Conceptualization, writing–review and editing. **K. Kostka:** Conceptualization, data curation, formal analysis, writing-review and editing. **T. Duarte-Salles:** Conceptualization, data curation, formal analysis, supervision, writing–review and editing.

## Acknowledgments

We would like to acknowledge the patients who suffered from or died of this devastating disease, and their families and carers. We would also like to thank the health care professionals involved in the management of COVID-19 during these challenging times, from primary care to intensive care units. This project has received support from the European Health Data and Evidence Network (EHDEN) project. EHDEN has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no. 806968. The JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. This research received partial support from the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), U.S. National Institutes of Health (NIH), U.S. Department of Veterans Affairs, Janssen Research & Development, and IQVIA. This work was also supported by the Bio Industrial Strategic Technology Development Program (20001234) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no.: H116C0992); Bill & Melinda Gates Foundation (INV-016201); and the Direcció General de Recerca i Innovació en Salut from the Department of Health of the Generalitat de Catalunya. E. Roel is supported by Instituto de Salud Carlos III (grant no. CM20/00174). W.U.R. Ahmed reports funding from the NIHR Oxford Biomedical Research Centre (BRC), Aziz Foundation, Wolfson Foundation, and the Royal College of Surgeons of England. A.G. Sena and L. Hester are employed by Janssen Research and Development, LLC. J. Jonnagaddala reports grants from the National Health and Medical Research Council (APP1192469). M. Gong and Y. Zhang report grants from National Key Research & Development Program of China (Project No. 2018YFC0116901). V. Subbian reports funding from the National Science Foundation, Agency for Health Care Research & Quality, National Institute of Health, and the State of Arizona. A. Prats-Uribe is supported by Fundacion Alfonso Martin Escudero and the Medical Research Council (grant nos. MR/K501256/1, MR/N013468/1). S.L. DuVall, K.E. Lynch, and M.E. Matheny report funding from the Department of Veterans Affairs (VA) Informatics and Computing Infrastructure (VINCI; VA HSR RES 13-457). No funders had a direct role in this study. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Clinician Scientist Award program, NIHR, Department of Veterans Affairs, or the U.S. Government, NHS, or the Department of Health, England. 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.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 25, 2021; revised April 26, 2021; accepted July 7, 2021; published first July 16, 2021.

## References

1. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol 2020;21:335–7.
2. Yu J, Ouyang W, Chua MLK, Xie C. SARS-CoV-2 transmission in patients with cancer at a tertiary care hospital in Wuhan, China. JAMA Oncol 2020;6:1108–10.
3. Rolston KVI. Infections in cancer patients with solid tumors: a review. Infect Dis Ther 2017;6:69–83.
4. Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. Lancet 2020;6736:1–13.
5. Lee LYW, Cazier JB, Starkey T, Briggs SEW, Arnold R, Bisht V, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. Lancet Oncol 2020;21:1309–16.
6. Reilev M, Kristensen KB, Pottegård A, Lund LC, Hallas J, Ernst MT, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. Int J Epidemiol 2020;49:1468–81.
7. Brar G, Pinheiro LC, Shusterman M, Swed B, Reshetnyak E, Soroka O, et al. COVID-19 severity and outcomes in patients with cancer: a matched cohort study. J Clin Oncol 2020;38:3914–24.
8. Westreich D, van Smeden M, Edwards JK. (OpenSAFELY): the Table 2 fallacy in a study of COVID-19 mortality risk factors. Epidemiology 2021;32:e1–2.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266




9. Tennant P, Murray EJ. The quest for timely insights into COVID-19 should not come at the cost of scientific rigor. Epidemiology 2021;32:e2.
10. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Informatics Assoc 2015;22:553–64.
11. 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.
12. Del Mar Garcia-Gil 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). Inform Prim Care 2012;19:135–45.
13. Datta S, Posada J, Olson G, Li W, O'Reilly C, Balraj D, et al. A new paradigm for accelerating clinical data science at Stanford Medicine. arXiv [Internet] 2020[cited 2020 Aug 20]. Available from: http://arxiv.org/abs/2003.10534.
14. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahão MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun 2020;11:5009.
15. Observational Health Data Sciences and Informatics. The Book of OHDSI [Internet]. 2019[cited 2020 Aug 20]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/.
16. World Health Organization. Cancer Today [Internet]. Int. Agency Res. 2018[cited 2020 Oct 22]. Available from: https://gco.iarc.fr/today/home.
17. Sena A, Kostka K, Schuemie Mjdposada. ohdsi-studies/Covid19CharacterizationCharybdis: Charybdis v1.1.1 - Publication Package. 2020[cited 2020 Dec 3]. Available from: https://zenodo.org/record/4033034.
18. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46:399–424.
19. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput 2009;38:1228–34.
20. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363–85.
21. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. JAMA 2020;323:2052–9.
22. de Azambuja E, Brandão M, Wildiers H, Laenen A, Aspeslagh S, Fontaine C, et al. Impact of solid cancer on in-hospital mortality overall and among different subgroups of patients with COVID-19: a nationwide, population-based analysis. ESMO Open 2020;5:e000947.
23. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. BMJ 2020;369:m1985.
24. Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020;369:m1966.
25. Fillmore NR, La J, Szalat RE, Tuck DP, Nguyen V, Yildirim C, et al. Prevalence and outcome of COVID-19 infection in cancer patients: a national veterans affairs study. J Natl Cancer Inst 2020;00:1–8.
26. Wang Q, Berger NA, Xu R. Analyses of risk, racial disparity, and outcomes among US patients with cancer and COVID-19 infection. JAMA Oncol 2020;44106:1–8.
27. Lee LYW, Cazier JB, Starkey T, Turnbull CD, UK Coronavirus Cancer Monitoring Project Team, Kerr R, et al. COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. Lancet 2020;395:1919–26.
28. Jutzeler CR, Bourguignon L, Weis CV, Tong B, Wong C, Rieck B, et al. Comorbidities, clinical signs and symptoms, laboratory findings, imaging features, treatment strategies, and outcomes in adult and pediatric patients with COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020;37:101825.
29. CDC - COVID-19. Evidence used to update the list of underlying medical conditions that increase a person's risk of severe illness from COVID-19 [Internet]. [cited 2020 Dec 3]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html.
30. World Health Organization. Coronavirus disease (COVID-19): situation report, 162 [Internet]. WHO. 2020. Available from: https://apps.who.int/iris/handle/10665/332970.
31. Saini KS, Tagliamento M, Lambertini M, McNally R, Romano M, Leone M, et al. Mortality in patients with cancer and coronavirus disease 2019: a systematic review and pooled analysis of 52 studies. Eur J Cancer 2020;139:43–50.
32. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun 2020;11:1–12.

Published OnlineFirst July 16, 2021; DOI: 10.1158/1055-9965.EPI-21-0266

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

# Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain

Elena Roel, Andrea Pistillo, Martina Recalde, et al.

*Cancer Epidemiol Biomarkers Prev* Published OnlineFirst July 16, 2021.

## 5.3. Study III: 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

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.

*International Journal of Cancer*. 2022;150(5):782-794
Doi: 10.1002/ijc.33846

Received: 12 July 2021 | Revised: 31 August 2021 | Accepted: 20 September 2021
DOI: 10.1002/ijc.33846

CANCER EPIDEMIOLOGY

IJC INTERNATIONAL JOURNAL of CANCER

# 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

Elena Roel[1,2] | Andrea Pistillo[1] | Martina Recalde[1,2] | Sergio Fernández-Bertolín[1] | María Aragón[1] | Isabelle Soerjomataram[3] | Mazda Jenab[3] | Diana Puente[1] | Daniel Prieto-Alhambra[4] | Edward Burn[1,4] | 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, Barcelona, Spain

[3]International Agency for Research on Cancer (IARC-WHO), Lyon, France

[4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK

**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, 08007 Barcelona, Spain.
Email: tduarte@idiapjgol.org

**Funding information**
This project was funded by the Health Department from the Generalitat de Catalunya with a grant for research projects on SARS-CoV-2 and COVID-19 disease organised by the Direcció General de Recerca i Innovació en Salut. This project has also received support from the European Health Data and Evidence Network (EHDEN) project. EHDEN received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The University of Oxford received a grant related to this work from the Bill & Melinda Gates Foundation (Investment ID INV-016201), and partial support from the UK National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. Elena Roel was supported by Instituto de Salud Carlos III (grant number CM20/00174). Daniel

## Abstract

The relationship between cancer and coronavirus disease 2019 (COVID-19) infection and severity remains poorly understood. We conducted a population-based cohort study between 1 March and 6 May 2020 describing the associations between cancer and risk of COVID-19 diagnosis, hospitalisation and COVID-19-related death. Data were obtained from the Information System for Research in Primary Care (SIDIAP) database, including primary care electronic health records from ~80% of the population in Catalonia, Spain. Cancer was defined as any primary invasive malignancy excluding non-melanoma skin cancer. We estimated adjusted hazard ratios (aHRs) for the risk of COVID-19 (outpatient) clinical diagnosis, hospitalisation (with or without a prior COVID-19 diagnosis) and COVID-19-related death using Cox proportional hazard regressions. Models were estimated for the overall cancer population and by years since cancer diagnosis (<1 year, 1-5 years and ≥5 years), sex, age and cancer type; and adjusted for age, sex, smoking status, deprivation and comorbidities. We included 4 618 377 adults, of which 260 667 (5.6%) had a history of cancer. A total of 98 951 individuals (5.5% with cancer) were diagnosed, and 6355 (16.4% with cancer) were directly hospitalised with COVID-19. Of those diagnosed, 6851 were subsequently hospitalised (10.7% with cancer), and 3227 died without being hospitalised (18.5% with cancer). Among those hospitalised, 1963 (22.5% with cancer) died. Cancer was associated with an increased risk of COVID-19 diagnosis (aHR: 1.08;

**Abbreviations:** 95% CI, 95% confidence interval; aHR, adjusted hazard ratio; BIC, Bayesian information criterion; CDM, Common Data Model; CI, cumulative incidence; COVID-19, coronavirus disease 2019; GP, general practitioner; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification; OHDSI, Observational Health Data Sciences and Informatics; OMOP, Observational Medical Outcomes Partnership; RT-PCR, reverse transcription polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SIDIAP, Information System for Research in Primary Care; SMD, standardised mean difference.

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

Prieto-Alhambra is funded through a National Institute for Health Research (NIHR) Senior Research Fellowship (Grant number SRF-2018-11-ST2-004). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. The funders of the study had no role in study design, data collection, analysis and interpretation, or writing of the report.

95% confidence interval [1.05-1.11]), direct COVID-19 hospitalisation (1.33 [1.24-1.43]) and death following hospitalisation (1.12 [1.01-1.25]). These associations were stronger for patients recently diagnosed with cancer, aged <70 years, and with haematological cancers. These patients should be prioritised in COVID-19 vaccination campaigns and continued non-pharmaceutical interventions.

**KEYWORDS**
cancer, COVID-19, electronic health record, fatality, SARS-CoV-2

**What's new?**

Studies addressing associations between cancer and severity of coronavirus disease 2019 (COVID-19) have focused primarily on hospitalized patients. Findings have been inconsistent, however, owing to varying cancer criteria, lack of representative samples, and other factors. Here, the natural history of COVID-19 in cancer patients during the first wave of the pandemic in 2020 in Spain was investigated in a large, representative cohort with a heterogenous cancer population. Patients with cancer were at increased risk of severe COVID-19. Risk was notably high among those over age 70 and those with recent cancer diagnosis, hematological cancer, or lung and bladder cancer.

## 1 | INTRODUCTION

Cancer is a leading cause of morbidity and death worldwide, with an estimated 19 million new cases and 10 million deaths in 2020.[1] Patients with cancer are often older and have multiple comorbidities and an impaired immunity due to the cancer itself and cancer therapies, thus increasing their susceptibility to infections.[2] As a result, patients with cancer have been considered a high-risk population for the novel coronavirus disease 2019 (COVID-19) since the beginning of the pandemic.[3] This disease, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), manifests with a varying degree of severity, ranging from asymptomatic to severe disease and death.[4]

Although there are a substantial number of publications addressing the relationship between cancer and COVID-19, these have shown conflicting results.[5] Some studies have found that patients with cancer have an increased risk of COVID-19 infection, hospitalisation and death compared to patients without cancer,[6-9] whereas others have reported null associations.[10-12] The majority of these studies were small, used different criteria to identify patients with cancer (eg, only active cancers, or solid cancers) and did not include representative samples (ie, restricted to hospital and/or laboratory-confirmed cases), which limits the generalizability of their findings and increases the risk of selection bias.[13]

Patients with cancer are a highly heterogeneous population that encompasses patients with different features, such as cancer type or phases of care since time of diagnosis (eg, under active treatment, active surveillance or cured). Understanding which patients with cancer are at the highest risk of COVID-19-infection of poor outcomes is essential to inform clinical care and to guide prevention strategies targeting this population. A large, population-based cohort study that includes a heterogeneous cancer population and that captures both COVID-19 incidence and COVID-19-related outcomes could address the limitations of the previous evidence. In our study, we aimed to describe the associations between cancer and the risks of COVID-19 diagnosis, hospitalisation with COVID-19 and COVID-19-related death, overall and by different population subgroups, using real-world data from Catalonia, Spain.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design, setting and data sources

We conducted a population-based cohort study from 1 March 2020 until 6 May 2020 (last date of data available), using data from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org), a primary care database from Catalonia, a north-eastern region in Spain. Spain has a universal primary care-based health system, in which general practitioners (GPs) are the first point of contact for care. As a consequence, GPs have diagnosed and managed the majority of COVID-19 cases since the beginning of the pandemic.[14] In addition, because GPs are responsible of issuing sick leaves, patients diagnosed with COVID-19 in other settings (eg, hospital emergency departments) were also bound to contact primary care providers during study follow-up.

The SIDIAP database includes anonymized primary care electronic health records collected since 2006 covering approximately six million people (80% of the population in Catalonia, Spain) and is representative in terms of age, sex and geographic distribution.[15] SIDIAP includes data on demographics, lifestyle information and disease diagnoses, among others and has been linked to SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) test results and hospital records (both from the public sector), as well as to regional

mortality data through unique ID linkage. In addition, SIDIAP has been mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), allowing us to apply common analytical tools developed by the open-science Observational Health Data Sciences and Informatics (OHDSI) network.[16]

## 2.2 | Study participants

We included all adults (aged 18 years or older) registered in the SIDIAP database as of 1 March 2020 (index date for all participants) with at least 1 year of prior history observation available. We excluded patients who had a record of a secondary cancer before a record of a primary cancer, patients with a clinical diagnosis or positive test result for COVID-19 prior to index date and patients hospitalised or living in a nursing home at index date (to include only patients representative of the community population).

## 2.3 | Multistate framework

To address our objectives, we employed a multistate framework that we have previously utilised to describe the risks of COVID-19 diagnosis, hospitalisation and death.[17] Multistate models can be used to describe processes where individuals transition from one health status to another, while separating baseline risk and covariate effects associated with each transition.[18] In our study, individuals started the follow-up at the general population and then could transition to three other states: diagnosed with COVID-19 (in an outpatient setting), hospitalised with COVID-19 and death. Six different transitions were possible: from the general population to either diagnosed with COVID-19, hospitalised with COVID-19 (ie, direct hospitalisation) or death; from diagnosed to either hospitalised with COVID-19 or death and from hospitalised with COVID-19 to death (Figure 1). We used this approach to provide a more comprehensive overview of patient's interactions with the health system, taking into account those who seek primary and hospital care.

For all the transitions, individuals were followed until the occurrence of a state of interest, the occurrence of a competing event or the end of the study period (6 May 2020). Because we were solely interested in COVID-19-related outcomes, we did not model the transition from the general population to death. However, we reported deaths occurring in the general population, which were considered as a competing event.

## 2.4 | Variables

The exposure of interest was cancer, which we defined as any diagnosis of a primary invasive solid or haematological cancer, excluding non-melanoma skin cancer, prior to the index date. We used the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) to identify cancer diagnoses: C00 to C96, except C44 (non-melanoma skin cancer) and C77-C79 (secondary cancers). Cancer types by anatomical location were identified using definitions previously validated in the SIDIAP database.[19] To avoid misclassification of primary cancers, we only considered the earliest cancer type registered for each patient. We stratified patients with cancer according to the number of years since the diagnosis to the index date into three groups (<1 year, 1-5 years and ≥5 years), because we lacked information on cancer status (ie, active, in remission) and cancer therapies. By doing this, we assumed that those diagnosed with cancer <5 years prior to the index date were more likely to have an active cancer and/or an ongoing cancer treatment (especially those diagnosed within 1 year prior), whereas those diagnosed ≥5 years prior would be mostly cancer survivors.

The covariates of interest were sex, age, smoking status, deprivation and comorbidities. We extracted participants' sex and age at

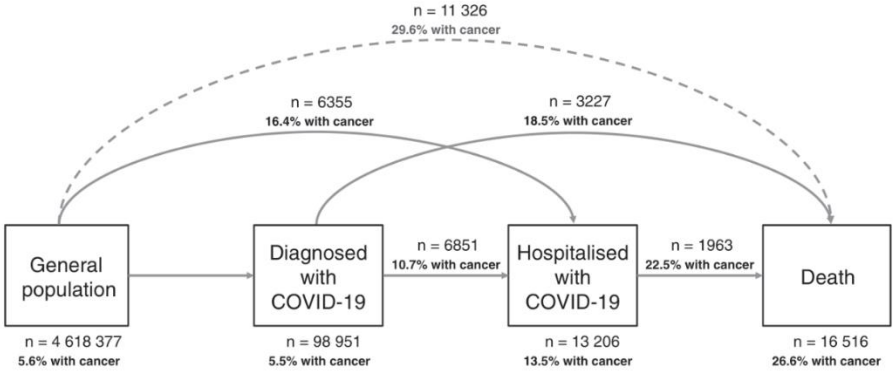



FIGURE 1 Overview of the multistate model used in our study

index date. Smoking status (never, former or current smoker) was assigned as the closest assessment to the index date recorded. Deprivation was assessed using the MEDEA deprivation index, which is calculated at the census tract level in urban areas of Catalonia.[20] MEDEA deprivation index is categorised in quintiles, with the first quintile representing the least deprived group and the fifth the most deprived. It also includes a rural category for individuals living in rural areas. Our comorbidities of interest were autoimmune conditions, chronic kidney disease, chronic obstructive pulmonary disease, dementia, heart disease, hyperlipidaemia, hypertension, obesity and type-2 diabetes. Comorbidities were defined as previously described based on medical diagnosis[17] and selected due to their relevance to the COVID-19 research field.[21] The definitions for each comorbidity can be consulted in a web application ("Index Event Breakdown" tab) available at https://livedataoxford.shinyapps.io/MultiStateCovidCohorts/.

Our outcomes of interest were an outpatient clinical diagnosis of COVID-19, a hospitalisation with COVID-19 and COVID-19-related death. We defined COVID-19 diagnoses based on a recorded clinical code for COVID-19 disease (ICD-10-CM: B34.2; B97.29). We did not require a positive RT-PCR test result in the definition of COVID-19 diagnoses due to testing restrictions during the first months of the pandemic.[17] For instance, at that time, tests were exclusively available at the hospital level, and only patients with severe symptoms and/or with underlying conditions were tested. We defined hospitalisation with COVID-19 as a hospital admission (with at least one-day hospital stay) where the patient had a COVID-19 diagnosis or a positive RT-PCR test result 21 days prior to admission up to 3 days after admission (to allow for a delay in diagnosis and minimise the risk of including hospital-acquired COVID-19 infections). We extracted deaths (from any cause) from region-wide mortality data, and by doing so, we included both deaths during hospitalisation and in the community. Deaths occurring following a COVID-19 event (diagnosis or hospitalisation) were considered as COVID-19-related deaths.

### 2.5 | Statistical analyses

We described participants' baseline characteristics, participants' time at risk at each state and numbers of events observed for each transition by cancer status (with or without cancer). To assess the relationship between the cancer and the risk of transitioning to a subsequent state in the multistate model, we estimated adjusted cause-specific hazard ratios (aHRs), with 95% confidence intervals (CIs), using Cox proportional hazard regressions for each transition.

First, we estimated models for all patients with cancer compared to patients without cancer adjusting for age, sex, the MEDEA deprivation index, smoking status and all the comorbidities of interest (main models). We used a directed acyclic graph to guide decisions on the control for confounding (Figure S1).[22] To check the proportional hazard assumptions for the variables included in the models, we visually inspected log-log survival curves. Missing data were handled as an additional category. Non-linearity in age and risks of transition was considered by fitting models with age as a linear term, with a polynomial of degree 2 (ie, quadratic), and with restricted cubic splines (with three, four or five knots).[23] We calculated the Bayesian information criterion (BIC) for each of those models, and we selected the models with the lowest BIC values.

Second, we estimated the relationship between cancer and COVID-19 outcomes adjusting for age and sex; and adjusting for age, sex, the MEDEA deprivation index and smoking status. Third, we further estimated our main models separately for <1-year, 1 to 5-year and ≥5-year cancer patients and stratified these models by sex (women or men), age (<70 and ≥70 years, 70 years was the median age of patients with cancer) and cancer type (haematological or solid cancer, as well as by solid cancer types). All models were relative to patients without cancer (cancer-free).

As sensitivity analyses, we re-estimated our main models: (a) stratifying by calendar time for transitions in which the proportionality assumption was violated, (b) restricting participants to never smokers, to avoid residual confounding by smoking and (c) after performing a multiple imputation of missing data (smoking status and MEDEA deprivation index) using predictive mean matching, with five imputations drawn. We also compared baseline characteristics of patients with and without missing data using standardised mean differences (SMD). We considered SMD $\geq|0.1|$ as a meaningful difference in the distribution of a given characteristic between the two groups.[24]

We used R version 3.6 for data analysis and visualisation. The R packages used in the analysis included mstate[25] and rms.[26] The analytic code is available at https://github.com/SIDIAP/COVID-19-cancer-multi-state.

## 3 | RESULTS

### 3.1 | Population included

A total of 4 618 377 adults were included. We excluded 104 022 individuals with less than a year of prior observation history; 1496 with a record of a secondary cancer before a record of a primary cancer, 303 with a COVID-19 diagnosis or positive SARS-CoV-2 test before index date, 40 421 living in a nursing home and 1138 hospitalised at the index date (Figure S2). Baseline characteristics of the population included are summarised in Table 1. In total, 260 667 (5.6%) patients had a prior diagnosis of cancer. Of these, 167 053 (64.1% of the cancer population) were diagnosed ≥5 years, 72 033 (27.6%) 1 to 5 years and 21 581 (8.3%) <1 year prior to the index date. Compared to cancer-free patients, those with cancer were older, more frequently former smokers and living in the least deprived areas of Catalonia. In addition, they had a higher burden of comorbidities, especially cardiovascular conditions (eg, 27.4% had heart disease vs 10.2% in cancer-free patients). When stratifying patients by age categories, we observed that the burden of comorbidities increased with age for both groups (Figure S3). Among patients with cancer, 239 030 (91.7%) and 21 637 (8.3%) had a solid and haematological cancer, respectively. The most frequent solid cancer types were breast (n = 58 611,

**TABLE 1** Baseline characteristics of the population included, by cancer status

| | Total population | Without cancer | With cancer | | | |
|---|---|---|---|---|---|---|
| | | | Overall | ≥5 years[a] | 1-5 years[a] | <1 year[a] |
| n | 4 618 377 | 4 357 710 | 260 667 | 167 053 | 72 033 | 21 581 |
| Age (median [IQR]) | 48 [36.0, 63.0] | 47 [35.0, 61.0] | 70 [59.0, 78.0] | 71 [61.0, 79.0] | 67 [57.0, 76.0] | 66 [56.0, 76.0] |
| Age categories (%) | | | | | | |
| 18-39 | 1 437 236 (31.1) | 1 427 705 (32.8) | 9531 (3.7) | 5555 (3.3) | 2974 (4.1) | 1002 (4.6) |
| 40-59 | 1 785 495 (38.7) | 1 727 443 (39.6) | 58 052 (22.3) | 32 909 (19.7) | 19 019 (26.4) | 6124 (28.4) |
| 60-69 | 615 198 (13.3) | 553 838 (12.7) | 61 360 (23.5) | 36 999 (22.1) | 18 786 (26.1) | 5575 (25.8) |
| 70-79 | 468 286 (10.1) | 393 504 (9.0) | 74 782 (28.7) | 50 205 (30.1) | 19 197 (26.7) | 5380 (24.9) |
| 80 or older | 312 162 (6.8) | 255 220 (5.9) | 56 942 (21.8) | 41 385 (24.8) | 12 057 (16.7) | 3500 (16.2) |
| Sex, female (%) | 2 361 230 (51.1) | 2 226 424 (51.1) | 134 806 (51.7) | 89 473 (53.6) | 35 060 (48.7) | 10 273 (47.6) |
| MEDEA deprivation index (%) | | | | | | |
| Quintile 1 (least deprived) | 714 183 (15.5) | 668 548 (15.3) | 45 635 (17.5) | 29 662 (17.8) | 12 392 (17.2) | 3581 (16.6) |
| Quintile 2 | 703 921 (15.2) | 662 113 (15.2) | 41 808 (16.0) | 26 971 (16.1) | 11 534 (16.0) | 3303 (15.3) |
| Quintile 3 | 697 074 (15.1) | 656 859 (15.1) | 40 215 (15.4) | 25 893 (15.5) | 11 114 (15.4) | 3208 (14.9) |
| Quintile 4 | 692 844 (15.0) | 654 775 (15.0) | 38 069 (14.6) | 24 488 (14.7) | 10 445 (14.5) | 3136 (14.5) |
| Quintile 5 (most deprived) | 687 062 (14.9) | 653 878 (15.0) | 33 184 (12.7) | 21 149 (12.7) | 9168 (12.7) | 2867 (13.3) |
| Rural | 832 256 (18.0) | 785 356 (18.0) | 46 900 (18.0) | 29 744 (17.8) | 13 073 (18.1) | 4083 (18.9) |
| Missing | 291 037 (6.3) | 276 181 (6.3) | 14 856 (5.7) | 9146 (5.5) | 4307 (6.0) | 1403 (6.5) |
| Smoking status (%) | | | | | | |
| Never smoker | 1 834 657 (39.7) | 1 736 604 (39.9) | 98 053 (37.6) | 64 646 (38.7) | 25 891 (35.9) | 7516 (34.8) |
| Former smoker | 772 875 (16.7) | 695 636 (16.0) | 77 239 (29.6) | 48 635 (29.1) | 22 576 (31.3) | 6028 (27.9) |
| Current smoker | 712 739 (15.4) | 686 159 (15.7) | 26 580 (10.2) | 15 702 (9.4) | 7901 (11.0) | 2977 (13.8) |
| Missing | 1 298 106 (28.1) | 1 239 311 (28.4) | 58 795 (22.6) | 38 070 (22.8) | 15 665 (21.7) | 5060 (23.4) |
| Comorbidities (%) | | | | | | |
| Autoimmune condition | 259 234 (5.6) | 235 347 (5.4) | 23 887 (9.2) | 15 474 (9.3) | 6526 (9.1) | 1887 (8.7) |
| Chronic kidney disease | 201 258 (4.4) | 165 751 (3.8) | 35 507 (13.6) | 24 922 (14.9) | 8339 (11.6) | 2246 (10.4) |
| Chronic obstructive pulmonary disease | 119 532 (2.6) | 98 365 (2.3) | 21 167 (8.1) | 13 281 (8.0) | 6001 (8.3) | 1885 (8.7) |
| Dementia | 42 504 (0.9) | 36 026 (0.8) | 6478 (2.5) | 4817 (2.9) | 1328 (1.8) | 333 (1.5) |
| Heart disease | 516 140 (11.2) | 444 733 (10.2) | 71 407 (27.4) | 47 851 (28.6) | 18 145 (25.2) | 5411 (25.1) |
| Hyperlipidaemia | 505 102 (10.9) | 458 565 (10.5) | 46 537 (17.9) | 30 173 (18.1) | 12 785 (17.7) | 3579 (16.6) |
| Hypertension | 687 358 (14.9) | 610 694 (14.0) | 76 664 (29.4) | 49 254 (29.5) | 21 195 (29.4) | 6215 (28.8) |
| Obesity | 1 144 442 (24.8) | 1 045 689 (24.0) | 98 753 (37.9) | 64 148 (38.4) | 26 800 (37.2) | 7805 (36.2) |
| Type-2 diabetes | 317 005 (6.9) | 275 132 (6.3) | 41 873 (16.1) | 26 913 (16.1) | 11 560 (16.0) | 3400 (15.8) |

(Continues)

**TABLE 1** (Continued)

| | Total population | Without cancer | With cancer | | | |
|---|---|---|---|---|---|---|
| | | | Overall | ≥5 years[a] | 1-5 years[a] | <1 year[a] |
| Age at cancer diagnosis, median [IQR] | – | – | 61 [50.3, 70.2] | 59 [48.1, 67.9] | 65 [54.3, 73.6] | 66 [55.5, 75.5] |
| Cancer type [ICD-10-CM code] (%) | | | | | | |
| Haematological | 21 637 (0.5) | | 21 637 (8.3) | 13 657 (8.2) | 6148 (8.5) | 1832 (8.5) |
| Leukaemia [C91-C95] | 7402 (0.2) | – | 7402 (2.8) | 4744 (2.8) | 2051 (2.8) | 607 (2.8) |
| Non-Hodgkin lymphoma [C82-C96] | 5111 (0.1) | – | 5111 (2.0) | 3776 (2.3) | 1031 (1.4) | 304 (1.4) |
| Hodgkin's lymphoma [C81] | 2724 (0.1) | – | 2724 (1.0) | 2133 (1.3) | 466 (0.6) | 125 (0.6) |
| Multiple myeloma [C90] | 2249 (0.0) | – | 2249 (0.9) | 1031 (0.6) | 916 (1.3) | 302 (1.4) |
| Other haematological [C96] | 4151 (0.1) | – | 4151 (1.6) | 1973 (1.2) | 1684 (2.3) | 494 (2.3) |
| Solid | 239 030 (5.2) | – | 239 030 (91.7) | 153 396 (91.8) | 65 885 (91.5) | 19 749 (91.5) |
| Breast [C50] | 58 611 (1.3) | – | 58 611 (22.5) | 40 074 (24.0) | 14 725 (20.4) | 3812 (17.7) |
| Prostate [C61] | 37 141 (0.8) | – | 37 141 (14.2) | 24 400 (14.6) | 10 165 (14.1) | 2576 (11.9) |
| Colorectal [C18-C21] | 36 071 (0.8) | – | 36 071 (13.8) | 21 669 (13.0) | 11 415 (15.8) | 2987 (13.8) |
| Bladder [C67] | 20 592 (0.4) | – | 20 592 (7.9) | 12 509 (7.5) | 6293 (8.7) | 1790 (8.3) |
| Skin melanoma [C43] | 12 956 (0.3) | – | 12 956 (5.0) | 8490 (5.1) | 3422 (4.8) | 1044 (4.8) |
| Kidney [C64] | 7911 (0.2) | – | 7911 (3.0) | 4522 (2.7) | 2630 (3.7) | 759 (3.5) |
| Lung [C33-C34] | 7569 (0.2) | – | 7569 (2.9) | 3080 (1.8) | 2948 (4.1) | 1541 (7.1) |
| Corpus uterus [C54-C55] | 7353 (0.2) | – | 7353 (2.8) | 4983 (3.0) | 1855 (2.6) | 515 (2.4) |
| Thyroid [C73] | 6449 (0.1) | – | 6449 (2.5) | 4579 (2.7) | 1500 (2.1) | 370 (1.7) |
| Head and neck [C00-C14] | 5770 (0.1) | – | 5770 (2.2) | 4042 (2.4) | 1323 (1.8) | 405 (1.9) |
| Cervix [C53] | 3979 (0.1) | – | 3979 (1.5) | 3035 (1.8) | 755 (1.0) | 189 (0.9) |
| Ovary [C56] | 3889 (0.1) | – | 3889 (1.5) | 2523 (1.5) | 997 (1.4) | 369 (1.7) |
| Stomach [C16] | 3628 (0.1) | – | 3628 (1.4) | 2210 (1.3) | 995 (1.4) | 423 (2.0) |
| Larynx [C32] | 3317 (0.1) | – | 3317 (1.3) | 2161 (1.3) | 874 (1.2) | 282 (1.3) |
| Brain and central nervous system [C70-C72, C75.1-C75.3] | 3313 (0.1) | – | 3313 (1.3) | 2216 (1.3) | 750 (1.0) | 347 (1.6) |
| Testis [C62] | 2763 (0.1) | – | 2763 (1.1) | 2073 (1.2) | 562 (0.8) | 128 (0.6) |
| Liver [C22] | 2051 (0.0) | – | 2051 (0.8) | 852 (0.5) | 818 (1.1) | 381 (1.8) |
| Bone and cartilage [C40-C41] | 1944 (0.0) | – | 1944 (0.7) | 1458 (0.9) | 371 (0.5) | 115 (0.5) |
| Pancreas [C25] | 1622 (0.0) | – | 1622 (0.6) | 568 (0.3) | 592 (0.8) | 462 (2.1) |
| Oesophagus [C15] | 763 (0.0) | – | 763 (0.3) | 349 (0.2) | 270 (0.4) | 144 (0.7) |
| Gallbladder [C23-C24] | 479 (0.0) | – | 479 (0.2) | 214 (0.1) | 181 (0.3) | 84 (0.4) |
| Other solid | 10 859 (0.2) | – | 10 859 (4.2) | 7389 (4.4) | 2444 (3.4) | 1026 (4.8) |

*Note*: – means not applicable. The MEDEA deprivation index is calculated at the census tract level in urban areas. Other solid cancers include other solid cancers, cancers of unspecified site [C76, C80] and more than one cancer (ie, patients that had more than one cancer recorded on the same date).

Abbreviations: ICD-10-CM, International Classification for Diseases, 10th revision Clinical Modification; IQR, interquartile range.

[a]Years since cancer diagnosis to the index date (1 March 2020).

**TABLE 2** Time at risk, absolute number of events and cumulative incidence, by cancer status

| | | From general population | | | | From diagnosed with COVID-19 | | | | From hospitalised with COVID-19 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | General population | | To diagnosed with COVID-19 | To hospitalised with COVID-19 | To death | | | To hospitalised with COVID-19 | To death | | | | |
| | n | Follow-up (days) Median (min, IQR, max) | Number of events (CI at 67 days) | Number of events (CI at 67 days) | Number of events (CI at 67 days) | n | Follow-up (days) Median (min, IQR, max) | Number of events (CI at 45 days) | Number of events (CI at 45 days) | n | Follow-up (days) Median (min, IQR, max) | Number of events (CI at 45 days) |
| Total population | 4 618 377 | 67 (1, 67 to 67, 67) | 98 951 (2.14%) | 6355 (0.14%) | 11 326 (0.25%) | 98 951 | 36 (0, 20 to 44, 66) | 6851 (7.19%) | 3227 (3.91%) | 13 206 | 37 (0, 27 to 43, 65) | 1963 (17.57%) |
| Patients without cancer | 4 357 710 | 67 (1, 67 to 67, 67) | 93 558 (2.15%) | 5312 (0.12%) | 7970 (0.18%) | 93 558 | 36 (0, 21 to 44, 66) | 6116 (6.79%) | 2631 (3.37%) | 11 428 | 37 (0, 28 to 43, 65) | 1522 (15.71%) |
| Patients with cancer | | | | | | | | | | | | |
| Overall | 260 667 | 67 (1, 67 to 67, 67) | 5393 (2.07%) | 1043 (0.40%) | 3356 (1.29%) | 5393 | 30 (0, 13 to 42, 65) | 735 (14.14%) | 596 (13.39%) | 1778 | 36 (0.5, 22 to 43, 58) | 441 (29.34%) |
| ≥5 years[a] | 167 053 | 67 (2, 67 to 67, 67) | 3464 (2.07%) | 670 (0.40%) | 1714 (1.03%) | 3464 | 30 (0, 13 to 42, 65) | 464 (13.91%) | 379 (13.13%) | 1134 | 36 (0.5, 23 to 43, 58) | 293 (30.55%) |
| 1-5 years[a] | 72 033 | 67 (1, 67 to 67, 67) | 1466 (2.04%) | 268 (0.37%) | 911 (1.27%) | 1466 | 30.5 (0, 13 to 43, 65) | 211 (14.85%) | 149 (12.09%) | 479 | 36 (1, 22 to 43, 58) | 110 (25.75%) |
| <1 year[a] | 21 581 | 67 (2, 67 to 67, 67) | 463 (2.15%) | 105 (0.49%) | 731 (3.39%) | 463 | 24 (0, 10 to 40, 64) | 60 (13.57%) | 68 (20.15%) | 165 | 35 (1, 22 to 42, 58) | 38 (32.64%) |

Abbreviations: CI, cumulative incidence; IRQ, interquartile range.

[a]Years since cancer diagnosis to the index date (1 March 2020).

22.5%), prostate (37 141, 14.2%), colorectal (36 071, 13.8%) and bladder (20 592, 7.9%).

### 3.2 | Occurrence of COVID-19 outcomes

Among the general population, 98 951 (2.1% cumulative incidence [CI] at 67 days) individuals were diagnosed with COVID-19, 6355 (0.1% CI) were directly hospitalised with COVID-19 and 11 326 (0.25% CI) died without a COVID-19 diagnosis/hospitalisation (Figure 1, Table 2). Among individuals diagnosed with COVID-19, 6851 (7.2% CI at 45 days) were hospitalised and 3227 (3.9% CI) died without a hospitalisation. Among those hospitalised, 1963 (18% CI at 45 days) died. Among the total cancer population (n = 260 667), 5393 (2.1% CI at 67 days) patients were diagnosed with COVID-19, 1043 (0.4%) were directly hospitalised with COVID-19 and 3356 (1.3%) died without a COVID-19 diagnosis/hospitalisation. Among those diagnosed with COVID-19, 735 (14.1% CI at 45 days) were subsequently hospitalised and 596 (13.4%) died without a hospitalisation. Among those hospitalised, 441 (29.3% CI at 45 days) died. Descriptive characteristics by state and transition are shown in Table S1. In brief, individuals diagnosed/hospitalised with COVID-19, as well as having a COVID-19-related death, were older, more frequently male and former smokers, and had more comorbidities than the general population.

### 3.3 | Risks of COVID-19 diagnosis, hospitalisation and death among patients with cancer

Compared to cancer-free patients, those with cancer had an increased risk of COVID-19 diagnosis (overall aHR: 1.08; 95% CI [1.05-1.11]), direct COVID-19 hospitalisation (1.33 [1.24-1.43]) and death following a COVID-19 hospitalisation (1.12 [1.01-1.25]) (Figure 2). Models using different adjustment strategies showed similar results to our main models (Figure S4).

In models stratified by years since cancer diagnosis, the risk of COVID-19 diagnosis was similar in <1-year, 1 to 5-year and ≥5-year cancer patients (Figure 2). As for the risk of direct COVID-19 hospitalisation, <1-year cancer patients had the highest risk (1.84 [1.52-2.23]), followed by 1 to 5-year cancer patients (1.32 [1.17-1.50]) and ≥5-year cancer patients (1.27 [1.17-1.38]). Increased risk of COVID-19-related death remained significant only in <1-year cancer patients, for both deaths following a COVID-19 diagnosis (1.81[1.42-2.31]) and following a COVID-19 hospitalisation (1.63 [1.18-2.26]).

Overall, in models stratified by sex, the associations between cancer and risk of COVID-19 diagnosis and death (following a diagnosis/hospitalisation) were moderately stronger in men, whereas the associations with risk of direct hospitalisation were moderately stronger in women (Figure 3, Table S2). In models stratified by age, we found a stronger association between cancer and COVID-19 outcomes in the subgroup of patients aged <70 years compared to those aged

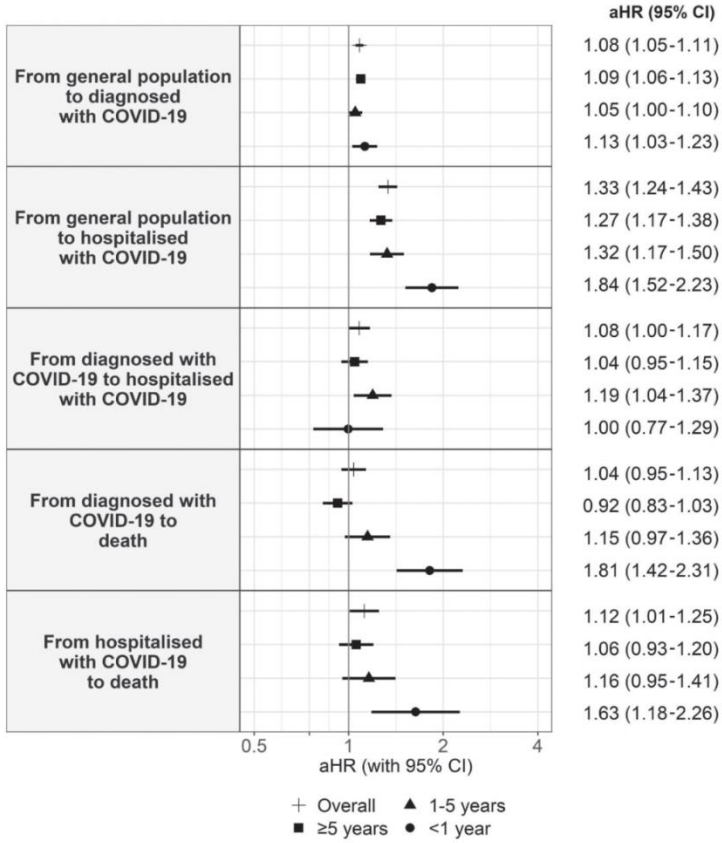



**FIGURE 2** Adjusted hazard ratios of COVID-19 outcomes in patients with cancer compared to patients without cancer, overall and by years since cancer diagnosis. Models are adjusted for age, sex, the MEDEA deprivation index, smoking status and comorbidities (autoimmune conditions, chronic kidney disease, chronic obstructive pulmonary disease, dementia, heart disease, hyperlipidaemia, hypertension, type-2 diabetes and obesity). aHR, adjusted hazard ratio; CI, confidence interval

≥70 years, aside from the risk of COVID-19 diagnosis (Figure 3, Table S3). Age differences were more pronounced in <1-year cancer patients. In addition, the associations between cancer and COVID-19-related death (either following a COVID-19 diagnosis or a hospitalisation) were only significant in the subgroup of patients aged <70 years. For example, the overall aHR for death following hospitalisation was 1.49 (1.10-2.01) in <70-year patients and 1.07 (0.95-1.20) in ≥70-year patients. In <1-year cancer patients, the aHR was 4.58 (2.47-8.50) in <70-year patients and 1.30 (0.88-1.90) in ≥70-year patients.

When stratifying patients by haematological or solid cancers, those with haematological cancers had a higher risk of COVID-19 outcomes (Figure 3, Table S4). These differences were more pronounced in <1-year cancer patients. For example, the overall aHR for having a direct COVID-19 hospitalisation was 2.51 (2.12-2.98) for patients with haematological cancers and 1.24 (1.15-1.33) for those with solid cancers. Among <1-year cancer patients, aHRs were 6.18 (4.31-8.86) for haematological cancers and 1.49 (1.19-1.87) for solid cancers. Patients with haematological cancers also had an increased risk of

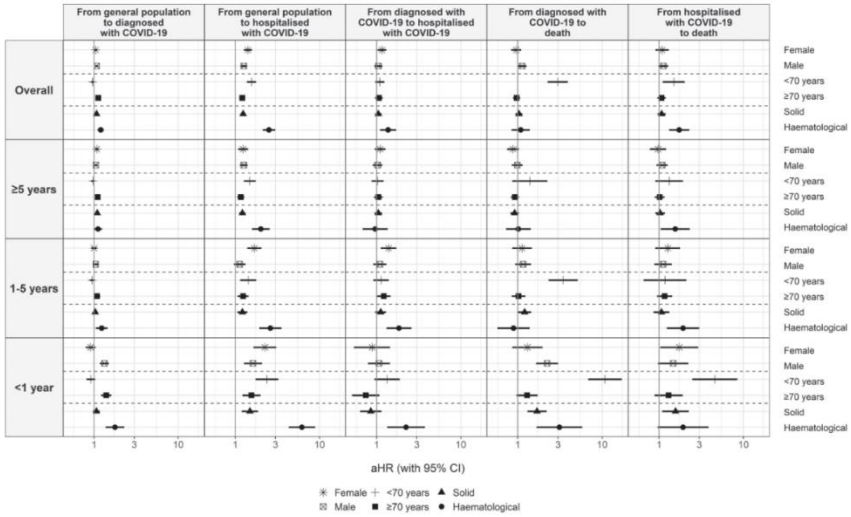



**FIGURE 3** Adjusted hazard ratios of COVID-19 outcomes in patients with cancer (overall and by years since cancer diagnosis) compared to patients without cancer, stratified by sex, age and cancer type (solid or haematological). Models are adjusted for sex (excepting models stratified by sex), age, the MEDEA deprivation index, smoking status and comorbidities (autoimmune conditions, chronic kidney disease, chronic obstructive pulmonary disease, dementia, heart disease, hyperlipidaemia, hypertension, type-2 diabetes and obesity). aHR, adjusted hazard ratio; CI, confidence interval

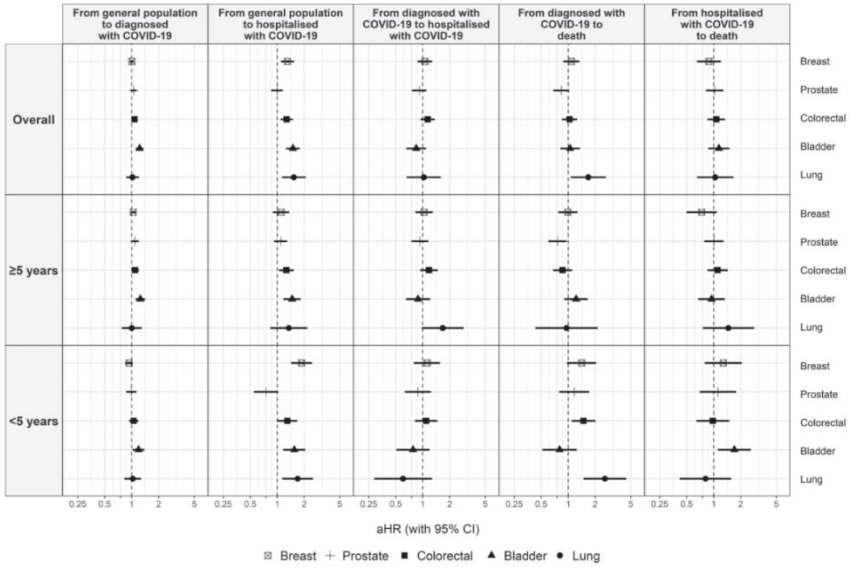



**FIGURE 4** Adjusted hazard ratios of COVID-19 outcomes in patients with cancer (overall and by years since the cancer diagnosis) compared to patients without cancer, stratified by solid cancer type. Models for specific cancer types include patients without cancer and patients with the cancer type of interest; models for prostate and breast cancer include only males and females, respectively. Models are adjusted for sex, age, smoking status, the MEDEA deprivation index, smoking status and comorbidities (autoimmune conditions, chronic kidney disease, chronic obstructive pulmonary disease, dementia, heart disease, hyperlipidaemia, hypertension, type-2 diabetes and obesity). aHR, adjusted hazard ratio; CI, confidence interval

COVID-19 hospitalisation following an outpatient diagnosis (overall 1.37 [1.10-1.71]; <1-year cancer patients: 2.24 [1.34-3.76]).

We also estimated the associations between cancer and COVID-19 outcomes by solid cancers. (Figure 4, Table S5). Due to small samples, models were estimated for breast, prostate, colorectal, bladder and lung cancer; overall and for <5-year (<1-year and 1-5-year categories combined) and ≥5-year cancer patients. Four cancer types were associated with having a direct COVID-19 hospitalisation: breast (1.30 [1.10-1.54]), colorectal (1.28 [1.10-1.49]), bladder (1.50 [1.26-1.79]) and lung (1.53 [1.13-2.08]) cancer; these associations were stronger in <5-year cancer patients. Lung cancer was associated with death following a COVID-19 diagnosis (1.68 [1.06-2.64]), with a stronger association in <5-year cancer patients (2.57 [1.49-4.46]). Bladder cancer was associated with death following a COVID-19 hospitalisation only in <5-year cancer patients (1.70 [1.11-2.60]).

### 3.4 | Sensitivity analysis

The assumption of proportionality was violated for age and years since cancer diagnosis for the risk of COVID-19 diagnosis (Figure S5). Thus, we stratified our model by years since cancer diagnosis and calendar time (Figure S6). The overall association was similar in March and April. However, in <1-year cancer patients, cancer was associated with a significant increased risk of COVID-19 diagnosis in April (1.41 [1.23-1.60]) but not in March (0.91 [0.80-1.05]).

In models restricted to never smokers (n = 1 834 657), the results were similar to those including all the population (Figure S7). Patients with missing data (n = 1 502 442) were younger and had fewer comorbidities than patients without missing data, but the distribution of cancer types was similar in both groups (Table S6). Despite these differences, imputed models showed similar results to the main models (Figure S8).

## 4 | DISCUSSION

In our population-based cohort study including 4 618 377 adults, a prior diagnosis of cancer was associated with an increased risk of COVID-19 outpatient (clinical) diagnosis, direct COVID-19 hospitalisation (without a prior outpatient diagnosis) and COVID-19-related death during the first wave of the COVID-19 pandemic in Catalonia, Spain. Overall, these associations were stronger in patients with a recent cancer diagnosis (<1 year), younger than 70 years and with haematological cancers. Lung and bladder cancers were also associated with higher risk of COVID-19 hospitalisation and death.

Prior studies investigating the risk of contracting SARS-CoV-2 in patients with cancer have reported conflicting results.[6,10,27,28] Even though we did not analyse the risk of COVID-19 infection per se, patients with cancer had a modestly increased risk of having an outpatient COVID-19 diagnosis, which was higher in <1-year cancer patients with haematological cancers. This is consistent with two studies from the United States (US) showing an increased risk of infection in patients with cancer, which was higher in those recently diagnosed and/or with haematological cancers.[6,27] Increased risk of diagnosis could be related to higher levels of interaction with healthcare services among patients with cancers, especially among those with a recent cancer diagnosis (thus, higher risk of being diagnosed with COVID-19 but also higher exposure to healthcare-associated infections), and to factors related to the cancer itself and/or cancer therapies (eg, haematological cancers, as well as treatment-related immunosuppression, thus increasing the risk of infection).[29]

Patients with cancer have also been reported to be at increased risk of COVID-19 severity, including hospitalisation and death.[6-9] We found that cancer was associated with a higher risk of direct hospitalisation, especially among <1-year cancer patients. Conversely, <1-year cancer patients had not an increased risk of subsequent hospitalisation (following an outpatient diagnosis). This counterintuitive finding could be explained by differences in care-seeking behaviours and/or in the clinical presentation of COVID-19. On the one hand, patients recently diagnosed with cancer have more interactions with hospital services and, therefore, could be more prone to seek care directly at the hospital level than the general population.[30] On the other hand, these patients might have a higher risk of rapidly developing severe COVID-19 symptoms due to their impaired immunity, thus more likely to be directly hospitalised. It is worth noting that although <1-year cancer patients had the highest risk of hospitalisation, this association remained significant in >5-year cancer patients (which mostly represent cancer survivors). This is consistent with a study showing that cancer survivors have higher risks of hospitalisation and death from influenza than cancer-free patients,[31] and could be related to long-term effects on the immune system of cancer therapies.

Conversely, the risk of COVID-19-related death was only significantly higher in <1-year cancer patients. Again, this could be due to factors related to the cancer itself (ie, the group of <1-year cancer patients might include individuals with more aggressive and active cancers) and/or cancer therapies. However, while some studies have shown that active cancer therapies increase the risk of COVID-19 death,[9] others have not.[8,32] These studies included different populations, cancer types or considered all different cancer therapies combined, which might have a different impact on COVID-19 outcomes. For instance, two meta-analyses reported an association between recent chemotherapy and increased COVID-19-related death, but a null association with recent surgery, radiotherapy, immunotherapy and targeted therapies.[33,34]

We found that the associations between cancer and direct hospitalisation and COVID-19-related deaths were more pronounced in patients younger than 70 years or with haematological cancers. Given that age is strongly associated with severe COVID-19 outcomes, cancer in older patients might not have a significantly worse impact as compared to cancer-free patients. In a study including 1187 patients with solid cancers and COVID-19, younger patients (<60 years) were also those with the highest risk of in-hospital mortality when compared to cancer-free patients.[35] Furthermore, increasing

evidence shows that patients with haematological cancers have a higher risk of poor COVID-19 outcomes.[6,7,9] The OpenSAFELY study reported an association between cancer and increased COVID-19 death, which was stronger in <1-year cancer patients and in those with haematological cancers.[7] Estimated aHRs for <1-year cancer patients were similar to ours for death following a COVID-19 diagnosis, with an aHR of 1.72 (1.50-1.96) (vs 1.69 [1.30-2.19] in our study) for solid cancer patients and an aHR of 2.80 (2.08-3.78) (vs 3.11 [1.67-5.81]) for haematological cancer patients. We also found a higher risk of hospitalisation and COVID-19-related death for lung and bladder cancers, both of which are strongly linked to tobacco smoking. While lung cancer has already been associated with poor COVID-19 outcomes,[36] to our knowledge, our study is the first showing an association with bladder cancer. However, these findings should be interpreted with caution considering the small sample sizes, which prevented us from performing analysis restricted to never smokers by specific cancer types.

Our study has several strengths. First, we used prospective data from a large and representative population covering almost all the population in Catalonia, and we included a heterogeneous cancer population. Second, by including patients with a clinical COVID-19 diagnosis, we avoided selection bias due to testing restrictions, or to (hypothetically) different testing patterns (ie, higher rates of testing in patients with cancer), although some cases might be false positives. Third, we performed our analysis across different cancer population groups, allowing us to identify those at highest risk of poor COVID-19 outcomes. Finally, our results were robust after restricting participants to never smokers and after multiple imputation of missing data, which lends credibility to our findings.

However, our study also has weaknesses. First, we did not have information on cancer stage nor specific-cancer therapy receipt and used instead years since cancer diagnosis as a proxy for active/inactive cancer. We also did not have information on the cause of death and considered as COVID-19-related deaths those occurring following a COVID-19 state. However, in patients with cancer, occurrence of death was substantially higher in those diagnosed (11.1%) and hospitalised (24.8%) with COVID-19 than in those without COVID-19 (1.3%), which suggests that we did capture deaths due to COVID-19. In addition, the proportion of deaths among hospitalised patients was in line with prior studies.[37] On the other hand, we cannot discard that some deaths in the general population might have occurred in undiagnosed COVID-19 cases, especially at the beginning of the pandemic. Second, due to the nature of our database, our results are not representative of asymptomatic or pauci-symptomatic COVID-19 cases that did not seek medical care. Third, our data spanned to May 2020, and therefore, our results are generalizable to the first wave of the pandemic. While changes over time might have changed SARS-CoV-2 virulence (eg, the emergence of new variants), it is unlikely that such changes have decreased the risk of severe disease among patients with cancer when compared to patients without cancer. Finally, routinely collected data often raise concerns about data quality, and some conditions, including cancer itself, may have been incompletely or inaccurately recorded. However, we used previously validated cancer codes,[19] and we included only individuals with at least 1 year of prior history available to comprehensively capture baseline characteristics.

Despite these weaknesses, our results highlight that patients with cancer are a vulnerable population for COVID-19 and, therefore, should be prioritised for vaccination against SARS-CoV-2. Unfortunately, the efficacy and effectiveness of COVID-19 vaccines in this subgroup population remain unknown. Indeed, patients with active cancer were excluded from randomised clinical trials,[38] and, to our knowledge, observational studies describing vaccine's effectiveness among patients with cancer are lacking to date. Emerging data suggest that these patients might have a weakened response to COVID-19 vaccines,[39,40] and recent studies have shown that COVID-19 vaccines are less effective among individuals immunocompromised.[41,42] As a result, the Centers for Disease Control and Prevention recently recommended a third mRNA-based vaccine dose among individuals immunocompromised, which include patients with ongoing treatment for haematological cancers or who have received a stem cell transplant within the last 2 years.[43] Further studies are needed to assess the effectiveness of COVID-19 vaccines among patients with cancer, overall and by oncologic features (eg, cancer type, cancer treatment), as well as to elucidate the utility of antibody testing[44] and booster vaccine doses. Meanwhile, these patients should also be protected with continued non-pharmaceutical interventions, infection control measures in healthcare settings and increased vaccination uptake among their caregivers and close contacts.

In conclusion, our population-based cohort study including a heterogeneous cancer population provides a comprehensive analysis of the associations between cancer and COVID-19 outcomes during the first wave of the pandemic in a Southern European region. Cancer was associated with an increased risk of COVID-19 diagnosis, hospitalisation and COVID-19-related death, with higher risks for patients diagnosed with cancer within the year prior, as well as those younger than 70 years and those with haematological cancers. Research is needed to address potential risk differences by specific cancer types, such as lung or bladder cancer, as well as to analyse the effect of subsequent COVID-19 waves. Notwithstanding that, our results highlight that patients with cancer are a vulnerable population for COVID-19. These patients, as well as their caregivers, should be prioritised in preventive strategies, including vaccination campaigns and continued non-pharmaceutical interventions.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the patients who suffered from or died of this devastating disease, and their families and carers. They would also like to thank the healthcare professionals involved in the management of COVID-19 during these challenging times, from primary care to intensive care units in the Catalan healthcare system. The analysis has made use of a range of open-source, free available tools provided by the OHDSI community. The phenotypes used in our study were developed or adapted from work performed during the OHDSI COVID-19 studyathon. The mapping of data to the OMOP CDM has been supported by a taskforce from EHDEN.

## CONFLICT OF INTEREST

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: DPA reports grant support from Les Laboratoires Servier; that his research group has received grants and advisory or speaker fees from Amgen, Astellas, AstraZeneca, Chesi-Taylor, Johnson and Johnson and UCD; and that Janssen, on behalf of Innovative Medicines Initiative-funded European Health Data Evidence Network and European Medical Information Framework consortium and Synapse Management Partners, have supported training programs, open to external participants, organised by his department. No other relationships or activities that could appear to have influenced the submitted work.

## DATA AVAILABILITY STATEMENT

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

## ETHICS STATEMENT

Our study was approved by the Clinical Research Ethics Committee of the IDIAPJGol (project code: 20/070-PCV). Informed consent of individual patients was not required as anonymised information was obtained from medical records.

## TRANSPARENCY STATEMENT

Elena Roel and Talita Duarte-Salles as guarantors of the study affirm that the study is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted and that any discrepancies from the study as planned have been explained.

## ORCID

*Elena Roel* https://orcid.org/0000-0002-1964-3546
*Martina Recalde* https://orcid.org/0000-0003-2256-618X
*Sergio Fernández-Bertolín* https://orcid.org/0000-0002-0029-6564
*Isabelle Soerjomataram* https://orcid.org/0000-0002-6017-741X
*Diana Puente* https://orcid.org/0000-0003-4725-537X
*Daniel Prieto-Alhambra* https://orcid.org/0000-0002-3950-6346
*Edward Burn* https://orcid.org/0000-0002-9286-1128
*Talita Duarte-Salles* https://orcid.org/0000-0002-8274-0357

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. https://doi.org/10.3322/caac.21660
2. Kumar D, Humar A. Respiratory viral infections in transplant and oncology patients. *Infect Dis Clin North Am*. 2010;24:395-412.
3. Liang W, Guan W, Chen R, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *Lancet Oncol*. 2020;21(3):335-337.
4. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708-1720.
5. Zhang Y, Han H, Tian Y, et al. Impact of cancer on mortality and severity of corona virus disease 2019: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. 2020;99(44):e23005.
6. Wang Q, Berger NA, Xu R. Analyses of risk, racial disparity, and outcomes among US patients with cancer and COVID-19 infection. *JAMA Oncol*. 2020;7(2):220-227.
7. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584:430-436.
8. Kuderer NM, Choueiri TK, Shah DP, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet*. 2020;6736:1-13.
9. Lee LYW, Cazier JB, Starkey T, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol*. 2020;21:1309-1316.
10. Angelis V, Tippu Z, Joshi K, Reis S, et al. Defining the true impact of coronavirus disease 2019 in the at-risk population of patients with cancer. *Eur J Cancer*. 2020;136:99-106.
11. Reilev M, Kristensen KB, Pottegård A, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. *Int J Epidemiol*. 2020;49:1468-1481.
12. Brar G, Pinheiro LC, Shusterman M, et al. COVID-19 severity and outcomes in patients with cancer: a matched cohort study. *J Clin Oncol*. 2020;38:3914-3924.
13. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11:1-12.
14. Prieto-Alhambra D, Balló E, Coma E, et al. Filling the gaps in the characterization of the clinical management of COVID-19: 30-day hospital admission and fatality rates in a cohort of 118 150 cases diagnosed in outpatient settings in Spain. *Int J Epidemiol*. 2020;49:1930-1939.
15. 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;138:617-621.
16. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-578.
17. 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;12:1-12.
18. Putter H, Fiocco M, Gekus RB. Tutorial in biostatistics: competing risk and multi-state models. *Stat Med*. 2007;26:2389-2430.
19. Recalde M, Manzano-Salgado CB, 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;11:1015-1024.
20. 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;22:179-187.
21. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA*. 2020;323:2052-2059.
22. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
23. Harrell FE. *Regression Modeling Strategies*. New York, NY: Springer; 2015.
24. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput*. 2009;38:1228-1234.

25. de Wreede LC, Fiocco M, Putter H. mstate: an R package for the analysis of competing risks and multi-state models. *J Stat Softw*. 2011;38:1-30.
26. Harrell Jr FE. rms: Regression modeling strategies. R package version 5.1-0. [Internet]. 2017 [cited April 6, 2021]; Available from: https://CRAN.R-project.org/package=rms
27. Wang QQ, Berger NA, Xu R. When hematologic malignancies meet COVID-19 in the United States: infections, death and disparities. *Blood Rev*. 2021;47:100775.
28. Bertuzzi AF, Marrari A, Gennaro N, et al. Low incidence of SARS-CoV-2 in patients with solid tumours on active treatment: an observational study at a tertiary cancer centre in Lombardy, Italy. *Cancers (Basel)*. 2020;12(9):2352.
29. Seth G, Sethi S, Bhattarai S, Saini G, Singh CB, Aneja R. SARS-CoV-2 infection in cancer patients: effects on disease outcomes and patient prognosis. *Cancers (Basel)*. 2020;12:1-16.
30. Lash RS, Bell JF, Reed SC, et al. A systematic review of emergency department use among cancer patients. *Cancer Nurs*. 2017;40:135-144.
31. Carreira H, Strongman H, Peppa M, et al. Prevalence of COVID-19-related risk factors and risk of severe influenza outcomes in cancer survivors: a matched cohort study using linked English electronic health records data. *EClinicalMedicine*. 2020;29–30:100656.
32. Fillmore NR, La J, Szalat RE, et al. Prevalence and outcome of COVID-19 infection in cancer patients: a National Veterans Affairs Study. *J Natl Cancer Inst*. 2021;113(6):691-698.
33. Park R, Lee SA, Kim SY, de Melo AC, Kasi A. Association of active oncologic treatment and risk of death in cancer patients with COVID-19: a systematic review and meta-analysis of patient data. *Acta Oncol (Madr)*. 2021;60:13-19.
34. Yekedüz E, Utkan G, Ürün Y. A systematic review and meta-analysis: the effect of active cancer treatment on severity of COVID-19. *Eur J Cancer*. 2020;141:92-104.
35. de Azambuja E, Brandão M, Wildiers H, et al. Impact of solid cancer on in-hospital mortality overall and among different subgroups of patients with COVID-19: a nationwide, population-based analysis. *ESMO Open*. 2020;5:e000947.
36. Garassino MC, Whisenant JG, Huang LC, et al. COVID-19 in patients with thoracic malignancies (TERAVOLT): first results of an international, registry-based, cohort study. *Lancet Oncol*. 2020;21:914-922.
37. Desai A, Gupta R, Advani S, et al. Mortality in hospitalized patients with cancer and coronavirus disease 2019: a systematic review and meta-analysis of cohort studies. *Cancer*. 2020;127(9):1459-1468.
38. Corti C, Curigliano G. Commentary: SARS-CoV-2 vaccines and cancer patients. *Ann Oncol*. 2021;32:569-571.
39. Monin L, Laing AG, Muñoz-Ruiz M, et al. Safety and immunogenicity of one versus two doses of the COVID-19 vaccine BNT162b2 for patients with cancer: interim analysis of a prospective observational study. *Lancet Oncol*. 2021;22(6):765-778.
40. Goshen-Lago T, Waldhorn I, Holland R, et al. Serologic status and toxic effects of the SARS-CoV-2 BNT162b2 vaccine in patients undergoing treatment for cancer. *JAMA Oncol*. 2021;1-8.
41. Chodick G, Tene L, Rotem RS, et al. The effectiveness of the two-dose BNT162b2 vaccine: analysis of real-world data. *Clin Infect Dis*. 2021;ciab438.
42. Tenforde MW, Patel MM, Ginde AA, et al. Effectiveness of severe acute respiratory syndrome coronavirus 2 messenger RNA vaccines for preventing coronavirus disease 2019 Hospitalizations in the United States. *Clin Infect Dis*. 2021;ciab687.
43. Centers for Disease Control and Prevention. COVID-19 Vaccines for Moderately to Severely Immunocompromised People | CDC [Internet]. [cited August 23, 2021]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations/immuno.html
44. Sun L, Warner JL, Parikh RB. Immune responses to SARS-CoV-2 among patients with cancer: what can seropositivity tell us? *JAMA Oncol*. 2021;7:1123-1125.


## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** 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. *Int. J. Cancer*. 2021;1-13. doi:10.1002/ijc.33846

## 5.4. Study IV: Socioeconomic Inequalities in COVID-19 Vaccination and Infection in Adults, Catalonia, Spain

Roel E, Raventós B, Burn E, Pistillo A, Prieto-Alhambra D, Duarte-Salles T.

*Emerging Infectious Diseases*. 2022;28(11):2243-2252.
Doi:10.3201/eid2811.220614

# Socioeconomic Inequalities in COVID-19 Vaccination and Infection in Adults, Catalonia, Spain

Elena Roel, Berta Raventós, Edward Burn, Andrea Pistillo, Daniel Prieto-Alhambra, Talita Duarte-Salles

Evidence on the impact of the COVID-19 vaccine rollout on socioeconomic COVID-19–related inequalities is scarce. We analyzed associations between socioeconomic deprivation index (SDI) and COVID-19 vaccination, infection, and hospitalization before and after vaccine rollout in Catalonia, Spain. We conducted a population-based cohort study during September 2020–June 2021 that comprised 2,297,146 adults ≥40 years of age. We estimated odds ratio of nonvaccination and hazard ratios (HRs) of infection and hospitalization by SDI quintile relative to the least deprived quintile, Q1. Six months after rollout, vaccination coverage differed by SDI quintile in working-age (40–64 years) persons: 81% for Q1, 71% for Q5. Before rollout, we found a pattern of increased HR of infection and hospitalization with deprivation among working-age and retirement-age (≥65 years) persons. After rollout, infection inequalities decreased in both age groups, whereas hospitalization inequalities decreased among retirement-age persons. Our findings suggest that mass vaccination reduced socioeconomic COVID-19–related inequalities.

The COVID-19 pandemic has caused an unprecedented global health crisis, resulting in >540 million cases worldwide as of July 2022 (*1*). However, the impact of the pandemic has not been uniform across or within countries (*2*). Disadvantaged populations, such as individuals with low socioeconomic status, display higher incidence rates of COVID-19 infection and hospitalization (*3*,*4*). To date, vaccines against SARS-CoV-2, the virus that causes COVID-19, are the cornerstone of the COVID-19 response. Yet, emerging evidence shows socioeconomic inequalities in COVID-19 vaccination coverage within countries with high access to vaccines, such as the United Kingdom or the United States (*5*–*8*). For instance, a report from May 2021 from the United Kingdom showed that vaccination coverage was 94% in the least areas and 84% in the most deprived areas (deprivation was measured using an index based on income, employment, education, health, crime, barriers to housing and services, and living environment) (*8*,*9*). Similarly, in the United States, vaccination coverage was lower (49%) among adults living in counties with the highest overall social vulnerability index (SVI) scores (based on socioeconomic status, household composition and disability, racial/ethnic minority status and language, and housing type and transportation) when compared to the coverage (59%) among adults living in counties with the lowest overall SVI scores in May 2021 (*10*). However, evidence is scarce regarding socioeconomic inequalities in COVID-19 vaccine uptake from other countries and the effect of the COVID-19 vaccine rollout on socioeconomic COVID-19–related outcomes inequalities.

In Spain, the COVID-19 vaccine rollout started on December 27, 2020. The first population groups eligible for vaccination were persons living in nursing homes and healthcare workers (*11*). Subsequently, other groups became eligible, taking into account age, starting with the eldest; underlying conditions, prioritizing persons with risk factors for COVID-19; and occupation, prioritizing essential workers. In Catalonia, a region located in northeast Spain, 52% of the population had received ≥1 dose of a COVID-19 vaccine as of June 30, 2021 (*12*). Determining patterns of socioeconomic inequalities in relation to COVID-19 vaccination and COVID-19 outcomes in Catalonia could provide valuable information to public health

Author affiliations: Universitat Autònoma de Barcelona, Barcelona, Spain (E. Roel, B. Raventós); Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Barcelona, Spain (E. Roel, B. Raventós, E. Burn, A. Pistillo, T. Duarte-Salles); Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxfordshire, United Kingdom (E. Burn, D. Prieto-Alhambra); Erasmus University Medical Center Department of Medical Informatics, Rotterdam, Netherlands (D. Prieto-Alhambra)

DOI: https://doi.org/10.3201/eid2811.220614

authorities to guide immunization efforts among vulnerable populations in Spain and in other countries with widespread access to vaccines.

We analyzed the association between a socioeconomic deprivation index (SDI) score based on place of residence (a proxy measure of socioeconomic status) and COVID-19 vaccination coverage 6 months after the start of vaccine rollout among adults ≥40 years of age living in urban areas of Catalonia. Subsequently, we analyzed the associations between SDI score and COVID-19 infection, hospitalization, and death, before and after the start of vaccine rollout. The Clinical Research Ethics committee of Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol) approved this study (project code 21/052-PCV), with no required written consent from participants.

## Methods

### Study Design and Data Source

We conducted a population-based cohort study during September 1, 2020–June 30, 2021, using primary care data from the Information System for Research in Primary Care (SIDIAP; https://www.sidiap.org) database, standardized to the Observational Medical Outcomes Partnership Common Data Model (*13*,*14*). SIDIAP contains pseudoanonymized electronic health records from ≈75% of the population in Catalonia, which has ≈7.5 million inhabitants, and is representative in terms of age, sex, and geographic distribution (*15*). SIDIAP includes data on sociodemographics, diagnoses, laboratory tests, medication use, and deaths. In addition, SIDIAP has been linked to the Catalan public health vaccine registry and to a population-based register of hospital discharge records from public and private hospitals of Catalonia (Conjunt Mínim Bàsic de Dades d'Alta Hospitalària, CMBD-AH) (E. Burn, et al., unpub. data, https://doi.org/10.1101/2021.11.23.21266734).

### Study Participants

We included 2,297,146 adults 40–110 years of age registered in SIDIAP as of September 1, 2020, after excluding those with <1 year of medical history available (n = 23,705), those with a previous COVID-19 infection (n = 125,111), those living in nursing homes (n = 31,091) and in rural areas (n = 513,386), and those with missing data on SDI (n = 307,038) (Figure 1). We included adults ≥40 years of age because those younger were not generally eligible for vaccination before mid-June 2021. We excluded persons living in rural areas, which included municipalities with <10,000 inhabitants and a population density <150 habitants/km$^2$ (*16*), because information on SDI was unavailable for these areas. We identified persons with a previous COVID-19 infection using SARS-CoV-2 positive tests or clinical COVID-19 diagnoses because SARS-CoV-2 tests were restricted to severe cases during the first months of the pandemic in Spain (*17*). We used Systematized Nomenclature of Medicine codes to identify COVID-19 diagnoses (Appendix Table 1, https://wwwnc.cdc.gov/EID/article/28/11/22-0614-App1.pdf).

To assess inequalities in COVID-19 vaccination coverage 6 months after the start of vaccine rollout (i.e., June 30, 2021), we restricted our analyses to persons with complete follow-up (vaccine coverage dataset, n = 2,258,866). We analyzed inequalities on COVID-19 outcomes for 2 time periods: 3 months before and 1–6 months after the start of vaccine rollout. For each period, we followed participants until the occurrence of the outcome of interest, end of study period, exit from database, or death, whichever occurred first. The period 3 months before vaccine rollout was September 1–December 26, 2020. The period 1–6 months after vaccine rollout was January 27–June 30, 2021; we excluded patients with a COVID-19 infection or lost before January 27, 2021 (n = 106,945), from analysis.

### Outcomes

We identified persons vaccinated against COVID-19 as those who had received a dose of any COVID-19 vaccine: BNT162b2 mRNA (Pfizer-BioNTech, https://www.pfizer.com), mRNA-1273 (Moderna, https://www.modernatx.com), ChAdOx1 nCoV-19 (Oxford-AstraZeneca, https://www.astrazeneca.com), or Ad.26.COV2.S (Janssen/J&J, https://www.janssen.com). The date of vaccination was the date of the first dose administration. We identified COVID-19 infections based on a positive SARS-CoV-2 antigen or reverse transcription PCR test, using the test date as the date of infection; we considered the first infection per person. We defined COVID-19 hospitalizations as hospitalizations with a positive SARS-CoV-2 test result between 21 days before and 3 days after the date of admission. We defined COVID-19–related deaths as deaths occurring ≤28 days after the date of infection.

### Variables

We measured SDI score using the Mortalidad en áreas pequeñas españolas y desigualdades socioeconómicas y ambientales (MEDEA) deprivation index (*16*). The MEDEA index was calculated for census tract urban

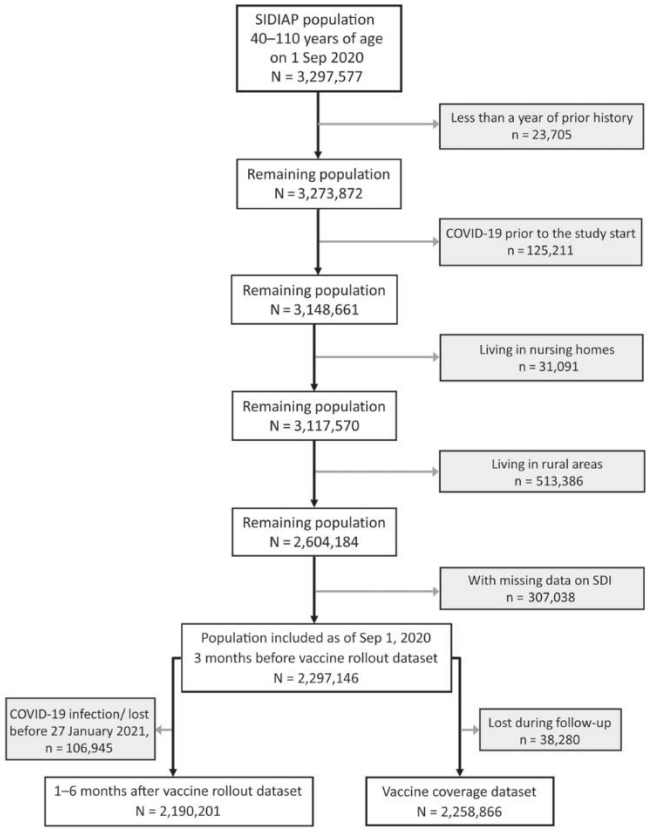



**Figure 1.** Flowchart showing the inclusion and exclusion criteria for study population in analysis of socioeconomic inequalities in COVID-19 vaccination and infection in adults, Catalonia, Spain. SDI, Socioeconomic Deprivation Index; SIDIAP, Information System for Research in Primary Care.

areas using information related to 5 indicators (related to work and education) from the 2001 national census in Spain. We linked the MEDEA deprivation index to each participant's most recent site of residence and categorized it into quintiles of socioeconomic deprivation, with the first quintile (Q1) representing the least deprived and the fifth (Q5) the most deprived area. We extracted age in years, sex, nationality by the country's geographic region, and comorbidities recorded before study start that were identified using Systematized Nomenclature of Medicine codes (Appendix Table 1). We categorized age into 2 groups: ≥65 (retirement age) and 40–64 years (working age).

### Statistical Analysis

We described participants' characteristics at baseline and by vaccination status, COVID-19 infection, hospitalization, and death over study follow-up period; we used counts and percentages for categorical variables and median and interquartile ranges (IQRs) for continuous variables. In accordance with information-governance requirements intended to protect confidentiality, we reported results with <5 persons as <5 rather than specific numbers. We also compared baseline characteristics of persons with and without missing data on SDI, and those with and without complete follow-up, using standardized mean differences (SMD). We

considered an absolute SMD ≥0.1 to be a meaningful difference in the distribution of a given characteristic between the groups compared (*18*). We generated charts of weekly cumulative vaccination coverages and incidence rates (IRs; cases/100,000 person-years) of COVID-19 infection, hospitalization, and death during September 1, 2020–June 30, 2021, by SDI quintile and age group. We used R version 4.1 (The R Project for Statistical Computing, https://www.r-project.org) for data curation, analysis, and visualization.

To assess the association between SDI quintile and nonvaccination, we performed crude and adjusted logistic regression models and calculated odds ratios (ORs) with 95% CIs by age group. We included persons with complete follow-up for these analyses (vaccine coverage dataset). To assess the association between SDI quintile and COVID-19 infection, hospitalization and death, we performed crude and adjusted Cox proportional-hazards models and calculated hazard ratios (HRs) with 95% CIs by age group and period using the 3 months before and 1–6 months after vaccine rollout datasets. We visually inspected log-log survival curves to check the proportional hazard assumptions for the variables included in the models. We did not estimate models in which the number of events per SDI quintile was <5. Models were relative to the least deprived quintile (Q1) and adjusted by age, sex, and nationality; we developed a directed acyclic graph to guide our modeling strategy (Appendix Figure 1) (*19*). Of note, rates of hospitalization and death were estimated among the total population rather than among those infected with COVID-19 to prevent collider bias (*20*).

In addition, we performed 3 sensitivity analyses. First, we reestimated our models for vaccination coverage after excluding persons with a COVID-19 infection during follow-up, because they were not eligible for vaccination until 6 months after the infection. Second, we reestimated our models for COVID-19 outcomes restricting our analyses to citizens of Spain because the proportionality assumption was violated for nationality and all the COVID-19 outcomes. Third, we estimated socioeconomic inequalities on COVID-19 outcomes for the time period 3–6 months after the start of vaccine rollout, March 27– June 30, 2021, after excluding those with a COVID-19 infection, deceased, or lost before March 27, 2021 (n = 137,663).

## Results

Among the 2,297,146 participants included, most (n = 1,518,851; 66.1%) were 40–64 years of age (median 57 years of age), were citizens of Spain (88.8%), and had few comorbidities (Table). Persons living in more deprived areas were younger, less frequently citizens of Spain, and had more comorbidities than those living in the least deprived ones (Appendix Table 2). Persons excluded because of missing data on SDI were slightly younger (median age 55 years), more frequently from Europe and North America, and less frequently from Asia and Oceania than those without missing data on SDI (Appendix Table 3). Compared with those in the vaccine coverage dataset (i.e., with complete follow-up), persons with incomplete follow-up (lost to follow-up) (n = 38,280; 1.7%) were older (median age 69 years), were less frequently citizens of Spain (80.3%), and had more comorbidities (Appendix Table 4). For 51.5% of that population, death was the reason patients were lost to follow-up.

### Vaccination Coverage and COVID-19 Infections, Hospitalizations, and Deaths at Study End

Six months after vaccine rollout, among those with complete follow-up (n = 2,258,866), 82.0% had been vaccinated. Vaccination coverage was highest among older persons (≥80 years; 92.6%), women (83.5%), those living in the least deprived areas (84.6% for Q1 vs. 76.7% for Q5), and those with comorbidities (e.g., 92.7% among persons with dementia) (Table). Vaccination coverage was particularly low among persons of other nationality: ≈60% for those from western Europe and America and <50% for those from Africa, Asia, and Oceania and from eastern Europe.

During September 1, 2020–June 30, 2021, a total of 134,966 (5.9%) persons were infected with COVID-19; of those, 16,921 (0.7%) were hospitalized for COVID-19, and 1,881 (0.1%) died (Table). Cases of COVID-19 were highest among younger persons, 40–49 years of age (6.8%), followed by those ≥80 years of age (4.9%); COVID-19 was also more common among migrants from Central and South America (9.1%) and Africa (7.5%) than for citizens of Spain (5.8%) and in the most deprived areas (6.8% for Q5) than the least deprived (5.3% for Q1). Conversely, hospitalizations were highest among the eldest (≥80 years; 1.5%), men (0.9%), those from Central and South America (1.1%), those with comorbidities (e.g., 1.8% among those with renal impairment), and those from the most deprived areas (0.9% for Q5 vs. 0.6% for Q1). Death rates were overall similar by sex, nationality, and SDI quintile but were higher among the eldest (0.6%) and those with comorbidities.

#### Trends in Vaccination Coverage and COVID-19 Infection, Hospitalization, and Death over Time

Among participants ≥65 years of age, vaccination coverage over time was similar across all SDI quintiles, whereas in those 40–64 years of age we observed a pattern of lower vaccination coverage in areas with increased socioeconomic deprivation (Figure 2). Regarding COVID-19 outcomes, IR of infection peaked in mid-October 2020 and mid-January 2021 and plateaued after March 2021. We observed a similar pattern for COVID-19 hospitalizations and deaths. Infection rates were higher among those 40–64 years of age, whereas hospitalization and death rates were higher among those ≥65 years of age. Overall, we observed a pattern of higher IR of infection and hospitalization in areas with increased socioeconomic deprivation among both age groups for the IR peaks. As for COVID-19 deaths, we found those living in the most deprived areas had the the higher IR for those peaks, without a clear pattern of increased IR with increased socioeconomic deprivation. After March 2021, differences by SDI quintile for all COVID-19 outcomes were less obvious, because IR of infection, hospitalization, and death were much lower.

#### Associations between SDI Quintile and Nonvaccination

Compared with persons ≥65 years of age living in the least deprived areas (Q1), those living in Q2, Q3, and Q4 areas had a lower probability of nonvaccination. In Q2 areas, OR was 0.97 (95% CI 0.95–1.00); in Q3 areas, 0.93 (95% CI 0.90–0.95); in Q4 areas, 0.90 (95% CI 0.88–0.93); and in Q5 areas, 1.01 (95% CI 0.99–1.04) (Figure 3; Appendix Figure 2). Conversely, among those 40–64 years of age, we found increased odds of nonvaccination for persons living in more deprived areas. For instance, when compared with those living

**Table.** Population characteristics in study of socioeconomic inequalities in COVID-19 vaccination and infection, Catalonia, Spain, 2020–2021*

| Characteristic | Population | Vaccinated† | Infected with COVID-19 | Hospitalized with COVID-19 | COVID-19–related death |
|---|---|---|---|---|---|
| Total | 2,297,146 (100.0) | 1,852,361 (82.0) | 134,966 (5.9) | 16,921 (0.7) | 1,881 (0.1) |
| Loss to follow-up | 38,280 (1.7) | 0 | 3,580 (9.4) | 1,779 (4.6) | 1,881 (4.9) |
| Median age, y (IQR) | 57 (48–69) | 59 (49–71) | 54 (47–66) | 66 (55–77) | 84 (76–89) |
| Age category, y | | | | | |
| 40–49 | 694,924 (30.3) | 481,716 (70.2) | 47,121 (6.8) | 2,330 (0.3) | 10 (0.0) |
| 50–59 | 582,558 (25.4) | 473,371 (82.1) | 38,092 (6.5) | 3,530 (0.6) | 64 (0.0) |
| 60–69 | 450,173 (19.6) | 385,155 (86.6) | 23,525 (5.2) | 3,815 (0.8) | 162 (0.0) |
| 70–79 | 345,152 (15.0) | 316,244 (93.2) | 15,221 (4.4) | 3,793 (1.1) | 402 (0.1) |
| ≥80 | 224,339 (9.8) | 195,875 (92.6) | 11,007 (4.9) | 3,453 (1.5) | 1,243 (0.6) |
| Sex | | | | | |
| F | 1,200,296 (52.3) | 987,415 (83.5) | 71,185 (5.9) | 7,262 (0.6) | 802 (0.1) |
| M | 1,096,850 (47.7) | 864,946 (80.4) | 63,781 (5.8) | 9,659 (0.9) | 1,079 (0.1) |
| Nationality | | | | | |
| Spain | 2,040,130 (88.8) | 1,726,192 (85.9) | 117,423 (5.8) | 14,864 (0.7) | 1,833 (0.1) |
| Africa | 69,086 (3.0) | 30,053 (44.8) | 5,161 (7.5) | 580 (0.8) | 13 (0.0) |
| Central & South America | 70,312 (3.1) | 40,287 (59.2) | 6,368 (9.1) | 740 (1.1) | 10 (0.0) |
| Asia & Oceania | 47,906 (2.1) | 21,888 (46.9) | 3,063 (6.4) | 443 (0.9) | 10 (0.0) |
| Eastern Europe | 34,803 (1.5) | 13,015 (38.4) | 1,674 (4.8) | 177 (0.5) | <5 |
| Western Europe & North America | 34,909 (1.5) | 20,926 (61.9) | 1,277 (3.7) | 117 (0.3) | 12 (0.0) |
| SDI quintile | | | | | |
| Q1 | 478,380 (20.8) | 397,672 (84.6) | 25,441 (5.3) | 2,748 (0.6) | 345 (0.1) |
| Q2 | 469,833 (20.5) | 387,994 (83.8) | 26,302 (5.6) | 3,123 (0.7) | 370 (0.1) |
| Q3 | 465,245 (20.3) | 378,990 (82.7) | 26,955 (5.8) | 3,393 (0.7) | 389 (0.1) |
| Q4 | 453,924 (19.8) | 364,672 (81.7) | 27,154 (6.0) | 3,604 (0.8) | 382 (0.1) |
| Q5 | 429,764 (18.7) | 323,033 (76.7) | 29,114 (6.8) | 4,053 (0.9) | 395 (0.1) |
| Comorbidities | | | | | |
| Asthma | 141,725 (6.2) | 118,256 (84.7) | 9,344 (6.6) | 1,373 (1.0) | 134 (0.1) |
| Autoimmune disease | 58,146 (2.5) | 50,527 (88.6) | 3,470 (6.0) | 607 (1.0) | 101 (0.2) |
| COPD | 119,845 (5.2) | 105,236 (91.4) | 6,497 (5.4) | 1,999 (1.7) | 403 (0.3) |
| Dementia | 30,223 (1.3) | 24,747 (92.7) | 2,082 (6.9) | 597 (2.0) | 314 (1.0) |
| Heart disease | 402,389 (17.5) | 353,597 (90.8) | 22,829 (5.7) | 5,696 (1.4) | 1,172 (0.3) |
| Hypertension | 775,420 (33.8) | 681,602 (90.0) | 43,425 (5.6) | 9,319 (1.2) | 1,498 (0.2) |
| Obesity | 515,509 (22.4) | 435,893 (85.9) | 34,914 (6.8) | 6,619 (1.3) | 626 (0.1) |
| Malignant neoplastic disease | 264,658 (11.5) | 233,327 (91.4) | 14,285 (5.4) | 3,165 (1.2) | 677 (0.3) |
| Renal impairment | 169,947 (7.4) | 149,516 (92.5) | 9,690 (5.7) | 3,131 (1.8) | 840 (0.5) |
| Type 2 diabetes | 288,188 (12.5) | 251,117 (89.7) | 17,689 (6.1) | 4,637 (1.6) | 758 (0.3) |

*Values are no. (%) except as indicated. Study population as of September 1, 2020, We noted characteristics overall and by vaccination, COVID-19 infection, hospitalization, and death status over follow-up period. Quintiles listed from least deprived (Q1) to most deprived (Q5). COPD, chronic obstructive pulmonary disease; IQR, interquartile range; SDI, Socioeconomic Deprivation Index.
†Among those with complete follow-up, n = 2,258,866.

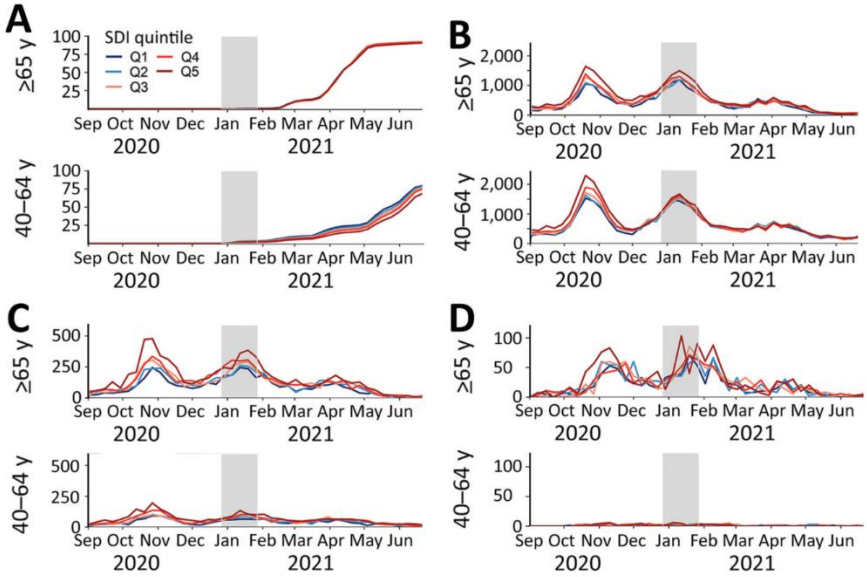



**Figure 2.** Vaccination coverage and incidence rates of COVID-19 infection, hospitalization, and death over time by SDI quintile and age group in study of socioeconomic inequalities in COVID-19 vaccination and infection, Catalonia, Spain, 2020–2021. Only persons with complete follow-up were included to estimate vaccination coverages. Gray area shows the first month after the start of vaccine rollout (December 27, 2020). Q1 represents the least deprived quintile, Q5 the most deprived. A) Vaccination coverage by age group, shown as percentage of population. B) COVID-19 infections by age group, shown as incidence rate per 100,000 person-years. C) COVID-19 hospitalizations, shown as incidence rate per 100,000 person-years. COVID-19–related deaths, shown as incidence rate per 100,000 person-years. Q, quintile; SDI, Socioeconomic Deprivation Index.

in Q1 areas, OR of nonvaccination was 1.01 (95% CI 1.00–1.02) in Q2 areas, 1.08 (95% CI 1.07–1.10) in Q3 areas, 1.11 (95% CI 1.10–1.13) in Q4 areas, and 1.33 (95% CI 1.31–1.35) in Q5 areas. Sensitivity analyses excluding persons with a COVID-19 infection before vaccination (n = 124,522) were consistent with our main analyses (Appendix Figure 3).

#### Association between SDI Quintile and COVID-19 Outcomes

Three months before vaccine rollout, we observed a pattern of increased HR of COVID-19 infection in more deprived areas in both age groups (Figure 4; Appendix Table 5). For example, among those ≥65 years of age, HR was 1.12 (95% CI 1.07–1.18) for those living in Q2 areas, 1.19 (95% CI 1.13–1.25) in Q3 areas, 1.26 (95% CI 1.20–1.32) in Q4 areas, and 1.54 (95% CI 1.46–1.61) in Q5 areas. A similar pattern was seen for COVID-19 hospitalizations among both age groups, with larger inequalities. Among persons ≥65 years of age, HR was 1.25 (95% CI 1.12–1.39) for those living in Q2 areas, 1.37 (95% CI 1.23–1.52) in Q3 areas, 1.53 (95% CI 1.38–1.70) in Q4 areas, and 1.99 (95% CI 1.80–2.19) in Q5 areas. Conversely, this pattern was not apparent for COVID-19–related deaths among persons ≥65 years of age; rates were only higher for those living in Q5 areas (HR 1.71 [95% CI 1.36–2.17]). We did not estimate models for death among persons 40–64 years of age because we observed <5 events in some SDI quintiles.

In the period 1–6 months after vaccine rollout, inequalities decreased in both age groups compared with the period before vaccine rollout (Figure 4: Appendix Table 5). Inequalities were still noticeable among those ≥65 years of age; HR was 1.08 (95% CI 1.02–1.14] for those living in Q2 areas, 1.09 (95% CI 1.03–1.15) in Q3 areas, 1.10 (95% CI 1.03–1.16) in Q4 areas, and 1.23 (95% CI 1.16–1.31) in Q5 areas. Conversely, among those 40–64 years of age, only those living in the most deprived

areas had higher rates of infection (Q5 HR 1.04 [95% CI 1.00–1.08]). Regarding hospitalizations, inequalities by SDI quintile remained in both age groups, although they decreased among those ≥65 years of age: HR was 1.17 (95% CI 1.04–1.32) for those living in Q2 areas, 1.27 (95% CI 1.14–1.43) in Q3 areas, 1.29 (95% CI 1.15–1.45) in Q4 areas, and 1.52 (95% CI 1.36–1.71) in Q5 areas. Similarly, rates of COVID-19-related deaths among those ≥65 years of age in Q5 areas moderately decreased; HR was 1.36 (95% CI 1.02–1.82).

In sensitivity analyses restricting participants to citizens of Spain, results were also consistent with our main analyses (Appendix Figure 4). In the period 3–6 months after vaccine rollout, results were overall similar to our main analysis, although among those ≥65 years of age, inequalities in hospitalizations were more apparent than 1–6 months after vaccine rollout. HR for hospitalizations 3–6 months after vaccine rollout were 1.33 (95% CI 1.10–1.60) for those living in Q2 areas, 1.47 (95% CI 1.23–1.77 in Q3 areas, 1.42 (95% CI 1.18–1.71) in Q4 areas, and 1.71 (95% CI 1.42–2.06) in Q5 areas (Appendix Table 6).

## Discussion

In this cohort study comprising >2 million adults living in urban areas of Catalonia, Spain, vaccination coverage was high (>80%) 6 months after the COVID-19 vaccine rollout. However, coverage differed by SDI quintile for place of residence; coverage was 85% in the least deprived areas and 77% in the most deprived areas. Among retirement-age persons (≥65 years), SDI quintile was not associated with vaccination, whereas among working-age persons (40–64 years), nonvaccination increased among those living in more deprived areas. Three months before vaccine rollout, we found a pattern of increased rates of COVID-19 infection and hospitalization among retirement-age and working-age persons living in more deprived areas. However, 6 months after rollout, socioeconomic inequalities in COVID-19 infection substantially decreased among both age groups, whereas inequalities in COVID-19 hospitalization moderately decreased only among retirement-age persons.

Surveys assessing inequalities in willingness to vaccinate (mostly conducted before vaccine rollout or shortly after) found conflicting results across countries (*21*–*23*). A study of 13,000 participants from 19 countries reported that younger age was associated with less willingness to vaccinate in the United Kingdom, Sweden, and Spain, whereas the opposite was observed in China (*22*). Conversely, higher education levels were associated with more willingness to vaccinate in the United States, France, and Germany, but not in Spain or the United Kingdom (*23*). Regarding COVID-19 vaccination coverage, studies are mostly limited to the United Kingdom (*7*,*24*,*25*) and the United States (*10*,*26*,*27*). However, these studies consistently found lower vaccination rates among persons with low socioeconomic status (*7*,*10*,*24*–*27*). This finding is also in line with prior evidence in relation to other vaccines (*28*,*29*). We found an association between higher socioeconomic deprivation and nonvaccination only among working-age persons. Differences

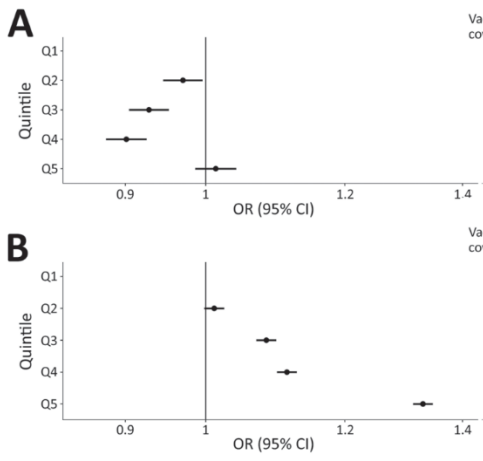



**Figure 3.** Odds ratios of nonvaccination 6 months after the start of COVID-19 vaccine rollout by Socioeconomic Deprivation Index quintile, stratified by age group, in study of socioeconomic inequalities in COVID-19 vaccination and infection, Catalonia, Spain, 2020–2021. A) OR for retirement-age persons ≥65 years of age. B) OR for working-age persons 40–64 years of age. Q1, the referent quintile, represents the least deprived areas; Q5, the most deprived. Persons with complete follow-up (n = 2,258,866) after vaccination were included. Models are adjusted for age, sex, and nationality. Dots indicate OR; bars, 95% CI. OR, odds ratio; Q, quintile.

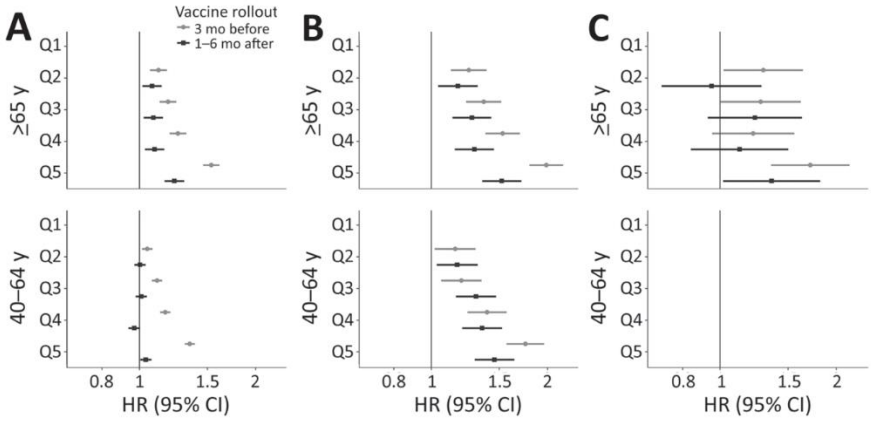



**Figure 4.** Fully adjusted hazard ratios of COVID-19 infection (A), hospitalization (B), and death (C) before and after vaccine rollout, by Socioeconomic Deprivation Index quintile and stratified by age group, in study of socioeconomic inequalities in COVID-19 vaccination and infection, Catalonia, Spain, 2020–2021. Q1, the referent quintile, represents the least deprived areas; Q5, the most deprived. Vaccine rollout started on December 27, 2020. Models before vaccine rollout are from September 1–December 26, 2020. Models after vaccine rollout are from January 27–June 30, 2021. All models are adjusted for age, sex, and nationality. Models in which the number of events for ≥1 deprivation area was <5 were not estimated. Dots indicate OR; bars, 95% CI. Q, quintile. HR, hazard ratio.

by age group could be related to working conditions (i.e., unavailability to miss work to vaccinate), as well as to an enhanced COVID-19 risk perception among older persons, who have a higher risk for severe disease (*22*,*30*). Unlike our study, UK studies also observed inequalities in coverage among the elderly (*7*,*25*). Differences in the development of the pandemic, the vaccination campaign, or cultural perspectives across countries might explain these discrepancies. Spain was severely hit by the first wave of the pandemic (*17*) and is one of the countries with the highest COVID-19 vaccination coverages (*31*). Furthermore, Spain is a country with traditionally high levels of vaccine confidence and with high vaccination coverages overall (*32*).

Inequalities among working-age persons are concerning, because those with low socioeconomic status are more likely to be exposed to infection because of poorer working and housing conditions and to develop severe disease because of poorer health status (*4*,*33*). Those findings are consistent with our findings before vaccine rollout, as well as with prior evidence from the United States and Europe, including Spain (*3*,*34*,*35*). In July–November 2020 the risk ratio of COVID-19 infection in residents of the poorest areas of Barcelona, the capital of Catalonia, was 1.67 (95% CI 1.41–196) in men and 1.71 (95% CI 1.44–1.99) in women, in line with our findings (*35*).

Despite inequalities in vaccination coverage, socioeconomic inequalities for COVID-19 infection decreased 6 months after vaccine rollout among both age groups, suggesting that vaccines reduced inequalities partly through mechanisms of herd immunity (*36*). Conversely, inequalities in hospitalizations decreased, although they still persisted, only among retirement-age persons. This finding highlights the importance of addressing vaccine inequalities among working-age persons. Persisting inequalities among the retirement-age persons might be related to differences in the risk for severe COVID-19 once infected because we found that those living in more deprived areas have more comorbidities and, thus, higher risk for complications (*33*). In addition to nationwide vaccination campaigns, strategies addressing structural inequalities are needed to reduce the burden of COVID-19–related outcomes among those most vulnerable (*6*).

The main strength of this study is the nature of our database, which encompasses ≈75% of the population of Catalonia. In addition, our data include a complete record of vaccines administered and of COVID-19 tests performed at public healthcare facilities. This study provides novel evidence regarding the associations between socioeconomic deprivation and COVID-19 infection, hospitalization, and death before and after the COVID-19 vaccine rollout in a country in southern Europe.

The first limitation of our study is that, although area-based indices of socioeconomic deprivation are widely used in epidemiologic studies, our results should be interpreted with caution considering the risks of ecologic bias. Second, we lacked information on occupation, which would have been of interest to have a better understanding of our results among working-age persons; a UK study reported lower vaccination coverage among persons working in manual occupations (*37*). Last, our results might not be generalizable to other contexts because of differences across countries, although they provide insights into the effects on socioeconomic COVID-19 inequalities of a mass vaccination campaign in a high-income country with high access to vaccination.

Despite socioeconomic inequalities in vaccination coverage, our results show that inequalities in COVID-19 infection and hospitalization in urban areas decreased but still persisted 6 months after the start of vaccine rollout in Catalonia. Our findings show that mass COVID-19 vaccination reduced COVID-19–related inequalities and emphasize the need to pursue efforts to vaccinate all population subgroups.

#### Acknowledgments

We thank all healthcare professionals in Catalonia who daily register information in the populations' electronic health records. We also want to thank 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.

This study was carried out as part of the doctoral program in methodology of biomedical research and public health at the Autonomous University of Barcelona. This work was supported by Instituto de Salud Carlos III (grant no. CM20/00174 to E.R. and CP21/00023 to T.D.S.) and by the National Institute for Health Research (grant no. SRF-2018-11-ST2-004 to D.P.A.). The funders of the study had no role in study design, data collection, analysis, and interpretation, or writing of the report.

D.P.A.'s research group has received grant support from Amgen, Chesi-Taylor, Novartis, and UCB Biopharma; D.P.A.'s department has received advisory or consultancy fees from Amgen, Astellas, AstraZeneca, Johnson and Johnson, and UCB Biopharma; as well as fees for speaker services from Amgen and UCB; Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programs organized by D.P.A.'s department and open for external participants. No other relationships or activities that could do so appear to have influenced the submitted work accuracy of the data analysis.

#### About the Author

Dr. Roel is a medical doctor specialized in preventive medicine and public health. She is currently a predoctoral researcher at Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain. Her thesis focuses on the use of real-world data to fill evidence gaps in the field of COVID-19.

#### References

1. World Health Organization. COVID-19 weekly epidemiological update. 2022 [cited 2022 Sep 22]. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-june-2022
2. Paremoer L, Nandi S, Serag H, Baum F. COVID-19 pandemic and the social determinants of health. BMJ. 2021;372:n129. https://doi.org/10.1136/bmj.n129
3. Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al.; OpenSAFELY Collaborative. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. Lancet. 2021;397:1711–24. https://doi.org/10.1016/S0140-6736(21)00634-6
4. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. Build back fairer: the COVID-19 Marmot review. London: Institute of Health Equity; 2020 [cited 2022 Sep 22]. https://www.health.org.uk/publications/build-back-fairer-the-covid-19-marmot-review
5. Hughes MM, Wang A, Grossman MK, Pun E, Whiteman A, Deng L, et al. County-level COVID-19 vaccination coverage and social vulnerability — United States, December 14, 2020–March 1, 2021. MMWR Morb Mortal Wkly Rep. 2021;70:431–6. https://doi.org/10.15585/mmwr.mm7012e1
6. Haque Z. Vaccine inequality may undermine the booster programme. BMJ. 2021;375:n3118. https://doi.org/10.1136/bmj.n3118
7. Curtis HJ, Inglesby P, Morton CE, MacKenna B, Green A, Hulme W, et al.; The OpenSAFELY Collaborative. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. Br J Gen Pract. 2021;72:e51–62. https://doi.org/10.3399/BJGP.2021.0376
8. Mutebi N. COVID-19 vaccine coverage and targeted interventions to improve vaccination uptake. 2021 [cited 2022 Feb 8]. https://post.parliament.uk/covid-19-vaccine-coverage-and-targeted-interventions-to-improve-vaccination-uptake
9. Ministry of Housing Communities and Local Government. National statistics: English indices of deprivation 2019. Vol. 2019, Ministry of Housing, Communities & Local Government. 2019 [cited 2022 Feb 8]. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
10. Barry V, Dasgupta S, Weller DL, Kriss JL, Cadwell BL, Rose C, et al. Patterns in COVID-19 vaccination coverage, by social vulnerability and urbanicity — United States, December 14, 2020–May 1, 2021. MMWR Morb Mortal Wkly Rep. 2021;70:818–24. https://doi.org/10.15585/mmwr.mm7022e1
11. Inter-territorial Council of the National Health System. COVID-19 vaccination strategy in Spain. Vol. 1. 2021 [cited 2022 Feb 8]. https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Actualizaciones_Estrategia_Vacunacion/docs/COVID-19_Actualizacion7_EstrategiaVacunacion.pdf

12. Institute of Statistics of Catalonia. COVID-19 vaccination. 2022 [cited 2022 Feb 8]. https://www.idescat.cat/indicadors/?lang=es&id=conj&n=14357
13. Recalde M, Rodrıguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data resource profile: the Information System for Research in Primary Care (SIDIAP). Int J Epidemiol. 2022 Apr 13 [Epub ahead of print]. https://doi.org/10.1093/ije/dyac068 https://doi.org/10.1093/ije/dyac068
14. Observational Health Data Sciences and Informatics. The book of OHDSI. 2019 [cited 2022 Feb 8]. https://ohdsi.github.io/TheBookOfOhdsi/TheBookOfOhdsi.pdf
15. Bolíbar B, Fina Avilés F, Morros R, Garcia-Gil MM, Hermosilla E, Ramos R, et al.; Grupo SIDIAP. SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research [in Spanish]. Med Clin (Barc). 2012;138:617–21.
16. Domínguez-Berjón MF, Borrell C, Cano-Serral G, Esnaola S, Nolasco A, Pasarín MI, et al. Constructing a deprivation index based on census data in large Spanish cities (the MEDEA project) [in Spanish]. Gac Sanit. 2008;22:179–87.
17. Henríquez J, Gonzalo-Almorox E, García-Goñi M, Paolucci F. The first months of the COVID-19 pandemic in Spain. Health Policy Technol. 2020;9:560–74. https://doi.org/10.1016/j.hlpt.2020.08.013
18. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput. 2009;38:1228–34. https://doi.org/10.1080/03610910902859574
19. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10:37–48. https://doi.org/10.1097/00001648-199901000-00008
20. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun. 2020;11:5749. https://doi.org/10.1038/s41467-020-19478-2
21. Sherman SM, Smith LE, Sim J, Amlôt R, Cutts M, Dasch H, et al. COVID-19 vaccination intention in the UK: results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. Hum Vaccin Immunother. 2021;17:1612–21. https://doi.org/10.1080/21645515.2020.1846397
22. Brailovskaia J, Schneider S, Margraf J. To vaccinate or not to vaccinate!? Predictors of willingness to receive COVID-19 vaccination in Europe, the U.S., and China. PLoS One. 2021;16:e0260230. https://doi.org/10.1371/journal.pone.0260230
23. Lazarus JV, Wyka K, Rauh L, Rabin K, Ratzan S, Gostin LO, et al. Hesitant or not? The association of age, gender, and education with potential acceptance of a COVID-19 vaccine: a country-level analysis. J Health Commun. 2020;25:799–807. https://doi.org/10.1080/10810730.2020.1868630
24. Perry M, Akbari A, Cottrell S, Gravenor MB, Roberts R, Lyons RA, et al. Inequalities in coverage of COVID-19 vaccination: a population register based cross-sectional study in Wales, UK. Vaccine. 2021;39:6256–61. https://doi.org/10.1016/j.vaccine.2021.09.019
25. Nafilyan V, Dolby T, Razieh C, Gaughan CH, Morgan J, Ayoubkhani D, et al. Sociodemographic inequality in COVID-19 vaccination coverage among elderly adults in England: a national linked data study. BMJ Open. 2021;11:e053402. https://doi.org/10.1136/bmjopen-2021-053402
26. Green-McKenzie J, Shofer FS, Momplaisir F, Kuter BJ, Kruse G, Bialal U, et al. Factors associated with COVID-19 vaccine receipt by health care personnel at a major academic hospital during the first months of vaccine availability. JAMA Netw Open. 2021;4:e2136582–2136582. https://doi.org/10.1001/jamanetworkopen.2021.36582
27. Cole MB, Raifman JR, Assoumou SA, Kim JH. Assessment of administration and receipt of COVID-19 vaccines by race and ethnicity in US federally qualified health centers. JAMA Netw Open. 2022;5:e2142698–e2142698. https://doi.org/10.1001/jamanetworkopen.2021.42698 PMID: 35006248 https://doi.org/10.1001/jamanetworkopen.2021.42698
28. Public Health England. National Immunisation Programme: health equity audit. 2021 [cited 2022 Feb 8]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957670/immnstn-equity_AUDIT_v11.pdf
29. Arat A, Burström B, Östberg V, Hjern A. Social inequities in vaccination coverage among infants and pre-school children in Europe and Australia—a systematic review. BMC Public Health. 2019;19:290. https://doi.org/10.1186/s12889-019-6597-4
30. Burn E, Tebé C, Fernandez-Bertolin S, Aragon M, Recalde M, Roel E, et al. The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nat Commun. 2021;12:777. https://doi.org/10.1038/s41467-021-21100-y
31. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of coronavirus (COVID-19) vaccinations. Nat Hum Behav. 2020;5:947–53 [cited 2022 Feb 8]. https://ourworldindata.org/covid-vaccinations
32. Antonini M, Eid MA, Falkenbach M, Rosenbluth ST, Prieto PA, Brammli-Greenberg S, et al. An analysis of the COVID-19 vaccination campaigns in France, Israel, Italy, and Spain and their impact on health and economic outcomes. Heal Policy Technol. 2022;11:100594. https://doi.org/10.1016/j.hlpt.2021.100594 https://doi.org/10.1016/j.hlpt.2021.100594
33. Campos-Matos I, Mandal S, Yates J, Ramsay M, Wilson J, Lim WS. Maximising benefit, reducing inequalities and ensuring deliverability: prioritisation of COVID-19 vaccination in the UK. Lancet Reg Heal Eur. 2021;2:100021. https://doi.org/10.1016/j.lanepe.2020.100021
34. Politi J, Martín-Sánchez M, Mercuriali L, Borras-Bermejo B, Lopez-Contreras J, Vilella A, et al.; COVID-19 Surveillance Working Group of Barcelona. Epidemiological characteristics and outcomes of COVID-19 cases: mortality inequalities by socio-economic status, Barcelona, Spain, 24 February to 4 May 2020. Euro Surveill. 2021;26:2001138. https://doi.org/10.2807/1560-7917.ES.2021.26.20.2001138
35. Marí-Dell'Olmo M, Gotsens M, Pasarín MI, Rodríguez-Sanz M, Artazcoz L, Garcia de Olalla P, et al. Socioeconomic inequalities in COVID-19 in a European urban area: Two waves, two patterns. Int J Environ Res Public Health. 2021;18:1–12. https://doi.org/10.3390/ijerph18031256
36. Randolph HE, Barreiro LB. Herd immunity: understanding COVID-19. Immunity. 2020;52:737–41. https://doi.org/10.1016/j.immuni.2020.04.012
37. Nafilyan V, Dolby T, Finning K, Morgan J, Edge R, Glickman M, et al. Differences in COVID-19 vaccination coverage by occupation in England: a national linked data study. Occup Environ Med. 2022 Sept 22 [Epub ahead of print]. https://doi.org/10.1136/oemed-2021-108140

Address for 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, 08007 Barcelona, Spain; email: tduarte@idiapjgol.org

## 5.5. Study V: Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain

Burn E,* Roel E,* Pistillo A, Fernández-Bertolín S, Aragón M, Raventós B, Reyes C, Verhamme K, Rijnbeek P, Li X, Strauss VY, Prieto-Alhambra D, Duarte-Salles T.

*Nature Communications*. 2022;13(1):7169.
Doi:10.1038/s41467-022-34669-9

nature communications

Article

https://doi.org/10.1038/s41467-022-34669-9

# Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain

Received: 20 August 2021

Accepted: 1 November 2022

Published online: 23 November 2022

Check for updates

A list of authors and their affiliations appears at the end of the paper

Population-based studies can provide important evidence on the safety of COVID-19 vaccines. Here we compare rates of thrombosis and thrombocytopenia following vaccination against SARS-CoV-2 with the background (expected) rates in the general population. In addition, we compare the rates of the same adverse events among persons infected with SARS-CoV-2 with background rates. Primary care and linked hospital data from Catalonia, Spain informed the study, with participants vaccinated with BNT162b2 or ChAdOx1 (27/12/2020-23/06/2021), COVID-19 cases (01/09/2020-23/06/2021) or present in the database as of 01/01/2017. We included 2,021,366 BNT162b2 (1,327,031 with 2 doses), 592,408 ChAdOx1, 174,556 COVID-19 cases, and 4,573,494 background participants. Standardised incidence ratios for venous thromboembolism were 1.18 (95% CI 1.06-1.32) and 0.92 (0.81-1.05) after first- and second dose BNT162b2, and 0.92 (0.71-1.18) after first dose ChAdOx1. The standardised incidence ratio for venous thromboembolism in COVID-19 was 10.19 (9.43-11.02). Standardised incidence ratios for arterial thromboembolism were 1.02 (0.95-1.09) and 1.04 (0.97-1.12) after first- and second dose BNT162b2, 1.06 (0.91-1.23) after first-dose ChAdOx1 and 4.13 (3.83-4.45) for COVID-19. Standardised incidence ratios for thrombocytopenia were 1.49 (1.43-1.54) and 1.40 (1.35-1.45) after first- and second dose BNT162b2, 1.28 (1.19-1.38) after first-dose ChAdOx1 and 4.59 (4.41- 4.77) for COVID-19. While rates of thrombosis with thrombocytopenia were generally similar to background rates, the standardised incidence ratio for pulmonary embolism with thrombocytopenia after first-dose BNT162b2 was 1.70 (1.11-2.61). These findings suggest that the safety profiles of BNT162b2 and ChAdOx1 are similar, with rates of adverse events seen after vaccination typically similar to background rates. Meanwhile, rates of adverse events are much increased for COVID-19 cases further underlining the importance of vaccination.

The advent of the coronavirus disease (COVID-19) vaccines has raised hopes of an end of the COVID-19 pandemic. As of June 2021, the European Medicines Agency (EMA) has now authorised four vaccines against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19[1]. Of these, two are mRNA vaccines: BNT162b2 mRNA (manufactured by Pfizer-BioNtech, approved on 21 December 2020) and mRNA-1273 (Moderna, 6 January 2021); and two are adenovirus-based vaccines: ChAdOx1 nCoV-19

e-mail: daniel.prietoalhambra@ndorms.ox.ac.uk; tduarte@idiapjgol.org

(Oxford–AstraZeneca, 29 January 2021), from now on ChAdOx1, and Ad.26.COV2.S (Janssen, 11 March 2021). These vaccines have shown a high degree of efficacy against symptomatic COVID-19 in randomised trials (70-95%);[2–5] and an increasing body of real-world evidence shows that they are highly effective in reducing infections, hospitalisations, and deaths[6,7]. However, concerns have been raised regarding their safety.

In March 2021, several European countries paused and/or suspended vaccination with ChAdOx1 after spontaneous reports of unusual thromboembolic events associated with thrombocytopenia among ChAdOx1 recipients[8]. As of 21 March, the EMA had reported 62 cases of cerebral venous sinus thrombosis (CVST) and 24 cases of splanchnic venous thrombosis (SVT) in the European Union and the United Kingdom in relation to ChAdOx1 (with 25 million doses administered to that date)[9]. Similarly, in the United States, 17 thromboembolic events with thrombocytopenia had been reported as of April 2021 in relation to Ad.26.COV2.S vaccine (with nearly 8 million doses administered)[10]. Subsequently, two population-based studies found increased rates of thromboembolic events among people vaccinated with one dose of ChAdOx1 when compared to expected rates in the general population[11,12]. Although far less reported, the occurrence of thrombocytopenia has also been seen among some recipients of mRNA vaccines[13]. With millions of people already having been vaccinated against COVID-19, population-based studies can provide important evidence on the safety of COVID-19 vaccines.

In this work we aimed to describe the incidence rates of thrombosis and thrombocytopenia after vaccination with BNT162b2 (first and second-dose) and ChAdOx1 (first-dose) and to compare these with the rates seen among the general population before the COVID-19 pandemic. In addition, we set out to estimate incidence rates of the same events for persons with COVID-19 to help contextualise our findings.

## Results

We included 2,613,774 people vaccinated against SARS-CoV-2 (2,021,366 with a first-dose of BNT162b2 and 592,408 with a first-dose of ChAdOx1) in the study. A total of 1,327,031 second doses were observed for BNT162b2 (65.7%), with 74% of these seen at exactly 21 days after the first, 1% seen on days 18–20 and the remaining 24% seen between days 22–28. In addition, we included 174,556 COVID-19 cases and 4,573,494 people in the general population cohort. People vaccinated with BNT162b2 were on average younger than those vaccinated with ChAdOx1, see Table 1, but with a much wider age distribution, Fig. 1. Indeed, in accordance with national guidelines, the majority of those vaccinated with ChAdOx1 (72.2%) were aged 60–69 years. Vaccinations over time are shown in Fig. 2, stratified by age group. Recipients of both BNT162b2 and ChAdOx1 were typically older than COVID-19 cases and those in the general population cohort. Supplementary Table 1 shows participants' characteristics by cohort and age category.

### Venous thromboembolism events

In the first 21 days following a first-dose of BNT162b2, we observed 182 instances of deep vein thrombosis (DVT), which compared with 176 expected events (standardised incidence rate [SIR]: 1.03 [95% CI 0.89–1.19]), and 154 occurrences of pulmonary embolism (PE), which compared with 123 expected events (1.25 [1.07-1.46]). After a second-dose of BNT162b2, however, 130 instances of DVT were observed which compared with 162 expected (0.80 [0.67–0.95]), while 116 PE were seen which compared with 115 expected (1.00 [0.84–1.20]). For the first-dose ChAdOx1 cohort, we saw 39 DVT, which compared with 43 expected (0.89 [0.65–1.22]), and 24 PE, which compared with 30 expected (0.78 [0.52–1.16]). Instances of splanchnic venous thrombosis were in line with expected rates for all the vaccinated cohorts, but were higher for the COVID-19 diagnosed cohort (2.64 [1.53–4.55]).

Overall, rates of venous thromboembolism (VTE, a composite of DVT and PE) were higher than expected after first-dose of BNT162b2 (1.18 [1.06–1.32]) but in line with expected rates after second-dose of BNT162b2 (0.92 [0.81–1.05]) and first-dose of ChAdOx1 (0.92 [0.81–1.05]). In comparison, while 61 occurrences of VTE would have been expected among COVID-19 cases in the absence of the disease, we saw 630 instances of VTE among this cohort (10.19 [9.43–11.02]), Table 2 and Fig. 3.

Incidence rates by age are shown in Fig. 4, with incidence rates increasing with age but broadly similar among those vaccinated as for the general population. Incidence rate ratios (IRR) by age group are shown in Fig. 5, with risks much increased among COVID-19 cases compared to the background population for all the age groups.

### Arterial thromboembolism events

Rates of arterial thromboembolism (ATE, a composite of myocardial infarction and ischaemic stroke) after vaccination were similar to expected rates for both vaccines. We observed 793 ATE after a first dose of BNT162b2, which compared with 780 expected (SIR: 1.02 [0.95–1.09]), 774 after a second-dose of BNT162b2, which compared with 741 expected (1.04 [0.97–1.12]), and 178 after a first dose of ChAdOx1, which compared with 168 expected (1.06 [0.91–1.23]). Conversely, rates of ATE for COVID-19 cases were much higher than expected, with 683 ATE seen among COVID-19 cases, which compared with 165 expected (4.13 [3.83–4.45]).

Incidence rates of ATE also increased by age (Fig. 4). However, when stratifying risks of ATE by age group, we observed higher than expected IRRs among first-dose BNT162b recipients aged 50–59 (IRR: 1.33 [1.11–1.58]) and 60-69 years (1.54 [1.09–2.13]), and second-dose BNT162b2 recipients aged 70–79 years (1.19 [1.06–1.33]) compared to the general population. These risks where much increased among COVID-19 cases for all the age groups.

### Thrombocytopenia

We observed more cases of thrombocytopenia than expected following both first- and second-dose of BNT162b2 and first-dose of ChAdOx1. Following a first-dose of BNT162b2, we observed 3186 cases, which compared with 2145 expected (SIR: 1.49 [1.43–1.54]). After second-dose, 2749 cases were seen, which compared with 1963 expected (1.40 [1.35–1.45]). Following a first-dose of ChAdOx1, we observed 754 cases, which compared with 588 expected (1.28 [1.19–1.38]). Among COVID-19 cases, 2476 cases were seen which compared to 560 expected (4.59 [4.41–4.77]).

When stratifying by age group, we found higher than background (expected) IRRs for thrombocytopenia after first-dose of BNT162b2 among individuals aged 50 years and older, Fig. 5. For example, for individuals aged 60–69 years, IRRs were 1.50 (1.25–1.80). Conversely, for first-dose of ChAdOx1, IRRs remained significant only for individuals aged 60–69 years (IRR: 1.30 [1.20–1.41]), although there was greater uncertainty around estimates for this vaccine among younger age groups.

Diagnoses of immune thrombocytopenia were though less or equal to expected for both vaccines. In total, 97 cases were seen after a first-dose of BNT162b2, which compared to 94 expected (SIR: 1.03 [0.84–1.26]), 61 after a second-dose of BNT162b2, which compared to 89 expected (0.69 [0.53–0.88]), and 12 after a first-dose of ChAdOx1, which compared to 24 expected (0.48 [0.27–0.85]). Conversely, 292 cases were seen among those COVID-19 cases, which compared with 22 expected (13.29 [11.85–14.91]).

### Thrombosis with thrombocytopenia

Rates of deep vein thrombosis with concomitant thrombocytopenia were in line with expected rates in all the vaccinated cohorts. Conversely, rates of pulmonary embolism with thrombocytopenia were

**Table 1 | Characteristics of study participants**

| | General population | BNT162b2 first-dose | BNT162b2 second-dose | ChAdOx1 first-dose | COVID-19 cases |
|---|---|---|---|---|---|
| *N* | 4,573,494 | 2,021,366 | 1,327,031 | 592,408 | 174,556 |
| Age | 48 [37–63] | 55 [45–75] | 70 [54–79] | 62 [59–65] | 47 [35–61] |
| Age: 20–29 | 572,054 (12.5%) | 41,288 (2.0%) | 32,011 (2.4%) | 25,029 (4.2%) | 27,629 (15.8%) |
| Age: 30–39 | 863,539 (18.9%) | 183,972 (9.1%) | 35,774 (2.7%) | 34,221 (5.8%) | 29,023 (16.6%) |
| Age: 40–49 | 979,378 (21.4%) | 527,768 (26.1%) | 105,665 (8.0%) | 48,145 (8.1%) | 38,071 (21.8%) |
| Age: 50–59 | 776,128 (17.0%) | 482,366 (23.9%) | 398,367 (30.0%) | 57,038 (9.6%) | 31,446 (18.0%) |
| Age: 60–69 | 599,531 (13.1%) | 69,304 (3.4%) | 59,135 (4.5%) | 427,885 (72.2%) | 19,023 (10.9%) |
| Age: 70–79 | 428,023 (9.4%) | 401,584 (19.9%) | 389,924 (29.4%) | 86 (0.0%) | 13,797 (7.9%) |
| Age: 80 or older | 354,841 (7.8%) | 315,084 (15.6%) | 306,155 (23.1%) | <5 | 15,567 (8.9%) |
| Sex: Male | 2,233,092 (48.8%) | 928,908 (46.0%) | 569,285 (42.9%) | 270,395 (45.6%) | 80,723 (46.2%) |
| Years of prior observation time | 11.0 [11.0–11.0] | 15.3 [15.2–15.4] | 15.3 [15.2–15.4] | 15.3 [15.2–15.3] | 14.8 [14.7–15.1] |
| **Comorbidities** | | | | | |
| Autoimmune disease | 78,636 (1.7%) | 50,255 (2.5%) | 41,345 (3.1%) | 14,485 (2.4%) | 3843 (2.2%) |
| Antiphospholipid syndrome | 988 (0.0%) | 1404 (0.1%) | 992 (0.1%) | 368 (0.1%) | 104 (0.1%) |
| Thrombophilia | 2708 (0.1%) | 2922 (0.1%) | 1939 (0.1%) | 679 (0.1%) | 264 (0.2%) |
| Asthma | 264,435 (5.8%) | 140,284 (6.9%) | 92,296 (7.0%) | 36,297 (6.1%) | 13,219 (7.6%) |
| Atrial fibrillation | 137,259 (3.0%) | 111,726 (5.5%) | 105,525 (8.0%) | 13,654 (2.3%) | 7339 (4.2%) |
| Malignant neoplastic disease | 338,633 (7.4%) | 243,671 (12.1%) | 219,790 (16.6%) | 60,079 (10.1%) | 15,408 (8.8%) |
| Diabetes mellitus | 464,169 (10.1%) | 286,446 (14.2%) | 245,320 (18.5%) | 80,496 (13.6%) | 20,803 (11.9%) |
| Obesity | 851,541 (18.6%) | 520,035 (25.7%) | 407,283 (30.7%) | 160,914 (27.2%) | 42,536 (24.4%) |
| Heart disease | 566,359 (12.4%) | 408,181 (20.2%) | 366,302 (27.6%) | 87,417 (14.8%) | 26,193 (15.0%) |
| Hypertensive disorder | 1,138,877 (24.9%) | 709,293 (35.1%) | 632,457 (47.7%) | 200,989 (33.9%) | 43,554 (25.0%) |
| Renal impairment | 229,003 (5.0%) | 195,554 (9.7%) | 184,698 (13.9%) | 21,368 (3.6%) | 12,887 (7.4%) |
| COPD | 165,700 (3.6%) | 109,761 (5.4%) | 100,699 (7.6%) | 30,281 (5.1%) | 7265 (4.2%) |
| Dementia | 72,171 (1.6%) | 55,270 (2.7%) | 52,793 (4.0%) | 1198 (0.2%) | 5295 (3.0%) |
| **Medication use (183 days prior to four days prior)** | | | | | |
| Non-steroidal anti-inflammatory drugs | 1,259,998 (27.6%) | 515,731 (25.5%) | 355,922 (26.8%) | 147,232 (24.9%) | 48,887 (28.0%) |
| Cox2 inhibitors | 26,822 (0.6%) | 17,364 (0.9%) | 13,342 (1.0%) | 6486 (1.1%) | 1099 (0.6%) |
| Systemic corticosteroids | 255,602 (5.6%) | 128,708 (6.4%) | 99,801 (7.5%) | 32,176 (5.4%) | 10,620 (6.1%) |
| Antithrombotic and anticoagulant therapies | 110,297 (2.4%) | 70,594 (3.5%) | 58,385 (4.4%) | 16,964 (2.9%) | 5885 (3.4%) |
| Lipid modifying agents | 80,526 (1.8%) | 46,175 (2.3%) | 38,984 (2.9%) | 18,339 (3.1%) | 2736 (1.6%) |
| Antineoplastic and immunomodulating agents | 56,526 (1.2%) | 25,965 (1.3%) | 18,655 (1.4%) | 7011 (1.2%) | 3101 (1.8%) |
| Hormonal contraceptives for systemic use | 40,488 (0.9%) | 14,959 (0.7%) | 6190 (0.5%) | 2595 (0.4%) | 3087 (1.8%) |
| Tamoxifen | 1207 (0.0%) | 713 (0.0%) | 525 (0.0%) | 187 (0.0%) | 56 (0.0%) |
| Sex hormones and modulators of the genital system | 51,800 (1.1%) | 20,498 (1.0%) | 10,108 (0.8%) | 4475 (0.8%) | 3592 (2.1%) |
| One or more condition of interest* | 1,449,480 (31.7%) | 880,600 (43.6%) | 716,224 (54.0%) | 249,505 (42.1%) | 65,949 (37.8%) |
| One or more medication of interest† | 1,361,157 (29.8%) | 567,169 (28.1%) | 392,729 (29.6%) | 159,770 (27.0%) | 53,952 (30.9%) |
| One or more condition/medication of interest*† | 2,244,217 (49.1%) | 1,148,815 (56.8%) | 866,134 (65.3%) | 327,718 (55.3%) | 94,795 (54.3%) |

Characteristics of the participants in the study cohorts used for the primary analyses. Participants were aged 20 years or older and had at least one year of prior history before index date in the database. Those in the general population were present in the database as of 01/01/2017. *Conditions of interest: autoimmune disease, antiphospholipid syndrome, thrombophilia, asthma, atrial fibrillation, malignant neoplastic disease, diabetes mellitus, obesity, or renal impairment. †Medications of interest: non-steroidal anti-inflammatory drugs, Cox2 inhibitors, systemic corticosteroids, hormonal contraceptives, tamoxifen, and sex hormones and modulators of the genital system.

higher than expected among first-dose BNT162b2 recipients, with 21 observed events which compared with 12 expected (SIR: 1.70 [1.11–2.61]). These rates were more in line with expected events for second-dose of BNT162b2 and ChAdOx1. Rates of VTE with thrombocytopenia after vaccination with BNT162b2 were low and in line with expected rates. We observed 27 such events after a first-dose of BNT162b2, which compared with 21 expected (1.24 [0.85–1.82]), and 20 after a second-dose, which compared with 20 expected (0.98 [0.63–1.52]). For ChAdOx1, 8 events were seen which compared to 6 expected (1.28 [0.64–2.56]).

As for rates of ATE with thrombocytopenia, these were lower than expected with 53 events observed after a first-dose of BNT162b2, which compared with 62 expected (0.86 [0.65–1.12]), and 59 after a second-dose, which compared with 59 expected (0.99 [0.76–1.27]). For ChAdOx1, 9 events were observed, which compared with 11 expected (0.77 [0.40–1.48]).

### Sensitivity analyses

We found similar results in sensitivity analysis after (1) removing the requirement for at least one year of prior history available in all cohorts; (2) stratifying participants by calendar time (before March 2021 and from March 2021 onwards); and (3) excluding individuals with prior COVID-19 among those vaccinated; see Supplementary Tables 2–4. All results from primary and

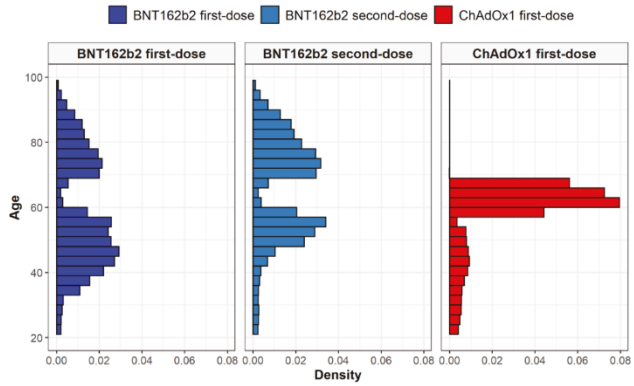



**Fig. 1 | Age profiles of people vaccinated against SARS-CoV-2.** Age distribution by vaccine and dose type. Dark blue - First-dose BNT162b2; Light blue - Second-dose BNT162b2; Red - First-dose ChAdOx1.

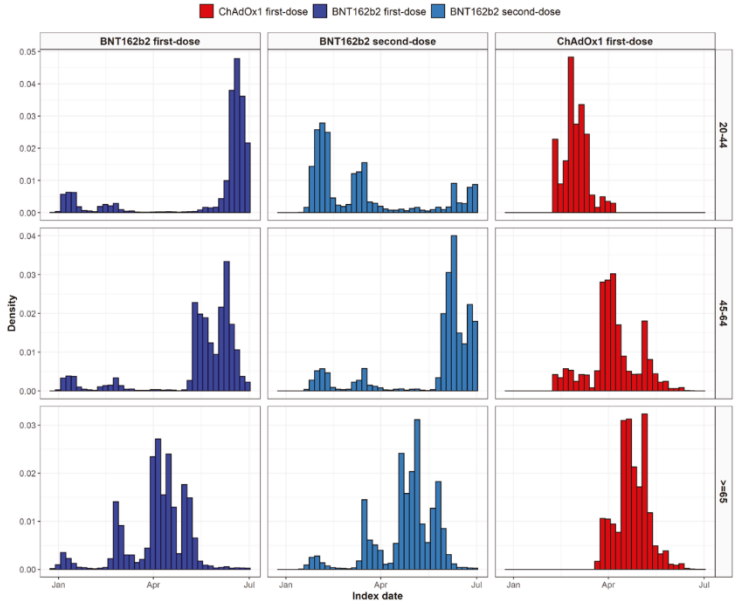



**Fig. 2 | Distribution of date of cohort entry among people vaccinated against SARS-CoV-2 by age group.** Distribution of date of cohort entry by vaccine type and age group. Dark blue - First-dose BNT162b2; Light blue - Second-dose BNT162b2; Red - First-dose ChAdOx1.

**Table 2 | Observed versus expected events among people vaccinated against SARS-CoV-2 or with a diagnosis of COVID-19**

| | *N* | Person-years | Observed events | Expected events | Standardised incidence ratio (95% CI) |
|---|---|---|---|---|---|
| **Deep vein thrombosis** | | | | | |
| ChAdOx1 first-dose | 590,599 | 33,928 | 39 | 43.9 | 0.89 (0.65–1.22) |
| BNT162b2 first-dose | 2,017,986 | 100,380 | 182 | 176.4 | 1.03 (0.89–1.19) |
| BNT162b2 second-dose | 1,324,426 | 78,253 | 130 | 162.6 | 0.80 (0.67–0.95) |
| COVID-19 case | 174,081 | 40,201 | 199 | 42.5 | 4.68 (4.07–5.38) |
| **Pulmonary embolism** | | | | | |
| ChAdOx1 first-dose | 590,761 | 33,938 | 24 | 30.7 | 0.78 (0.52–1.16) |
| BNT162b2 first-dose | 2,018,466 | 100,406 | 154 | 123.2 | 1.25 (1.07–1.46) |
| BNT162b2 second-dose | 1,324,890 | 78,280 | 116 | 115.8 | 1.00 (0.84–1.20) |
| COVID-19 case | 173,905 | 40,110 | 500 | 28 | 17.86 (16.37–19.50) |
| **Splanchnic vein thrombosis** | | | | | |
| ChAdOx1 first-dose | 591,119 | 33,959 | 8 | 7.3 | 1.09 (0.54–2.18) |
| BNT162b2 first-dose | 2,020,842 | 100,542 | 14 | 18.4 | 0.76 (0.45–1.29) |
| BNT162b2 second-dose | 1,326,894 | 78,401 | 9 | 16.7 | 0.54 (0.28–1.04) |
| COVID-19 case | 174,421 | 40,298 | 13 | 4.9 | 2.64 (1.53–4.55) |
| **Venous thromboembolism (deep vein thrombosis or pulmonary embolism)** | | | | | |
| ChAdOx1 first-dose | 590,307 | 33,911 | 60 | 65.5 | 0.92 (0.71–1.18) |
| BNT162b2 first-dose | 2,015,999 | 100,266 | 313 | 264.6 | 1.18 (1.06–1.32) |
| BNT162b2 second-dose | 1,322,747 | 78,151 | 227 | 246.1 | 0.92 (0.81–1.05) |
| COVID-19 case | 173,608 | 40,031 | 630 | 61.8 | 10.19 (9.43–11.02) |
| **Myocardial infarction** | | | | | |
| ChAdOx1 first-dose | 590,016 | 33,894 | 72 | 76.1 | 0.95 (0.75–1.19) |
| BNT162b2 first-dose | 2,016,950 | 100,320 | 280 | 267.3 | 1.05 (0.93–1.18) |
| BNT162b2 second-dose | 1,323,543 | 78,197 | 272 | 247.3 | 1.10 (0.98–1.24) |
| COVID-19 case | 173,975 | 40,191 | 118 | 63.7 | 1.85 (1.55–2.22) |
| **Ischaemic stroke** | | | | | |
| ChAdOx1 first-dose | 589,870 | 33,884 | 106 | 95.6 | 1.11 (0.92–1.34) |
| BNT162b2 first-dose | 2,012,434 | 100,055 | 521 | 530.5 | 0.98 (0.90–1.07) |
| BNT162b2 second-dose | 1,319,429 | 77,948 | 515 | 510.9 | 1.01 (0.92–1.10) |
| COVID-19 case | 173,390 | 39,983 | 577 | 106.5 | 5.42 (5.00–5.88) |
| **Arterial thromboembolism (myocardial infarction or ischaemic stroke)** | | | | | |
| ChAdOx1 first-dose | 588,689 | 33,815 | 178 | 168.2 | 1.06 (0.91–1.23) |
| BNT162b2 first-dose | 2,008,196 | 99,814 | 793 | 780.7 | 1.02 (0.95–1.09) |
| BNT162b2 second-dose | 1,315,777 | 77,726 | 774 | 741.9 | 1.04 (0.97–1.12) |
| COVID-19 case | 172,905 | 39,868 | 683 | 165.4 | 4.13 (3.83–4.45) |
| **Immune thrombocytopenia** | | | | | |
| ChAdOx1 first-dose | 590,919 | 33,947 | 12 | 24.9 | 0.48 (0.27–0.85) |
| BNT162b2 first-dose | 2,019,395 | 100,459 | 97 | 94.3 | 1.03 (0.84–1.26) |
| BNT162b2 second-dose | 1,325,627 | 78,325 | 61 | 89 | 0.69 (0.53–0.88) |
| COVID-19 case | 174,145 | 40,188 | 292 | 22 | 13.29 (11.85–14.91) |
| **Thrombocytopenia** | | | | | |
| ChAdOx1 first-dose | 577,534 | 33,156 | 754 | 588.5 | 1.28 (1.19–1.38) |
| BNT162b2 first-dose | 1,952,174 | 96,641 | 3186 | 2145.40 | 1.49 (1.43–1.54) |
| BNT162b2 second-dose | 1,264,798 | 74,678 | 2749 | 1963.20 | 1.40 (1.35–1.45) |
| COVID-19 case | 167,218 | 38,379 | 2476 | 539.9 | 4.59 (4.41–4.77) |
| **Deep vein thrombosis with thrombocytopenia** | | | | | |
| BNT162b2 first-dose | 2,020,823 | 100,541 | 12 | 12.8 | 0.94 (0.53–1.65) |
| BNT162b2 second-dose | 1,326,869 | 78,400 | 7 | 12 | 0.58 (0.28–1.22) |
| COVID-19 case | 174,409 | 40,295 | 25 | 3 | 8.32 (5.62–12.32) |
| **Pulmonary embolism with thrombocytopenia** | | | | | |
| BNT162b2 first-dose | 2,020,780 | 100,538 | 21 | 12.4 | 1.70 (1.11–2.61) |
| BNT162b2 second-dose | 1,326,837 | 78,398 | 15 | 11.7 | 1.29 (0.78–2.13) |
| COVID-19 case | 174,370 | 40,280 | 61 | 2.9 | 21.37 (16.63–27.46) |
| **Venous thromboembolism (deep vein thrombosis or pulmonary embolism) with thrombocytopenia** | | | | | |
| ChAdOx1 first-dose | 591,096 | 33,957 | 8 | 6.3 | 1.28 (0.64–2.56) |

**Table 2 (continued) | Observed versus expected events among people vaccinated against SARS-CoV-2 or with a diagnosis of COVID-19**

| | N | Person-years | Observed events | Expected events | Standardised incidence ratio (95% CI) |
|---|---|---|---|---|---|
| BNT162b2 first-dose | 2,020,650 | 100,531 | 27 | 21.7 | 1.24 (0.85–1.82) |
| BNT162b2 second-dose | 1,326,723 | 78,391 | 20 | 20.4 | 0.98 (0.63–1.52) |
| COVID-19 case | 174,337 | 40,272 | 78 | 5 | 15.48 (12.40–19.32) |
| **Arterial thromboembolism (myocardial infarction or ischaemic stroke) with thrombocytopenia** | | | | | |
| ChAdOx1 first-dose | 591,014 | 33,952 | 9 | 11.7 | 0.77 (0.40–1.48) |
| BNT162b2 first-dose | 2,020,074 | 100,497 | 53 | 62 | 0.86 (0.65–1.12) |
| BNT162b2 second-dose | 1,326,182 | 78,358 | 59 | 59.8 | 0.99 (0.76–1.27) |
| COVID-19 case | 174,297 | 40,269 | 58 | 12.6 | 4.62 (3.57–5.98) |

For each event of interest, the number of people contributing to the analysis from the target population, their person-years contributed, and the number of observed events are given. Expected events are estimated using indirect standardisation to the general population. Standardised incidence ratios (SIRs) with 95% confidence intervals (CIs) were estimated. Events with fewer than 5 occurrences were omitted to protect patient confidentiality.

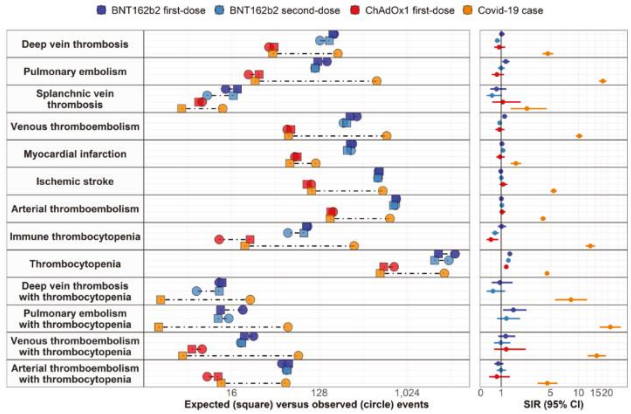



**Fig. 3 | Expected versus observed events among those vaccinated against SARS-CoV-2 and those with a COVID-19 diagnosis.** Expected events for each of the study cohorts based on indirect standardisation using rates from the general population between 2017 and 2019 are compared with the number of observed events seen in each cohort on the panels on the left. Corresponding standardised incidence ratios (SIRs) with 95% confidence intervals (95% CI) are shown in the panels on the right.

secondary analyses can be explored at: https://livedataoxford.shinyapps.io/SidiapCovidVaccinationStudy/.

## Discussion

In this study including 2,613,774 people vaccinated against SARS-CoV-2 in Catalonia, Spain, risks of thrombosis and thrombosis with thrombocytopenia after vaccination against COVID-19 were generally similar to background risks for the general population. We observed a potential safety signal for pulmonary embolism after a first dose of BNT162b2, with 31 more events than expected among 2 million vaccine recipients (SIR: 1.25 [1.07–1.46]). In addition, more cases of thrombocytopenia were seen than were expected following both first and second doses of BNT162b2 and first-dose of ChAdOx1. Rates of pulmonary embolism with concomitant thrombocytopenia were also increased among first-dose BNT162b recipients. In comparison, rates of VTE, ATE, and thrombocytopenia among COVID-19 cases were far higher than background rates. For instance, risks of VTE, ATE, and thrombocytopenia were 10-, 4- and 5-fold higher for COVID-19 cases. It is worth noting that differences exist in the socio-demographics and clinical characteristics of recipients of both vaccines compared to each other and compared to the general population. Confounding by indication may therefore, at least in part, explain the observed safety signals.

To date, population-based studies reporting thromboembolic events among SARS-CoV-2 vaccine recipients are scarce. A cohort study from Denmark and Norway which included 281,264 people aged between 18 and 65 years vaccinated with ChAdOx1 reported a 2-fold increase in VTE (SIR: 1.97 [1.50–2.54]) and a 3-fold increase in thrombocytopenia (SIR: 3.02 [1.76–4.83]) within 28 days of vaccination[11]. VTE rates were largely driven by cerebral venous sinus thrombosis (CVST) events, and were higher among younger age groups (18-44 years SIR: 2.99 [1.94–4.42]). Further, in line with our results, Pottegård et al. did not observe increased rates of ATE. Differences between ours and Pottegård's results might be partially explained by the characteristics of ChAdOx1 vaccine recipients, who were mostly aged 60–69 years in our study.

In a nested case-control study including hospital data from Scotland, increased rates of idiopathic thrombocytopenic purpura (ITP),

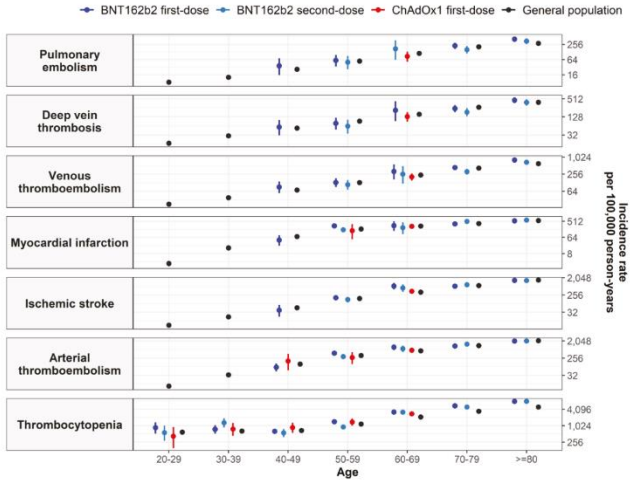



**Fig. 4 | Background and post-vaccine rates of thromboembolic events and thrombocytopenia by age.** Events with less than 5 occurrences have been omitted for privacy reasons. Point estimates with 95% confidence intervals.

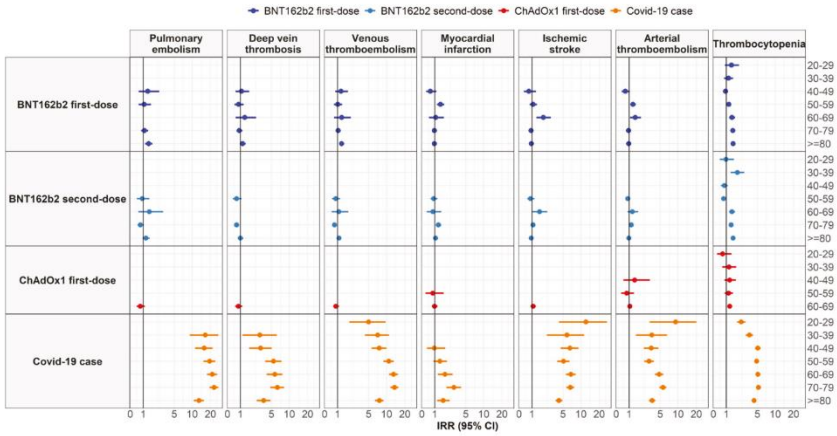



**Fig. 5 | Incidence rate ratios (IRRs) for thromboembolic events and thrombocytopenia by age.** Events with less than 5 occurrences have been omitted for privacy reasons. Point estimates with 95% confidence intervals.

ATE and haemorrhagic events were observed among 1.7 million people with a first-dose of ChAdOx1[12]. These findings were subsequently confirmed for ITP in a post hoc self-controlled case series (SCCS) analysis, but not for ATE or haemorrhagic events. In contrast to our results, the authors did not find increased rates of VTE among 821,052 people vaccinated with a first-dose of BNT162b2. Discrepancies between their and our results could be related to the study design, as well as to different sample sizes. As recently raised by Schuemie et al., case-control studies are prone to substantial bias when using retrospective databases and therefore must be interpreted with caution[14]. Further, our study included a larger cohort of BNT162b recipients but fewer and older ChAdOx1 vaccinees.

Another SCCS study from the UK found increased risks of hospital admissions or deaths associated with thromboembolism events

among vaccine recipients and individuals with SARS-CoV-2[15]. ChAdOx1, for which 19.6 million recipients were included, was associated with increased risks of thrombocytopenia, VTE, and CVST 8–14 days after first-dose vaccination, while BNT162b, with 9.5 million recipients, was associated with increased risks of ATE, ischaemic stroke, and CVST 15–21 days after first-dose vaccination. Interestingly, although we did not find an association between BT162b and ATE overall, we observed increased risks of ATE among BNT162b2 recipients aged 50-69 years. As also seen in our study, risks for these events were much higher among individuals with SARS-CoV-2. Again, sample sizes and the characteristics of ChAdOx1 recipients differed from ours, and only severe events leading to hospitalisation or death were captured in their study.

In a cohort study conducted by our group using primary care data from the UK, we also found increased risks of pulmonary embolism (SIR: 1.21 [1.07–1.36]) among 1.6 million BNT162b first-dose recipients[16]. Risks of pulmonary embolism were also seen to be increased among 1.8 million ChAdOx1 recipients. Rates of thrombocytopenia were also higher than expected for ChAdOx1, with similar SIR to those in Catalonia (SIR 1.25 [1.19–1.31]), but not for BNT162b recipients. However, rates of immune thrombocytopenia were higher than expected for both vaccines, with SIRs of 2.01 [1.27–3.19] for ChAdOx1 and 1.74 [1.05–2.89] for BNT162b2. Likewise, the differences between the two studies might be related to differences in study populations, with the UK study including younger and three-times as many ChAdOx1 recipients, but fewer BNT162b vaccinees. Differences could also be related to lack of hospital data in the UK study.

The mechanisms underlying a potential association between SARS-CoV-2 vaccines and thromboembolic events are currently under investigation. Shortly after the first signal alerts for the ChAdOx1 vaccine, case reports were published describing a potentially new immune disorder named vaccine-induced immune thrombotic thrombocytopenia (VITT) among ChadOx1 recipients[17,18]. According to the author's descriptions, this disorder manifests with atypical thrombotic events (e.g. CVST, SVT) associated with thrombocytopenia 5 to 15 days following vaccination and could be mediated by platelet-activating autoantibodies against platelet factor 4 (PF4). Regarding SARS-CoV-2 mRNA vaccines, some authors have suggested that the inflammatory response following vaccination might increase macrophage-mediated clearance and/or diminish platelet production, thus leading to thrombocytopenia[13]. Such mechanisms have been previously postulated in relation to ITP following viral infections (including SARS-CoV-2)[19], as well as following vaccination against other virus (e.g., varicella-zoster, measles-mumps-rubella)[20,21]. We did find increased risks of pulmonary embolism with thrombocytopenia among BNT162b2 recipients and increased rates of thrombocytopenia among BNT162b2 and ChadOx1 recipients. However, risks of thrombocytopenia and TTS were far higher among COVID-19 cases. Further research is needed to confirm our observations in other large population-based cohorts and to establish its pathogenesis.

The main strength of our study is its large sample size and representativeness, which allowed us to assess the incidence of rare adverse events in a real-world setting. In addition, to our knowledge, this is the largest study to-date reporting thromboembolic events following BNT162b2 second-dose. Our study was underpinned by a well-established primary care database that has previously been used for various post-authorisation safety studies[22,23], linked to hospital diagnoses, and was designed in collaboration with the study funder, the EMA. Finally, for the sake of transparency and reproducibility, we have made our protocol, analytical code and full result set publicly available.

However, our study also has several limitations. First, vaccinated cohorts differed substantially from the general population and from one another. Our data mirrored the different, and changing, nationwide guidelines for the provision of BNT162b2 and ChAdOx1 in Spain. Although the distribution of comorbidities and medication was broadly similar across cohorts when stratifying by age categories, vaccine recipients were generally in slightly worse health than the general population. Therefore, even though we calculated SIRs to allow comparisons between cohorts, we cannot exclude that residual confounding by indication might have influenced our results. In addition, the majority of ChAdOx1 vaccine recipients were older than 60 years in our study, with prior evidence suggesting a stronger association between this vaccine and VTE among individuals younger than 50 years[11,15]. This could have prevented us from observing an association between ChAdOx1 and VTE. Secondly, thromboembolic events were identified using routinely collected data, and therefore the lack of formal adjudication of outcomes might have introduced measurement error. Events might also be underestimated in all cohorts, especially rare events with complex diagnoses, such as immune thrombocytopenia or CVST, which we were unable to identify. Although underestimation of events was likely non-differential across cohorts for events before March 2021, we cannot exclude that detection bias following the EMA's signal alert report might have led to differential measurement error between the vaccinated and comparator cohorts from March onwards. However, our results were consistent in analysis stratified by time period for the first-dose BNT162b2 and ChadOx1 cohorts. Thirdly, individuals with a thromboembolic event after a first-dose were by design excluded from an analysis of the same event after a second-dose and this exclusion could be considered as leading to depletion of susceptibles for the second-dose cohort.

In summary, in this study of over two million people vaccinated against SARS-CoV-2 in Catalonia, Spain, the BNT162b2 and ChAdOx1 vaccines have been seen to have similar safety profiles. Safety signals for pulmonary embolism following BNT162b vaccination, as well as for thrombocytopenia following BNT162b2 and ChAdOx1 vaccination have been identified. Our results must be interpreted with caution considering the risk of residual confounding by indication. Regardless of the vaccine used, the increase in rates of thrombosis among persons with COVID-19 is far higher than any potential safety signal seen among persons vaccinated.

## Methods

### Study design and data source

We conducted a population-based cohort study using data from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org), a primary care database from Catalonia. This study was approved by the Clinical Research Ethics Committee of the IDIAPJGol (project code: 21/054-PCV). SIDIAP includes pseudo-anonymized primary care electronic health records from 80% of the population in Catalonia and is representative of the general population in terms of age, sex, and geographic distribution[24] SIDIAP was linked to discharge data from hospitals in Catalonia, with this data including both diagnosis and procedures registered during hospital admissions. The database has been mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), a structure that facilitates safety surveillance in observational health care databases[25].

### Setting

The national COVID-19 vaccination campaign was launched in Spain on 27 December 2020. Spain established priority subpopulation groups eligible for vaccination, differentiating the general population and essential workers[26]. Among the general population, vaccination began with those considered most at risk of severe disease (people living in nursing homes, older people, and people with risk factors for COVID-19), whereas among essential workers vaccination began with healthcare professionals followed by others (such as teachers and police officers). While the BNT162b2 vaccine has consistently been used across both of these priority groups, guidelines for ChAdOx1 have

changed over time. For instance, ChAdOx1 was initially restricted to essential workers aged 55 years or younger, but its use became restricted to those aged 60 to 65 and, subsequently, from 60 to 69 years after safety concerns emerged in March. Of note, we did not include mRNA-1273, Ad.26.COV2.S and second-dose ChAdOx1 recipients in this study due to small sample sizes.

### Study participants and follow-up time

Three vaccination cohorts were identified: BNT162b2 first-dose, BNT162b2 second-dose, and ChAdOx1 first-dose. Individuals vaccinated were required to have received their vaccine between 27 December 2020 and 23 June 2021 (a week prior to the end of data availability), be aged ≥20 years at time of vaccination (with very few people under 20 vaccinated during the study period), and have at least a year of prior history available (so as to identify events of interest prior to vaccination). In addition, those with a second-dose of BNT162b2 were required to have received this dose between 18 to 28 days following their first.

A historical comparator general population cohort was identified. Individuals present in the database as of 1 January 2017 were included in this cohort, with this date used as their index date. In addition, a cohort of COVID-19 cases was identified between 1 September 2020 and 23 June 2021 (i.e. from the second-wave onwards in Spain). These individuals were identified on the basis of a positive PCR test for SARS-CoV-2. Only incident cases were included, with persons with a clinical diagnosis of COVID-19 or positive test for SARS-CoV-2 prior to 1 September 2020 excluded from this cohort and only the first-recorded positive PCR test used for those persons included. COVID-19 cases were also required to have not had a vaccination against COVID-19 prior to their index date. As with the cohorts of people vaccinated, individuals were required to be aged ≥20 years and have at least a year of prior history available.

Additional cohorts were generated for sensitivity analyses. We created analogous cohorts including all individuals aged ≥20 years with no requirement for the amount of prior history available. Subsequently, we excluded any individuals with a COVID-19 diagnosis prior to the index date for the vaccinated cohorts. Vaccination cohorts were also stratified by calendar time; before 1 March 2021 and from 1 March 2021 onwards.

For each specific outcome and cohort, we excluded individuals with an occurrence of the outcome the year prior to the index date. All cohorts were followed from index date to whichever came first of: end of follow-up (21 days for people vaccinated, 90 days for those diagnosed with COVID-19, and 31 December 2019 for the background general population cohort), end of data collection (26 May 2021), exit from the database, or occurrence of the outcome of interest. In addition, follow-up was censored at time of second-dose for the BNT162b2 and ChAdOx1 first-dose in the few cases where it occurred before 21 days.

### Outcomes

Venous thromboembolic events included deep vein thrombosis (DVT), pulmonary embolism (PE), and the composite venous thromboembolism (VTE, which included DVT and PE). We also assessed portal vein thrombosis and splanchnic venous thrombosis (SVT). Arterial thromboembolism (ATE) events included myocardial infarction and ischaemic stroke. We also identified stroke in general, for which we included both ischemic and haemorrhagic stroke. Thrombocytopenia was identified using diagnostic codes or a measurement of a platelet count between 10,000 and 150,000 platelets/microliter. Finally, thromboembolic events with concomitant thrombocytopenia were identified, where thrombocytopenia was seen in 10 days before and after the thromboembolic event.

### Statistical analyses

We first summarised the characteristics of individuals included in each cohort (socio-demographics, baseline comorbidities and drug prescriptions), with counts and percentages for categorical variables and median and interquartile ranges (IQR) for continuous variables. For each cohort and outcome, we described the total number of events observed and calculated incidence rates (IR) per 100,000 person-years, with exact 95% confidence intervals (CI), overall and stratified by age and sex. We calculated crude incidence rate ratios (IRRs), with 95% CI, for the vaccinated and COVID-19 cohorts compared against the background general population cohort, both overall and stratified by age and sex. We also estimated the number of events expected among the vaccinated and COVID-19 cohorts using indirect standardisation (10 year age bands), with the general population cohort as the standard population. We calculated standardised incidence ratios (SIRs) with 95% CI dividing the number of events observed by the number of events expected. A SIR above 1 indicates that the observed rate for a specific outcome was higher than what was expected in the population and was taken to indicate a safety signal for a given vaccine cohort.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## 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 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) or by contacting SIDIAP (sidiap@idiapjgol.info). Source data are provided with this paper.

## Code availability

The analytic code to perform the study is available at https://github.com/SIDIAP/CovidVaccinationAdverseEventsStudy (https://doi.org/10.5281/zenodo.6583871).

## References

1. European Medicines Agency. COVID-19 vaccines: authorised. https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorised (2021).
2. Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
3. Baden, L. R. et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).
4. Voysey, M. et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**, 99–111 (2021).
5. Sadoff, J. et al. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N. Engl. J. Med.* **384**, 2187–2201 (2021).
6. Dagan, N. et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021).
7. Vasileiou, E. et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. *Lancet* **397**, 1646–1657 (2021).
8. Wise, J. Covid-19: European countries suspend use of Oxford-AstraZeneca vaccine after reports of blood clots. *BMJ* **372**, n699 (2021).
9. European Medicines Agency. COVID-19 vaccine safety update Vaxzevria vaccine - 14 April 2021. https://www.ema.europa.eu/en/documents/covid-19-vaccine-safety-update/covid-19-vaccine-safety-update-vaxzevria-previously-covid-19-vaccine-astrazeneca-14-april-2021_en.pdf (2021).

10. Shay, D. K. et al. Safety Monitoring of the Janssen (Johnson & Johnson) COVID-19 Vaccine — United States, March–April 2021. *MMWR Recomm. Rep.* **70**, 680–684 (2021).
11. Pottegård, A. et al. Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: Population based cohort study. *BMJ* **373**, n1114 (2021).
12. Simpson, C. R. et al. First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland. *Nat. Med.* 1–8 https://doi.org/10.1038/s41591-021-01408-4 (2021).
13. Lee, E. J. et al. Thrombocytopenia following Pfizer and Moderna SARS-CoV-2 vaccination. *Am. J. Hematol.* **96**, 534–537 (2021).
14. Schuemie, M. J. et al. A plea to stop using the case-control design in retrospective database studies. *Stat. Med.* **38**, 4199–4208 (2019).
15. Hippisley-Cox, J. et al. Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: Self-controlled case series study. *BMJ* **374**, n1931 (2021).
16. Burn, E. et al. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2: a population-based cohort analysis. *medRxiv* (2021) https://doi.org/10.1101/2021.07.29.21261348.
17. Schultz, N. H. et al. Thrombosis and thrombocytopenia after ChAdOx1 nCoV-19 vaccination. *N. Engl. J. Med.* **384**, 2124–2130 (2021).
18. Greinacher, A. et al. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *N. Engl. J. Med.* **384**, 2092–2101 (2021).
19. Dotan, A. & Shoenfeld, Y. Perspectives on vaccine induced thrombotic thrombocytopenia. *J. Autoimmun.* **121**, 102663 (2021).
20. Perricone, C. et al. Immune thrombocytopenic purpura (ITP) associated with vaccinations: a review of reported cases. *Immunol. Res.* **60**, 226–235 (2014).
21. Cecinati, V., Principi, N., Brescia, L., Giordano, P. & Esposito, S. Vaccine administration and the development of immune thrombocytopenic purpura in children. *Hum. Vaccines Immunother.* **9**, 1158–1162 (2013).
22. Robinson, D. E. et al. Safety of oral bisphosphonates in moderate-to-severe chronic kidney disease: A Binational Cohort Analysis. *J. Bone Miner. Res.* **36**, 820–832 (2021).
23. Giner-Soriano, M. et al. Effectiveness and safety of drugs used for stroke prevention in a cohort of non-valvular atrial fibrillation patients from a primary care electronic database. *Pharmacoepidemiol. Drug Saf.* **26**, 97–107 (2017).
24. Bolíbar, B. et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Med. Clin.* **138**, 617–621 (2012).
25. Marc Overhage, J., Ryan, P. B., Reich, C. G., Hartzema, A. G. & Stang, P. E. Validation of a common data model for active safety surveillance research. *J. Am. Med. Inform. Assoc.* **19**, 54–60 (2012).
26. Grupo de Trabajo Técnico de Vacunación COVID-19 de la Ponencia de Programa y Registro de Vacunaciones. *Actualización 7 - Estrategia de vacunación frente a COVID- 19 en España*. https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/docs/COVID-19_Actualizacion7_EstrategiaVacunacion.pdf (2021).


## Acknowledgements

This study was funded by the European Medicines Agency in the form of a competitive tender (Lot ROC No EMA/2017/09/PE). This document expresses the opinion of the authors of the paper, and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties. E.R. was supported by Instituto de Salud Carlos III (grant number CM20/00174) and conducted this study as part of the doctoral program in methodology of biomedical research and public health at the Autonomous University of Barcelona. D.P.A. is funded through a National Institute for Health Research (NIHR) Senior Research Fellowship (Grant number SRF-2018-11-ST2-004).

## Author contributions

E.B., X.L., V.Y.S., D.P.A. and T.D.S. led study design. E.R., A.P., S.F.B., B.R., M.A. and T.D.S. led data collection and processing and analysis. K.V. and C.R. provided clinical input and contributed to the identification of study outcomes. P.R. led the coordination of the project and contracting. E.B., E.R., and D.P.A. led the drafting of the manuscript. All authors were involved in the interpretation of the results, and the critical review and approval of the manuscript.

## Competing interests

D.P.A.'s research group has received research grants from the European Medicines Agency, from the Innovative Medicines Initiative, from Amgen, Chiesi, and from UCB Biopharma; and consultancy or speaker fees from Astellas, Amgen and UCB Biopharma. The remaining authors declare no competing interests'

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34669-9.

**Correspondence** and requests for materials should be addressed to Daniel Prieto-Alhambra or Talita Duarte-Salles.

**Peer review information** *Nature Communications* thanks Toshiaki Iba, Eero Pukkala and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Edward Burn[1,2,5], Elena Roel[1,3,5], Andrea Pistillo[1], Sergio Fernández-Bertolín[1], Maria Aragón[1], Berta Raventós[1,3], Carlen Reyes[1], Katia Verhamme[4], Peter Rijnbeek[4], Xintong Li[2], Victoria Y. Strauss[2], Daniel Prieto-Alhambra[2,4,6] ✉ & Talita Duarte-Salles[1,6] ✉**

[1]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain. [2]Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK. [3]Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Barcelona, Spain. [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands. [5]These authors contributed equally: Edward Burn, Elena Roel. [4]These authors jointly supervised this work: Daniel Prieto-Alhambra, Talita Duarte-Salles. ✉e-mail: daniel.prietoalhambra@ndorms.ox.ac.uk; tduarte@idiapjgol.org

# 6. DISCUSSION

In this Thesis, we first described the baseline characteristics and the occurrence of 30-day outcomes of interest among individuals with COVID-19 infection living with obesity and with cancer history. Then, we investigated the associations between cancer and socioeconomic deprivation with COVID-19 infection and severity as well as the associations between socioeconomic deprivation and COVID-19 vaccine uptake. Finally, we investigated the risks of thromboembolic events following COVID-19 infection and vaccination. All the studies were conducted using RWD. In the Results section, we presented the results of the studies included in this Thesis, as well as the discussion of the findings of each specific objective of this Thesis. In this section, we provide an overview of the main findings, contributions to the literature, and methodological considerations, propose some recommendations for future research and discuss the implications of our findings for public health.

## 6.1. Main findings and contributions to the literature

### 6.1.1. Characteristics of COVID-19 cases

In **Studies I** and **II**, we described in depth the baseline characteristics and 30-day outcomes of interest of COVID-19 patients with obesity (**Study I**) and cancer history (**Study II**). To our knowledge, these studies contributed to the literature by providing the most comprehensive description of COVID-19 patients with obesity and with cancer to date. In **Study I**, we included over 600,000 patients from six different databases from Spain, the UK, and the US and compared more than 10,000 medical conditions between patients living with and without obesity. Patients living with obesity had more frequently over one thousand different health conditions (including, but not limited to, obesity-related comorbidities) when compared to patients

without obesity. This highlights the importance of carefully considering potential confounders when attempting to analyse the associations between obesity and health outcomes, such as COVID-19-related outcomes. Individuals living with obesity had also higher hospitalisation and intensive services requirement rates than those without obesity. We also reported 30-day rates of adverse events of interest among hospitalised COVID-19 patients, such as acute kidney injury or cardiac arrhythmia, among others.

The prevalence of obesity among diagnosed and hospitalised COVID-19 patients was larger than expected when compared to the prevalence of obesity in the countries that contributed to the study. Although our study did not attempt to draw any causal relation between obesity and COVID-19 severity, our results highlighted that individuals with obesity were a high-risk population for severe COVID-19 during the first wave of the pandemic. Later RWD studies addressing the associations between overweight/obesity and COVID-19 outcomes have consistently reported that individuals living with obesity (as well as with overweight) have increased risks of severe COVID-19 after accounting for potential confounders.[115-117]

In **Study II**, we described the characteristics of 366,050 patients diagnosed and 119,597 patients hospitalised with COVID-19 and with cancer history using data from eight different databases from Spain and the US. Additionally, we compared patients with cancer history hospitalised with COVID-19 to patients with cancer history hospitalised with influenza, as a benchmark. We found that breast and prostate cancers were the most common cancers among COVID-19 patients, which is in line with the most frequent cancer types in the general population.[224,225] However, we provided novel evidence showing that non-Hodgkin lymphoma, leukaemia, and multiple myeloma were more frequent than expected among both COVID-19 cohorts, thus suggesting that haematological cancers might be associated with higher risks of COVID-19 infection and severity. We later corroborated this

hypothesis in **Study III** (see sections 5.3. Study IIII and 6.1.2. Factors associated with COVID-19 infection and severity). Even though patients hospitalised with COVID-19 and with influenza had an overall similar distribution of age and comorbidities, we also provided novel evidence showing that those with COVID-19 experienced more frequently adverse outcomes than patients with influenza, such as ARDS or death. In consequence, our results suggested that COVID-19 is a more severe disease for cancer patients when compared to seasonal influenza, which is consistent with evidence from the general population.[226,227] However, our results should be interpreted with caution, since factors such as influenza vaccine and antivirals, as well as the pressure on hospitals during the first wave of the pandemic might have influenced our results (see sections 5.2. Study II and 6.2. Methodological considerations)

### 6.1.2. Factors associated with COVID-19 infection and severity

In **Studies III** and **IV**, we aimed to assess the associations between two different exposures of interest (cancer history and socioeconomic deprivation, respectively) and COVID-19 related outcomes. In **Study III**, we used a multistate framework to analyse the relation between cancer history and COVID-19 diagnosis, hospitalisation, and death during the first wave of the pandemic. We included 4.6 million adults living in Catalonia, Spain, of which 260,667 (5.6%) had a prior diagnosis of cancer. Cancer history was associated with increased risks of an outpatient COVID-19 diagnosis, direct COVID-19 hospitalisation (without a prior outpatient diagnosis), and COVID- 19-related death, after accounting for potential confounders. These associations were stronger for patients with a recent cancer diagnosis (<1 year), aged <70 years, and with haematological cancers. In analyses stratified by solid cancer type, breast, colorectal, lung, and bladder cancers were also associated with increased risks of

direct hospitalisation, with stronger associations among patients with a recent cancer diagnosis (<5 years). Bladder and lung cancers were also associated with increased risks of COVID-19-related death.

This study provided additional evidence showing that patients with cancer, especially those with a recent cancer diagnosis or with a haematological cancer, have higher risks of COVID-19 infection and severity. Although a few small studies in early stages of the pandemic found null associations between cancer and COVID-19 outcomes,[111,184,185] our results are in line with a body of consolidated evidence that shows that patients with cancer have increased risks of severe COVID-19.[228–231] This might be due to an impaired immune response, both at the humoral and cellular level, since patients with cancer have lower seroconversion rates and a lower proportion of SARS-CoV-2 $CD4^+$ and $CD8^+$ T cells following infection when compared to individuals without cancer.[230,232] SARS-CoV2- T cells and antibody titres are particularly low among patients with haematological cancers, as well as among patients undergoing anticancer therapy, which could explain the stronger association between cancer and severe COVID-19 outcomes among these groups.[232–235] In line with our findings, a few studies have also reported a stronger association between cancer and COVID-19-related death among younger age patients.[228,236,237] This could be due to differences by age group in terms of cancer types (with more indolent cancers among the eldest, e.g., prostate cancer) as well as anticancer treatments (with a higher probability of receiving immunosuppressive therapies among the youngest). Similarly, other reports have also shown that lung cancer patients are particularly at risk of severe COVID-19 outcomes, which could be related to prior tobacco exposure and chronic lung damage.[228,238] Likewise, bladder cancer is also heavily associated with tobacco exposure, this could explain our results for COVID-19-related death among cancer bladder patients. A meta-analysis study found increased risks of COVID-19-related death among

individuals with genitourinary cancers, although results where non-significant (Relative Risk: 1.11 [95% CI: 1.00; 1.24]).[228]

In **Study IV**, we analysed the relation between socioeconomic deprivation and COVID-19 infection, hospitalisation, and death among 2.3 million individuals living in urban areas of Catalonia, Spain, before and after the rollout of COVID-19 vaccines. To our knowledge, this study is the first to provide evidence on the impact of mass COVID-19 vaccination on infection inequalities. Before vaccine rollout, higher deprivation levels were associated with increased risks of COVID-19 infection and hospitalisation among working-age (40-64 years) and retirement-age (≥65 years) individuals. This is consistent with prior studies from other high-income countries,[127,128,130,135] as well as with two studies from Barcelona showing higher infection rates in more deprived areas.[112,239] Despite disparities in vaccine uptake among working-age individuals (see section 6.1.3. Vaccine uptake), socioeconomic inequalities in COVID-19 infections decreased six months after the introduction of COVID-19 vaccines among both age groups, suggesting that herd immunity might have contributed to the reduction of inequalities.[240] Conversely, inequalities in COVID-19 hospitalisation only decreased (but persisted) among retirement-age individuals. Since individuals living in more deprived areas had a poorer health status, we believe that persistent inequalities among retirement-age individuals might be due to disparities in the risk of developing severe COVID-19 once infected.[127]

### 6.1.3. Vaccine uptake

In **Study IV**, we also provided novel evidence regarding socioeconomic inequalities in COVID-19 vaccine uptake in Catalonia. To our knowledge, our study was the first to provide vaccination coverage rates by quintile of socioeconomic deprivation index in Catalonia. Interestingly, our results differed by age group. While among retirement-age individuals

vaccination coverage was similar across all deprivation quintiles (ranging from 90.6% to 91.8%), among working-age individuals it ranged from 70.6% in the most deprived quintile to 80.6% in the least deprived quintile. When accounting for potential confounders, we observed a pattern of higher odds of non-vaccination with higher levels of socioeconomic deprivation among working-age individuals, but not among retirement-age individuals. We hypothesised that differences in working conditions might influence inequalities in vaccine uptake among working-age individuals. The use of different strategies to invite individuals for vaccination might also explain age differences. Whereas individuals aged 70 years and older were contacted by and vaccinated at their primary care centres, those younger had to book themselves their vaccine appointment through a dedicated web page. In addition, they were mostly vaccinated at large vaccination points specifically deployed for the vaccination campaign.[241]

Large observational studies from the US and the UK have consistently found a lower vaccine uptake among individuals living in more deprived areas or with low SES.[191,192,242–248] Interestingly, a study that compared US cities with paid sick leave to those without observed inequalities in COVID-19 vaccine uptake among working-age individuals (with higher uptake in cities with paid sick leave) but not among retirement-age individuals.[244] The authors suggested that fear to miss work due to vaccine side effects might influence vaccine uptake. Alas, large and representative studies from other countries on this matter are still lacking. In Catalonia, to our knowledge, there is only one study reporting to some extent socioeconomic inequalities in COVID-19 coverage.[241] Malmusi et al described COVID-19 vaccine uptake among individuals aged 60-69 years living in Barcelona by basic health area (BHA). As of May 2021, vaccine uptake ranged from 40% in the BHA with the greatest levels of deprivation to 72% in the least deprived BHAs.[241] After implementing a targeted intervention to address this issue (which included setting up support points to help people to book a

vaccine appointment as well as walk-in vaccination sites without appointment), inequalities decreased. As of November 2021, vaccine coverage by BHA ranged from 82% to 95% among this age group.

### 6.1.4. Vaccine safety

In **Study V**, we compared incidence rates of thromboembolic events and thrombocytopenia following COVID-19 infection and vaccination to historical background rates. Over 2 million vaccinees and 170,000 individuals with COVID-19 infection were compared to 4.5 million historical controls. We found increased risks of thrombocytopenia following first and second dose of BNT162b2 vaccination as well as following first dose of ChAdOx1, with SIRs of 1.49 [95% CI: 1.43; 1.54], 1.40 [1.35; 1.45], and 1.28 [1.19; 1.38], respectively. We also observed increased rates of VTE and of pulmonary embolism with concomitant thrombocytopenia following first dose vaccination with BNT162b2, with SIRs of 1.18 [1.06; 1.32] and 1.70 [1.11; 2.61]. Conversely, rates of thrombocytopenia as well as VTE and ATE, were 5, 4, and 10 times higher than expected among COVID-19 cases, respectively. Residual confounding might explain, at least partially, our results among those vaccinated since vaccinees had a slightly worse health condition than historical controls.

The main contributions to the literature of this study were: i) the inclusion of individuals infected with COVID-19 to provide context to the findings among the vaccinees; ii) the inclusion of second-dose BNT162b2 vaccinees (with most studies limited to first doses); and iii) the inclusion of thromboembolic events with concomitant thrombocytopenia as an individual outcome (since this outcome has been scarcely reported in other studies due to its rarity as well as to the complexity of its definition).[163,249] The use of Catalan data is also an asset, since the majority of studies published to this date on this matter have been conducted in the UK[195,250-254] and in Nordic countries.[194,255] Table 1 provides a

summary of the main findings of RWD studies reporting the associations between COVID-19 vaccines and thrombosis or thrombocytopenia.

Overall, ChAdOx1 has been more frequently associated with increased risks of thrombosis than mRNA vaccines, particularly with VTE and with CVST. However, a few studies have also reported increased risks of VTE following BNT162b2 vaccination, in line with our results.[254,255] Importantly, absolute risks of thromboembolic events following COVID-19 vaccination have consistently been reported to be low.[171,254] Thrombocytopenia has also been associated with ChAdOx1 and, to a lesser extent, with BNT162b2 vaccines. In a multinational study including data from six countries, ChAdOx1 was associated with a pooled 30% increase of thrombocytopenia when compared to BNT162b2.[249] Increased risks of thrombocytopenia have been also documented previously in the context of other vaccines (e.g., against influenza,[256] hepatitis B,[257] and measles, mumps, and rubella),[258,259] thus suggesting that vaccines might trigger immune-mediate processes as part of the immune response that can potentially result in autoimmune reactions.[254] Inconsistent results across studies might be related to the use of different methodologies and outcome definitions, as well as to differences in the COVID-19 vaccination rollout. For example, while ChAdOx1 was extensively used across different age groups in the UK, it was restricted to individuals aged 60-69 years in Spain after March 2021. This might have prevented us from observing an association between ChAdOx1 and thromboembolic events, since prior studies suggest that this association is stronger among younger individuals.[194,250,254]

Table 1: Summary of real-world data studies assessing the associations between COVID-19 vaccines and thrombosis and/or thrombocytopenia syndromes

| | VTE | ATE | CVST or CVT | Thrombocytopenia | Immune thrombocytopenia | Thrombocytopenia & concomitant event | Bleeding |
|---|---|---|---|---|---|---|---|
| **ChAdOx1, first dose** | | | | | | | |
| Study V [260] | | | | | | | |
| Pottegård [194] | | | | | | | |
| Simpson [195] | | * | | *# | | | * |
| Hippisley-Cox [250] | | | | | | | |
| Kerr [251] | | | | | | | |
| Whiteley [252] | | | | | | | |
| Dag Berild[$] [255] | | | | | | | |
| Torabi [253] | | | | | | | |
| Burn [254] | | | | | | | |
| **ChAdOx1, second dose** | | | | | | | |
| Torabi [253] | | | | | | | |
| Burn [254] | | | | | X | | |
| **Ad26.COV2.S** | | | | | | | |
| Ashrani [261] | | | | | | | |
| **BNT162b2, first dose** | | | | | | | |
| Study V [260] | | | | | | + PE | |
| Simpson [195] | | | | # | | | |
| Hippisley-Cox [250] | | | | | | | |
| Barda [262] | | | | | | | |
| Kerr [251] | | | | | | | |
| Whiteley [252] | | | | | | | |
| Dag Berild[$] [255] | | | | | | | |
| Torabi [253] | | | | | | | |
| Burn [254] | | | | | | | |
| **BNT162b2, second dose** | | | | | | | |
| Study V [260] | | | | | | | |
| Torabi [253] | | | | | | | |
| Burn [254] | | | | | | + DVT | |
| **mRNA-1273** | | | | | | | |
| Dag Berild[$] [255] | | | | | | | |

**Notes:** Red boxes represent positive associations and green boxes null or negative associations. *These risks were identified using a case-control study design but were attenuated in a post-hoc self-controlled case series (SCCS) analysis. # Excluding Immune Thrombocytopenia. $ Dag Berild et al analyse together first and second doses of each vaccine. Abbreviations: VTE: venous thromboembolism, ATE: arterial thromboembolism; CSVT: cerebral venous sinus thrombosis; CVT: cerebral venous thrombosis.
**Source:** own elaboration with data from 11 real-world studies that use the following methodological approaches: historical comparator method,[194,254,260] case-control analysis,[195] SCCS analysis,[195,250,251,253,255] and contemporary cohort analysis.[252,262]

As for risks of thrombosis and thrombocytopenia following COVID-19 infection, consistent evidence shows that COVID-19 is associated with increased risks of VTE and ATE, as well as of thrombocytopenia, as seen in Table 2. Importantly, these risks are much higher following COVID-19 infection than following vaccination, thus suggesting that the benefits of COVID-19 vaccines far outweigh the risks of adverse events, at least regarding TTS.

Table 2: Summary of real-world data studies assessing the association between COVID-19 infection and thrombosis and/or thrombocytopenia syndromes

| | VTE | ATE | CVST or CVT | Thrombocytopenia | Immune thrombocytopenia | Thrombocytopenia & concomitant event | Bleeding |
|---|---|---|---|---|---|---|---|
| **Study V** [260] | | | | | | +VTE, ATE | |
| Hippisley-Cox [250] | | | | | | | |
| Barda [262] | | | | | | | |
| Torabi [253] | | | | | | +VTE | |
| Burn [254] | | | | | | | |

**Notes:** Red boxes represent positive associations and green boxes null or negative associations. Abbreviations: VTE: venous thromboembolism, ATE: arterial thromboembolism; CSVT: cerebral venous sinus thrombosis; CVT: cerebral venous thrombosis.
**Source:** own elaboration with data from 5 real-world studies that use the following methodological approaches: historical comparator method,[254,260] controlled case series analysis,[250,253] and contemporary cohort analysis.[252,262]

## 6.2. Methodological considerations

### 6.2.1. Study design and data sources

The studies of this Thesis were cohort studies underpinned by prospectively collected longitudinal records, which is a major strength. The use of RWD also allowed us to conduct our studies in a timely manner. **Studies I** and **II** reported more than 10,000 medical characteristics and included primary care, hospital, and claims records from different countries, which increased the generalisability and reliability of our findings while also enabling us to reflect the heterogeneity of various clinical settings. By

conducting a federated analysis, we were able to maintain patient confidentiality while also speeding up the results-obtention process. **Studies III, IV** and **V** were population-based studies underpinned by the SIDIAP database, a well-established database from Catalonia, Spain, which is representative of the region's population in terms of age, sex, and geographic distribution.[212] As a result, our findings are generalisable to the population living in Catalonia as well as to other countries/regions with similar sociodemographic characteristics. The use of Catalan data is also an asset since most studies in the field of COVID-19 have been conducted in the US and the UK, and evidence from our country remains scarce. Because we included a large number of participants in all the Studies (from 300,000 in **Study II** to 4.6 million in **Study III**), we were able to analyse rare outcomes such as thromboembolic events (**Study V**) as well as to explore the relation between exposures (e.g., cancer in **Study III**) and outcomes of interest (e.g., COVID-19 hospitalisation in **Study III**) in specific subgroup populations of interest (e.g., patients with haematological cancer in **Study III**). Unlike other RWD databases, SIDIAP includes information on socioeconomic deprivation, nationality, and some lifestyle factors (e.g., smoking), and therefore we were able to include these variables in our analyses when appropriate. The linkage of SIDIAP data to hospital records also allowed us to comprehensively capture events of interest that might not be adequately recorded in the primary care setting, such as thromboembolic events (**Study V**). Since all the studies had a relatively short follow-up time (from 21 days in **Study V** to 5 months in **Study IV**), our results are less prone to loss of follow-up bias. The use of a multi-state framework in **Study III** also allowed us to give a more thorough perspective of patients' interactions with the healthcare system while minimising the risks of collider bias. Finally, for the sake of reproducibility, we made our analytical code available in **Studies I, II, III,** and **V**.

However, our studies also have limitations. First, RWD is not collected for research purposes and data quality issues may arise.

Comorbidities and outcomes might be underestimated due to incomplete recording (see section 6.2.2. Assessment of variables),[263,264] although by including only individuals with at least one year of prior medical history available we attempted to comprehensively capture baseline conditions. In addition, SIDIAP includes several internal procedures to ensure data quality and numerous studies have been conducted to validate disease diagnoses, such as cancer, heart diseases, or dementia, among others.[265–267] Another limitation of RWD is that clinical standards and coding practices change over time and, therefore, findings must be interpreted with caution when comparing rates over different time periods, such as when comparing patients with COVID-19 to patients with influenza (**Study II**) or when comparing IR of events among exposed cohorts to background population rates (**Study V**). The comparison between COVID-19 and influenza patients in **Study II** is also limited by the fact that influenza vaccines and antiviral drugs might have reduced the risks of complications among influenza patients. Additionally, the overwhelming pressure on hospitals during the first wave of the pandemic might have contributed to poorer outcomes among COVID-19 patients.

In **Studies I** and **II**, some of the databases included were not representative of the general population (e.g., the Veterans Affairs database includes mostly veteran males aged ≥65 years). Additionally, some patients from the US might have been included in more than one database (e.g., in a hospital and a claims database). Unfortunately, we were unable to report the degree of overlap across data sources because patient-level data was not shared across data partners for confidentiality purposes. As for **Studies III, IV,** and **V** (SIDIAP-based), even though all the population registered at primary care centres from ICS is included in SIDIAP, diagnoses and procedures among persons that seek care at the private healthcare system are not captured. This might result in an underreporting of disease diagnoses, particularly acute conditions. However, chronic conditions are more likely to be well captured in SIDIAP for all the population

since general practitioners oversee long-term medical prescriptions. In **Study V,** we used a historical rate comparison study design to identify vaccine safety signals, a method sensitive but rather unspecific, with high type 1 error.[268] Lastly, our studies included data from a relatively short time (roughly March-June 2020 for **Studies I, II,** and **III**, and March 2020 to June 2021 for **Studies IV** and **V**). Therefore, our results related to COVID-19 outcomes are not generalisable to SARS-CoV-2 variants that emerged later, such as Delta or Omicron.

### 6.2.2. Assessment of variables

#### i. COVID-19 infection, hospitalisation, death, and vaccination

The use of different definitions to identify cases of COVID-19 infection is a strength of this Thesis. In **Studies I, II,** and **III,** we used COVID-19 diagnoses as well as test results to identify individuals infected with COVID-19 during the first wave of the pandemic. We used this approach to avoid selection bias due to testing restrictions or to hypothetical different testing patterns over time and across subgroup populations and contexts. However, due to the low positive predictive value of COVID-19 symptoms and signs, we might have incurred in misclassification bias and considered false positives as COVID-19 cases. In **Studies I, II,** and **III,** misclassification bias might have been differential, with a lower probability of false positives among exposed individuals since patients with obesity and with cancer were more likely to be infected and to become severely ill.[127,131,136] This would have biased our results away from the null in **Study III.** Since **Studies IV** and **V** were conducted when SARS-CoV-2 testing was widely available, we adapted our definition and considered as COVID-19 cases only those with a confirmed infection. However, our database included test results from the public sector, and thus we might have missed cases tested at private facilities, thus also incurring in misclassification bias. This

would have biased our results away from the null in **Study IV**, since individuals living in more affluent areas use more frequently private healthcare services. However, when the studies of this Thesis were conducted, COVID-19 cases were required to isolate following public health guidelines and in-person workers diagnosed with COVID-19 had to obtain a sickness certificate from primary care providers to be able to isolate at home. Thus, most COVID-19 cases diagnosed in the private sector were captured in the SIDIAP database through sick leaves records. Another limitation is that although public healthcare (including SARS-CoV-2 tests performed at healthcare centres) is free of charge in Catalonia, vulnerable subgroup populations might have been less likely to get tested due to access barriers, such as cultural or geographic barriers. This would have biased our results towards the null.[269] Lastly, we were unable to capture asymptomatic cases as well as cases that did not seek for care, and therefore our findings are only generalisable to symptomatic COVID-19 patients.

We used COVID-19 hospitalisations as an indicator of severe COVID-19 in **Studies I** to **IV**. However, because we lacked information on the cause of hospitalisation, COVID-19 hospitalisations were identified considering the number of days elapsed between a COVID-19 diagnosis and a hospital admission. Therefore, we might have classified as severe COVID-19 cases patients that had COVID-19 but that were hospitalised for other reasons. This misclassification bias might have been differential, particularly in **Study III** since patients with cancer are more likely to be hospitalised than patients without cancer due to cancer-related treatments and complications. Likewise, individuals living in more deprived areas were more likely to be hospitalised for other reasons since they exhibit a higher prevalence of chronic diseases.

In **Studies IV** and **V,** we identified individuals vaccinated against COVID-19. Our data included a complete record of vaccines administered in Catalonia. Therefore, the risk of misclassification

bias is particularly low for this variable, since only vaccines administered in other countries would not have been captured.

### ii. Additional exposures of interest

In **Study I**, we compared patients with COVID-19 living with obesity to those living without obesity. We used an obesity definition that combined an ever-record of a diagnosis of obesity with anthropomorphic measurements (BMI and weight). Despite of this, we might have incurred in misclassification bias due to underreporting of obesity, as well as changes of BMI/body weight over time or missing BMI and body weight measurements (e.g., in SIDIAP only 70% of the population has at least one measurement of BMI).[212] However, the prevalence of obesity in SIDIAP is in line with representative studies from Spain and individuals with a recorded BMI measurement have similar characteristics to the general SIDIAP population; suggesting that the risk of misclassification bias for BMI is low, at least in SIDIAP.[270,271] Furthermore, prior research indicates that adult BMI trends are relatively stable over time, with a tendency to rise with age.[272,273] In consequence, most adults with an ever-record of obesity are likely to remain obese over time. The use of BMI as an indicator of obesity has also some drawbacks, such as its inability to discern between body fastness and lean body mass, and inconsistent accuracy in measuring body fatness by sex, age, and ethnicity.[274,275] Lastly, we were unable to report our results according to different levels of obesity, or overweight, because we lacked data on BMI ranges or BMI as a continuous variable.

In **Studies II** (multi-database study) and **III** (SIDIAP-based study), cancer history was the main exposure of interest, which we identified using records of cancer diagnoses. In **Study II**, we used SNOMED codes to describe the frequency of cancer subtypes. We decided to describe ranks of cancer types, rather than to compare prevalence numbers across databases, because we acknowledge that the sensitivity of the codes used might differ by database. Four out of the eight databases included had been previously used in the oncology field: SIDIAP, CUIMC, Optum,

and the Veterans Affairs database.[202,276,277] Additionally, cancer diagnoses in SIDIAP have been validated against two Catalan cancer registries.[265] Sensitivities for the most common solid cancer types (prostate, breast, colorectal, lung, and bladder, the solid cancer types of interest in **Study III)** were high (>75%); however PPVs were lower, ranging from 47% (bladder) to 76% (lung). Sensitivities for haematological cancers were overall slightly lower (aside from multiple myeloma, with a sensitivity of 80%) and were of 64%, 67%, and 68% for Hodgkin and non-Hodgkin lymphoma and leukaemia, respectively. This is likely due to the nature of the SIDIAP database, which is based on primary care records. Thus, cancer types that are common and mostly managed at the primary care level are more likely to be recorded than more rare cancers that are mostly managed at the hospital level. Therefore, although SIDIAP captures adequately cancer diagnoses, we might have incurred in misclassification bias, especially for rare cancer types. Conversely, hospital databases from **Study II** might capture better these cancers. For example, in CUIMC (a hospital-database from the US included in **Study II**), SNOMED cancer codes showed high sensitivity (99%) and specificity (99.9%).[277]

In **Study III**, we also lacked information on cancer stage or cancer treatments, although we used years since cancer diagnosis as a proxy of active cancer treatment (since individuals recently diagnosed are more likely to be under treatment than those diagnosed years ago). Due to small sample sizes, we were unable to explore the association between less frequent cancer types and COVID-19 outcomes, as well as to provide results stratified by sex for the solid cancer types that we analysed. Lastly, we handled missing data on smoking and socioeconomic status as an additional category, an approach that can lead to biased results.[278] However, in sensitivity analyses using multiple imputation of missing data we found consistent results with our main results.

In **Study IV**, socioeconomic deprivation was the main exposure of interest. We used an area-based index of deprivation as a proxy of individual SES, an approach that is limited due to the risks of ecological fallacy. Thus, our results must be interpreted with caution considering that the inequalities observed could be due to both individual and area differences (e.g., environmental conditions, healthy food availability, access to healthcare services). However, there is evidence showing that individual-level and area-level indicators of SES are moderately to highlight correlated.[279] In addition, our deprivation index was based on census-tract data from 2001, and therefore might be outdated. However, prior literature suggests that deprivation levels do not fluctuate significantly between censuses.[280,281] We also restricted our analyses to individuals living in urban areas (roughly 88% of the SIDIAP population), and thus our results are not generalisable to the rural population. The lack of information on occupation also prevented us from investigating whether the differences found are related to different working conditions.

### iii. Additional outcomes of interest

We reported COVID-19-related deaths in **Studies III** and **IV** which we defined as deaths of any-cause following a COVID-19 event (**Study III**) or withing 28 days after a COVID-19 diagnosis (**Study IV**). However, some of these deaths might not be due to COVID-19 infection, although our results in **Study III** suggest that we mostly captured deaths caused (at least partially) by the infection. For instance, death rates were remarkably higher among patients with cancer with COVID-19 when compared to patients with cancer without COVID-19.

In **Study V**, we described the incidence rates of thrombosis, thrombocytopenia, and TTS following COVID-19 vaccination and infection. To identify these outcomes, we used algorithms developed in collaboration with the EMA. However, thromboembolic events have not been validated previously in the SIDIAP database and measurement error might be an issue.[220] Since concerns regarding vaccine's safety emerged in

early March 2021, we cannot exclude that measurement error might have been differential between vaccinated (exposed) cohorts and the background comparator cohort. For example, a study from Norway reported that the use of health care services among healthcare workers vaccinated against COVID-19 increased after early March 2021.[282] However, we found consistent results when stratifying by time period in sensitivity analyses. Lastly, we assessed TTS instead of VITT because: i) VITT was described for the first-time following adenovirus-based COVID-19 vaccination and we were interested in comparing rates of thromboembolic events following vaccination to historical background rates, and ii) identifying VITT in real-world data is challenging since antibodies against PF4 are not routinely measured.[163] However, to date there is still not a clear consensus on how to identify TTS using RWD.[163] Lastly, the identification of TTS in RWD is likely subject to high amounts of false positives, since TTS might reflect only a coincidence between two common clinicals events (simultaneous thrombosis and thrombocytopenia) that can be attributed to underlying comorbidities (e.g., cancer) rather than vaccine-induced TTS.[163,164]

### 6.2.3. Confounding

In **Studies III, IV,** and **V**, we analysed the associations between different exposures and outcomes of interest. In **Studies III** and **IV,** we used a DAG to guide our modelling strategy, whereas in **Study V,** we used SIRs to compare exposed to background (non-exposed) cohorts. Despite our methods, confounding might have biased our results. For example, in **Study III,** although we adjusted our models by smoking categories (never-smoker, smoker, former smoker), our results might have been biased due to residual confounding by smoking amount. However, sensitivity analyses restricted to never-smokers results were comparable to the entire study population. Due to small sample sizes, we were unable to restrict our analysis to never smokers for solid cancer types, and thus our findings for bladder and lung cancer must be

interpreted with caution. Similarly, additional lifestyles variables not included in SIDIAP, such as physical activity, might be associated with both our exposure (cancer) and COVID-19 outcomes. However, the fact that models performed using different adjustment strategies showed similar results is reassuring, and the inclusion of other covariates would probably not have had a major impact in our results.

In **Study V**, we used the historical comparator method to compare standardised incidence rates of thromboembolic events after COVID-19 vaccination and infection. Although we standardised rates by age, we cannot exclude that confounding by indication might have biased our results.[283] Confounding by indication occurs when an indication (e.g., the indication to receive a treatment or vaccine) is associated with both the exposure (e.g., being vaccinated) and the outcome (e.g., having a thromboembolic event). In the context of COVID-19 vaccines, vaccination was particularly recommended among individuals with underlying conditions that have been previously associated with severe COVID-19. In addition, several of these conditions (e.g., cancer, obesity, cardiovascular diseases) are also associated with increased risks of thromboembolic events. Since vaccinees had a slightly worse baseline health status than the background population, confounding by indication might explain the safety signals observed for pulmonary embolism among BNT162b2 recipients, as well as for thrombocytopenia for BNT162b2 and ChAdOx1 recipients.

## 6.3. Recommendations for future research

### 6.3.1. Epidemiology of COVID-19

#### i. Characteristics of COVID-19 cases

**Studies I** and **II** described in-depth baseline characteristics and 30-days outcomes among individuals with COVID-19 and obesity or cancer. However, these studies were conducted prior to the advent of SARS-CoV-2 VOC and the COVID-19 vaccine rollout. Hence, the results from these studies are not generalisable to the current situation as of early 2023. Nowadays, the predominant SARS-CoV-2 variant is Omicron and 69% of the global population has received at least one dose of a COVID-19 vaccine as of early January 2023.[284] To understand the current impact of the COVID-19 pandemic, further studies are needed on the characteristics of patients with COVID-19, overall and by vaccination or reinfection status, as well as by underlying conditions of interests.

Another topic that is gaining increased importance are long-term complications of COVID-19, particularly long COVID.[75] According to reports from the UK, 3.4% of the population self-reported long Covid as of December 2022,[285] and concerns have been raised regarding the impact of this condition on patients as well as on health systems and on labour productivity.[75,286] To date, most studies on this topic include relatively small sample sizes, are based on surveys/questionnaires and/or include only hospitalised COVID-19 cases, and often lack a comparator for benchmark.[287–291] In addition, studies characterising long COVID among specific subgroup populations, such as individuals with obesity or cancer, are still lacking.

A few studies have compared long-lasting symptoms/manifestations among COVID-19 cases to influenza cases, suggesting that the risks of long-lasting complications are higher after COVID-19 infection.[77,292] These studies were conducted before the advent Omicron and did not explore whether the characteristics of individuals presenting with long-

lasting complications differed by vaccination status. Interestingly, a study from the UK conducted with self-reported data found that the risks of developing long Covid were lower during the Omicron predominance than during the Delta predominance.[293] There is also some evidence suggesting that the prevalence of long-lasting symptoms might be lower among individuals previously vaccinated against COVID-19 when compared to those non-vaccinated.[294–296] Additional research is needed on the characteristics of individuals with long COVID, overall and by SARS-CoV-2 variant and by vaccination or reinfection status, as well as by baseline conditions, to comprehend the long-term effects of COVID-19 infection and to inform healthcare decision-making.

### ii. Factors associated with COVID-19 infection and severity

In **Study III**, we investigated the associations between cancer and COVID-19 infection and severity. However, we only included data from the first wave of the pandemic (March-May 2020). Therefore, as discussed previously, results might not be generalisable to other waves of the pandemic in Spain, particularly following the emergence of SARS-CoV-2 VOCs such as Alpha, Delta, or Omicron. More importantly, **Study III** was conducted prior to the COVID-19 vaccine rollout. Although cancer was an exclusion criterion in RCTs assessing COVID-19 VE and safety,[297] patients with cancer were generally prioritised for COVID-19 vaccination since they were considered a high-risk population.[298] As seen in **Study IV**, as of June 2021, 92% of individuals aged ≥40 years with a history of cancer had received at least one dose of a COVID-19 vaccine in Catalonia. In consequence, studies are needed addressing the relation between cancer and the risks of breakthrough COVID-19 infection and severity (breakthrough infections are defined as infections following vaccination against a particular disease). Similarly, as reinfections are becoming increasingly common, studies addressing this relation among individuals with a prior infection history are also needed. To date, evidence on these matters remains scarce. A study from the US

reported that patients with cancer had higher risks of breakthrough infections than patients without cancer, these risks were higher among those with haematological cancers and those treated with immunosuppressive therapies or bone marrow transplantation.[299]

Increased risks of breakthrough infections might be related to a lower VE among the cancer population. For instance, a study from the US reported that VE against COVID-19 was 58% [95% CI:39; 72] following two doses vaccination[300] whereas VE among the general population was estimated around 90%.[149-153] Interestingly, VE was lower among those with haematological cancers, those treated with systemic anticancer therapy three months prior to vaccination, and those treated with chemotherapy.[300] However, this study had important limitations. First, the generalisability of the study was limited, since 95% of the study population were males and 75% were aged ≥70 years. Second, vaccine protection was ascertained only 3 months after vaccination, and some studies suggest that vaccine immunity wanes faster among patients with cancer.[301] Third, the study was also conducted prior to the advent of Omicron, which is associated with reduced VE.[155,156] Lastly, the study did not capture booster doses. Interestingly, a study from the UK showed that vaccine boosters increase VE against COVID-19 infection, hospitalisation, and death among patients with and without cancer, although booster VE was lower among cancer patients recently diagnosed or treated with anticancer therapies, as well as those with haematological cancers.[302] Despite the effect of booster doses, patients with cancer remained a high-risk population for severe COVID-19 when compared to the general population, with OR of hospitalisation of 3.38 [95% CI: 3.03; 3.77] and of death of 3.01 [2.48; 3.65]. However, this study was also conducted prior to the advent of Omicron and different behaviours between cancer patients and controls might have biased study results.[302] In summary, only a few studies from the UK and the US have provided evidence on COVID-19 VE among patients with cancer. More research is needed on this matter,

especially regarding the duration of protection and the role of booster doses in the context of different predominant SARS-CoV-2 variants, overall and by cancer subtypes.

In **Study IV**, we found that socioeconomic inequalities in COVID-19 infection and hospitalisation in Catalonia decreased among individuals aged ≥40 years after the start of the COVID-19 vaccine rollout, providing novel evidence suggesting that mass COVID-19 vaccination can contribute to the reduction of COVID-19 inequalities. However, our study was not generalisable to younger nor rural populations and included only data up to 6 months after the start of the vaccine rollout in Catalonia. We also used an ecological variable as a proxy of SES and did not explore other axes of inequalities, such as ethnicity. Thus, more research is needed to corroborate our findings in other countries and to overcome our limitations.

### iii. Vaccine uptake

In **Study IV**, we also found inequalities in COVID-19 vaccination uptake among individuals aged 40-64 years living in Catalonia. This is consistent with evidence from the US and the UK,[191,192,242–248] but evidence from other countries is still lacking. Studies addressing this issue among children and adolescents are also lacking, although emerging evidence from the US show disparities in COVID-19 vaccine uptake in these age groups.[303,304] Concerningly, in Spain, vaccine uptake among children remains low.[305] As of November 2022, only 56% of children aged 5-12 years had received at least one dose of a COVID-19 vaccine (41% in Catalonia). Investigating potential disparities in COVID-19 vaccine uptake withing this age group could help understand factors that influence vaccination and guide future immunization efforts.

Our study was also conducted in early stages of the vaccination campaign and was limited to first-dose coverage. Nowadays, COVID-19 boosters are recommended for the elderly, high-risk populations, and healthcare workers. Yet, booster coverage is far

lower than primary vaccination coverage. As of November 2022, only 59% and 17% of individuals aged >80 years and 60-69 years, respectively, had received a booster in Catalonia.[306] Other countries have also reported low booster coverages. [307] However, studies addressing socioeconomic inequalities in booster coverage are still lacking, with few exceptions. A study from the US published in January 2023 reported important ethnic inequalities in COVID-19 booster coverage, as well as lower coverage among those with lower education or living in poverty.[307] Since booster doses are expected to play a key role in the COVID-19 response in the future years, studies addressing the influence of socioeconomic factors (including, but not limited to, SES) on booster vaccine uptake are also needed to inform vaccination strategies.

### iv. Vaccine safety

Although a substantial number of RWD studies have attempted to analyse the risks of thromboembolic events following COVID-19 vaccination, these have found inconsistent results.[194,195,250–255,262] However, these studies were single-database studies conducted in different settings and used different methods and outcomes definitions. Thus, comparing results across these studies has important limitations. Future studies addressing this issue should consider the inclusion of multiple databases to provide comparable results across settings and to increase generalisability. For example, the EMA and the European Medicines Regulatory Network Data Analysis recently launched an initiative to support regulatory decision-making using RWD, the Data Analysis and Real-World Interrogation Network (DARWIN EU®).[308] DARWIN-EU's goal is to conduct studies that provide reliable and timely evidence on the use, safety, and effectiveness of medicines using a federated network of real-world healthcare database from Europe. Currently, this federated network includes eight databases from the primary and hospital care setting and from six countries, namely Finland, France, Estonia, Spain (2 databases, one of which is SIDIAP), United

Kingdom, and the Netherlands (2 databases).[309] The participating databases have all ben standardised to the OMOP CDM and analyses will be conducted through a federated network approach. The DARWIN-EU initiative is a 5-year project which is currently as its very early stages, the first study results are expected to be published in 2023.

Studies assessing the risks of thromboembolic events following COVID19 vaccination focused mostly on primary and homologous vaccination schemes and included mostly ChAdOx1 and BNT162b2 vaccinees. However, COVID-19 vaccination schemes are becoming increasingly diverse and might include repeated booster doses and heterologous vaccination schemes (i.e., when an individual receives different COVID-19 vaccines, either within the primary vaccination scheme or booster vaccines different from the primary scheme). Although there is evidence suggesting that boosters and heterologous vaccination is safe, studies on this matter are still limited. Since booster doses are expected to continue to play a key role in the COVID-19 response in the foreseeable future, studies assessing the risks of adverse events associated with heterologous vaccination and booster vaccination are required. Future studies should also analyse if risks of adverse events are higher among specific population groups.

### 6.3.2. Using real-world data for epidemiological research

Throughout the pandemic, RWD has been key to provide reliable and timely evidence in the field of COVID-19.[310] As shown by its increased use in epidemiological studies[310] and by the recognition of their importance by regulatory agencies such as the EMA (e.g., DARWIN-EU)[308] and the US Food and Drug Administration,[311] it is expected that the use of RWD will continue to increase in the future years. Although real-world data (RWD) has many strengths, such as its generalizability, large sample size, and accessibility, the quality of the data is often a major concern

due to inaccurate or incomplete recording, as well as the lack of information on relevant health variables like lifestyle and socioeconomic factors. Thus, it is crucial to develop and implement mechanisms to enhance the quality of RWD, not just for research purposes but also to better inform public health and clinical decision-making.

For example, there is some evidence suggesting that providing feedback to clinicians, financial incentives, using customised templates to collect information on specific conditions, and developing specific protocols and training programmes, could improve the data recording process. [260,261] These mechanisms are non-mutually exclusive and could be implemented in combination. Further strategies could also be explored and developed in close collaboration with healthcare professionals through participatory processes. The analysis of free text could also significantly enhance the completeness of RWD, since a lot of information about the state of patients is recorded as free text.[312] With the expanding field of natural language processing (NLP , which is a branch of artificial intelligence), automated solutions can be developed to extract information from free text.[312] In the field of health, there are examples of NLP systems that capture symptoms and ascertain smoking history from EHR, [313,314] extract information from radiological reports,[315] and identify cancer stage from pathology reports.[316] However, the currently available NLP systems focus on a highly specific area (e.g., extract symptoms) and are considerably context-dependent.[312] The development of a more comprehensive NLP system that can extract diverse information (e.g., extract symptoms and cancer staging through a unique algorithm) across various settings and institutions remains a challenge.[312]

Another way of improving the completeness of RWD is through linkage of multiple data sources,[317] as seen in **Study V** where we used SIDIAP data linked to hospital records to better capture our outcomes. However, as discussed in the section 6.2. Methodological considerations, we lacked important information

that could be obtained through linkage with other data sources. For instance, SIDIAP could be linked to the Spanish registry of deceases from the National Institute for Statistics (in Spanish, *Instituto Nacional de Estadística*) to capture causes of mortality, to taxes data to capture socioeconomic variables, such as income or working status, as well as to other health data sources. Undoubtedly, datasets that depict the whole spectrum of patient care and the conditions in which individuals live would provide better insights into the population's health. However, data linkage also poses several challenges, such as safeguarding citizens' privacy, evaluating the quality of the linkage, and addressing data fragmentation issues within and between organizations.[317] The use of an unique personal identifier could potentially make data linkage easier.[318] For example, in Denmark each resident has a long-life unique personal identifier that can be used to link administrative information (e.g., the Danish Civil Registration System, including information on migration or on civil status; or the Danish Registry of causes of death) as well as multiple health data sources (e.g., cancer registry, birth registry, prescriptions registry, hospital discharge database, biobanks). [318]

Lastly, validation studies assessing data quality are also needed. For example, our outcomes of interest in **Study V** were thromboembolic events, thrombocytopenia, and TTS. However, there is evidence showing that the identification of these events in RWD is challenging and might be particularly prone to false positives.[163] Consequently, a study aiming to validate thrombosis diagnoses and thrombocytopenia in SIDIAP would be of extreme value for subsequent research. More broadly, validation studies are an important asset to evaluate the fitness of purpose of RWD, particularly when including exposures or outcomes not previously investigated with a specific database.

## 6.4. Implications for public health

Overall, this Thesis emphasises the prominent role that RWD can play to inform public health policies, particularly in the context of an emergent disease when timely evidence is crucial. Large observational studies such as **Studies I, II,** and **III** have been fundamental to identify vulnerable subgroup populations and subsequently tailor prevention strategies to protect them. For example, in early stages of the pandemic the UK recommended high-risk individuals, such as patients with cancer, to shield (i.e., minimising face-to-face interactions and isolating at home as much as possible).[319] Nationwide vaccination campaigns also prioritised high-risk individuals. In Spain, individuals with a recent haematological cancer or with an uncontrolled cancer as well as individuals with a solid metastatic cancer or undergoing chemotherapy were prioritised for vaccination.[298] RWD studies, such as **Studies IV** and **V,** have also provided evidence on COVID-19 vaccine's uptake and safety in the real-world setting and informed vaccination campaigns.

**Studies II** and **V** also highlight the importance of generating reliable evidence to counter misconceptions and misinformation, which can cause significant damage on the population's health.[320,321] At the onset of the pandemic, the threat posed by the emergence of SARS-CoV-2 was to some extent underestimated, since COVID-19 was often compared to influenza due to their similarities in terms of symptoms and modes of transmission.[320,322] This misjudgement might have delayed the COVID-19 response and influenced individual's self-protection behaviours.[320] By showing that COVID-19 is more severe than influenza, **Study II** contributed to shift the perception of COVID-19 among policy-makers and the public. As for COVID-19 vaccines, a significant number of fake news about vaccine's side effects have circulated in online social media, with harmful consequences.[321,323] Indeed, being exposed to COVID-19 vaccine fake news (and believing these to be true) has been associated with increased vaccine hesitancy. Thus, evidence-based studies

on the safety of COVID-19 vaccine's safety, such as **Study V**, can contribute to reduce vaccine's hesitancy provided that the results are widely disseminated.

**Studies I, II,** and **III** also demonstrate the intersection between NCDs and infectious diseases such as COVID-19.[324,325] Indeed, we observed that individuals with underlying NCDs (obesity and cancer) were more vulnerable to COVID-19. This highlights the need to tackle NCDs to reduce the burden of the COVID-19 pandemic and increase resilience against future epidemics. Preventing NCDs is also crucial to improve the population's health since NCDs are the leading cause of mortality worldwide, accounting for 74% of the total number of deaths.[326] The WHO estimates that 80% of cases of diabetes, heart disease, and stroke, and 40% of cancers could be prevented by addressing preventable risk factors, such as alcohol, drug, and tobacco use, dietary habits, physical inactivity, and exposure to environmental hazards (e.g., air pollution, radiation).[326,327] Globally, these factors are responsible for over 30 million of deaths every year, according to estimates from the Global Burden of Disease study 2019.[328]

Addressing these risk factors requires actions at different levels since their underlying causes are multifactorial and multifaceted. For example, to prevent the harmful effects of tobacco (the second leading risk factor of preventable death worldwide),[328,329] the WHO recommends interventions both at the policy level (e.g., increase tobacco taxes, ban smoking in public places) and the individual level (e.g., counselling on cessation of tobacco, pharmacological treatments for smoking cessation).[330] Unfortunately, the COVID-19 pandemic has also had deleterious consequences on the prevention and management of NCDs due to the disruption of health services (e.g., delays in surgical procedures, cessation of screening programmes, interruption of long-term treatments).[325]

More broadly, our findings in **Study IV** also underscore the importance of analysing socioeconomic inequalities in health

outcomes as well as of designing and implementing health policies taking into consideration the influence of the social determinants of health (SDOH, i.e., the conditions in which people are born, grow, work, live, and age, and people's access to power, money and resources, as defined by the WHO).[331] The links between the SDOH and health status had been well-established prior to the pandemic, with some studies suggesting that 40% of health outcomes (both NCDs and infectious diseases) might be influenced by the SDOH.[332]

Preventable risk factors of NCDs are also heavily influenced by the SDOH. For example, individuals with low SES are more likely to smoke, have an unhealthy diet and insufficient physical activity, and to be exposed to air pollution.[333] Unsurprisingly, the SDOH have also led to unfair and avoidable differences in health across population groups throughout the pandemic.[127,331] As shown in **Study IV** individuals living in more deprived areas had higher incidence rates of COVID-19 infections and hospitalisations and lower COVID-19 vaccine uptake.

In addition to the importance of targeted interventions to promote vaccination among vulnerable populations, our results emphasise that strategies addressing the structural determinants of socioeconomic inequalities are urgently needed. These include poverty, unemployment, lack of education, food insecurity, poor working and housing conditions, discrimination (e.g., ageism, racism, sexism), violence, and barriers to healthcare access, among many other drivers of inequalities. However, the indirect effects of the COVID-19 pandemic (e.g., job losses, school closures) have disproportionately affected those more vulnerable, exacerbating existing inequalities. Ultimately, in the aftermath of the COVID-19 pandemic, building back fairer health systems will require a strong political commitment with health equity and the development of intersectoral policies.[127]

# 7. CONCLUSIONS

1. People with underlying conditions such as obesity or cancer (particularly those with recent cancer diagnosis and with haematological cancers) as well as people with low socioeconomic status have higher risks of being infected with SARS-CoV-2 and developing severe COVID-19.

   *Les persones amb malalties cròniques, com l'obesitat o el càncer (especialment aquelles amb un diagnòstic recent de càncer o que tenen un càncer hematològic) i les persones amb un baix nivell socioeconòmic tenen més riscos d'infectar-se pel SARS-CoV-2 i de desenvolupar complicacions.*

2. Despite socioeconomic inequalities in COVID-19 vaccine uptake, six months after the start of the COVID-19 vaccine rollout in Catalonia, Spain, inequalities in COVID-19 infection and hospitalisation decreased among individuals aged 40 years or older living in urban areas. Yet, socioeconomic inequalities persisted and were more pronounced for hospitalisations.

   *Malgrat desigualtats socioeconòmiques en la cobertura de la vacuna contra la COVID-19, sis mesos després de l'inici de la campanya de vacunació a Catalunya, van disminuir les desigualtats en les infeccions i hospitalitzacions per COVID-19 en persones de 40 anys o més residents en zones urbanes. Tot i així, les desigualtats van persistir, especialment pel que fa a les hospitalitzacions.*

3. Notwithstanding previous concerns over COVID-19 vaccines safety, we found that thromboembolic events following COVID-19 infection far outweigh the risks of thromboembolic events following vaccination with BNT162b2 and ChAdOx1.

> *Tot i certes preocupacions sobre la seguretat de les vacunes contra la COVID-19, vam observar que els riscos d'esdeveniments tromboembòlics després de la infecció per COVID-19 eren molt superiors als riscos de patir aquests esdeveniments després de la vacunació amb BNT162b2 i ChAdOx1.*

4. Three years after the emergence of SARS-CoV-2, there are still important research gaps related to COVID-19, including on the long-term consequences of the infection, the risks of reinfections or breakthrough infections among high-risk populations, the associations between socioeconomic factors and COVID-19 vaccination uptake, and the safety of booster vaccine doses.

> *Tres anys després de l'aparició del SARS-CoV-2, encara hi ha aspectes relacionats amb la COVID-19 poc coneguts, incloent-hi les conseqüències a llarg termini de la infecció, els riscos de reinfecció o d'infecció postvacunació en poblacions de risc, l'associació entre factors socioeconòmics i la cobertura vacunal, i la seguretat de les dosis de record de les vacunes contra la COVID-19.*

5. RWD have been key to provide reliable and timely evidence throughout the COVID-19 pandemic and their use in the field of epidemiology is expected to increase. However, RWD quality remains a challenge that must be addressed.

    *Les dades del món real han estat clau per generar evidència de manera ràpida i fiable durant la pandèmia de la COVID-19, i es preveu que augmenti el seu ús en el camp de l'epidemiologia en els pròxims anys. No obstant això, la qualitat de les dades del món real continua sent un repte que cal dirimir.*

6. The findings of this Thesis reinforce the importance of tackling non-communicable diseases, preventable risk factors, and, more broadly, socioeconomic inequalities to reduce the burden of the COVID-19 pandemic and improve the population's health.

    *Els resultats d'aquesta Tesi reforcen la importància d'abordar l'impacte de les malalties cròniques i els factors de risc evitables així com, de manera més general, l'impacte de les desigualtats socioeconòmiques per reduir la càrrega de la pandèmia de la COVID-19 i millorar la salut de la població.*

# 8. REFERENCES

1. Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451(7181):990-993. doi:10.1038/nature06536

2. Morens DM, Fauci AS. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell*. 2020;182(5):1077-1092. doi:10.1016/j.cell.2020.08.021

3. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*. 2020;382(8):727-733. doi:10.1056/nejmoa2001017

4. Zhou P, Yang X Lou, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7

5. World Health Organization. Listings of WHO's response to COVID-19. World Health Organization. Published 2020. Accessed April 7, 2022. https://www.who.int/news/item/29-06-2020-covidtimeline

6. Porta MS. *A Dictionary of Epidemiology*. Oxford University Press; 2014. doi:10.1093/acref/9780199976720.001.0001

7. World Health Organization. *COVID-19 Weekly Epidemiological Update - Edition 125.*; 2023. Accessed February 14, 2023. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-january-2023

8. European Centre for Disease Prevention and Control. Rapid risk assessment: Cluster of pneumonia cases caused by novel coronavirus, Wuhan, China - January 2020. Published online 2020. Accessed April 1, 2022. https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-cluster-pneumonia-cases-caused-novel-coronavirus-wuhan

9. Gorbalenya AE, Baker SC, Baric RS, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology 2020 5:4*. 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z

10. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. Accessed April 12, 2022. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

11. World Health Organization. WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV). Published June 30, 2020. Accessed April 7, 2022. https://www.who.int/director-

general/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov)

12. Spiteri G, Fielding J, Diercke M, et al. First Cases of Coronavirus Disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. *The COVID-19 Reader*. 2020;2019(February):95-104. doi:10.4324/9781003141402-11

13. Nadeau SA, Vaughan TG, Scire J, Huisman JS, Stadler T. The origin and early spread of SARS-CoV-2 in Europe. *Proceedings of the National Academy of Sciences*. 2021;118(9). doi:10.1073/pnas.2012008118

14. Linde P. Sanidad confirma en La Gomera el primer caso de coronavirus en España. El País. Published January 31, 2020. Accessed April 14, 2022. https://elpais.com/sociedad/2020/01/31/actualidad/1580509404_469734.html

15. Henríquez J, Gonzalo-Almorox E, García-Goñi M, Paolucci F. The first months of the COVID-19 pandemic in Spain. *Health Policy Technol*. 2020;9(4):560-574. doi:10.1016/j.hlpt.2020.08.013

16. Centro Nacional de Epidemiología. Panel COVID-19. Accessed January 17, 2023. https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos

17. Equipo COVID-19. RENAVE. Centro Nacional de Epidemiología. *Situación de COVID-19 En España a 24 de Febrero de 2023.*; 2023. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Nueva_estrategia_vigilancia_y_control.pdf

18. Zhang SX, Arroyo Marioli F, Gao R, Wang S. A Second Wave? What Do People Mean by COVID Waves? – A Working Definition of Epidemic Waves. *Risk Manag Healthc Policy*. 2021;Volume 14:3775-3782. doi:10.2147/RMHP.S326051

19. Zhang SX, Arroyo Marioli F, Gao R, Wang S. A second wave? What do people mean by COVID waves? – A working definition of epidemic waves. *Risk Manag Healthc Policy*. 2021;14:3775-3782. doi:10.2147/RMHP.S326051

20. Coma E, Guiriguet C, Mora N, et al. Impact of the COVID-19 pandemic and related control measures on cancer diagnosis in Catalonia: a time-series analysis of primary care electronic health records covering about five million people. *BMJ Open*. 2021;11(5):e047567. doi:10.1136/BMJOPEN-2020-047567

21. Coma E, Mora N, Méndez L, et al. Primary care in the time of COVID-19: monitoring the effect of the pandemic and the lockdown measures on 34 quality of care indicators calculated for 288 primary care practices covering about 6 million people in Catalonia. *BMC Fam Pract*. 2020;21(1):208. doi:10.1186/s12875-020-01278-8

22. Pérez de la Sota E, Piñón M, Quintana E, Mestres CA. COVID 19—A Spanish perspective. *J Card Surg*. 2021;36(5):1624-1631. doi:10.1111/JOCS.14994

23. Our World in Data. Coronavirus (COVID-19) Cases. Accessed January 17, 2023. https://ourworldindata.org/covid-cases

24. Villanueva-Cañas JL, Gonzalez-Roca E, Unanue AG, et al. Implementation of an open-source robotic platform for SARS-CoV-2 testing by real-time RT-PCR. *PLoS One*. 2021;16(7 July). doi:10.1371/journal.pone.0252509

25. Panteli D, Reichebner C, Rombey T, et al. Health care patterns and policies in 18 European countries during the first wave of the COVID-19 pandemic: an observational study. *Eur J Public Health*. 2022;32(4):557-564. doi:10.1093/eurpub/ckac059

26. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *www.thelancet.com*. 2020;396:535. doi:10.1016/S0140-6736(20)31483-5

27. Centro de Coordinación de Alertas y Emergencias Sanitarias. *Actualización Nº 102. Enfermedad Por El Coronavirus (COVID-19). 11.05.2020.*; 2020. Accessed January 2, 2023. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_102_COVID-19.pdf

28. Ward H, Atchison C, Whitaker M, et al. SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat Commun*. 2021;12(1):1-8. doi:10.1038/s41467-021-21237-w

29. Havers FP, Reed C, Lim T, et al. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern Med*. 2020;180(12):1776-1786. doi:10.1001/jamainternmed.2020.4130

30. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet*. 2020;396(10247):313-319. doi:10.1016/S0140-6736(20)31304-0

31. Glassman A, Kenny C, Yang G. *COVID-19 Vaccine Development and Rollout in Historical Perspective*. Report no 607. Center for Global Development; 2022. Accessed January 13, 2023. https://www.cgdev.org/publication/covid-19-vaccine-development-and-rollout-historical-perspective

32. European Medicines Agency. EMA recommends Nuvaxovid for authorisation in the EU. Published 2021. Accessed May 19, 2022. https://www.ema.europa.eu/en/news/ema-recommends-first-covid-19-vaccine-authorisation-eu

33. European Medicines Agency. Spikevax (previously COVID-19 Vaccine Moderna). Accessed May 19, 2022. https://www.ema.europa.eu/en/medicines/human/EPAR/spikevax

34. European Medicines Agency. Vaxzevria (previously COVID-19 Vaccine AstraZeneca). Accessed May 19, 2022. https://www.ema.europa.eu/en/medicines/human/EPAR/vaxzevria-previously-covid-19-vaccine-astrazeneca#authorisation-details-section

35. European Medicines Agency. Jcovden (previously COVID-19 Vaccine Janssen). Accessed May 19, 2022. https://www.ema.europa.eu/en/medicines/human/EPAR/jcovden-previously-covid-19-vaccine-janssen

36. European Medicines Agency. Nuvaxovid. Accessed May 19, 2022. https://www.ema.europa.eu/en/medicines/human/EPAR/nuvaxovid

37. Valdés I, Mouzo J. España empieza a vacunar bajo la amenaza de una tercera ola. *El País*. Published online December 28, 2020. https://elpais.com/sociedad/2020-12-27/espana-empieza-a-vacunar-bajo-la-amenaza-de-una-tercera-ola.html

38. Consejo Interterritorial del Sistema Nacional de Salud. *Estrategia de Vacunación Frente a COVID-19 En España. Actualización 6*. Vol 1.; 2021. Accessed February 8, 2022. https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Actualizaciones_Estrategia_Vacunacion/docs/COVID-19_Actualizacion6_EstrategiaVacunacion.pdf

39. Ministerio de Sanidad. *GIV COVID-19 Gestión Integral de La Vacunación COVID-19. Informe de Actividad*. Accessed January 13, 2023. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Informe_GIV_comunicacion_20211230.pdf

40. Our World in Data. COVID-19 Data Explorer. Accessed January 13, 2023. https://ourworldindata.org/explorers/coronavirus-data-

explorer?zoomToSelection=true&time=2021-12-27&facet=none&pickerSort=desc&pickerMetric=total_vaccinations_per_hundred&Metric=People+vaccinated&Interval=Cumulative&Relative+to+Population=true&Color+by+test+positivity=false&country=AUT~BEL~BGR~CYP~CZE~DEU~DNK~ESP~EST~FIN~FRA~GRC~HRV~HUN~IRL~ITA~LTU~LUX~LVA~MLT~NLD~POL~PRT~ROU~SVK~SVN~SWE~EuropeanUnion

41. Fang FC, Naccache SN, Greninger AL. The Laboratory Diagnosis of Coronavirus Disease 2019-Frequently Asked Questions. *Clinical Infectious Diseases*. 2020;71(11):2996-3001. doi:10.1093/cid/ciaa742

42. Scialo F, Daniele A, Amato F, et al. ACE2: The Major Cell Entry Receptor for SARS-CoV-2. *Lung*. 2020;198(6):867-877. doi:10.1007/s00408-020-00408-4

43. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology 2020 19:3*. 2020;19(3):155-170. doi:10.1038/S41579-020-00468-6

44. World Health Organization. Tracking SARS-CoV-2 variants. World Health Organization. Published 2021. Accessed June 14, 2022. https://www.who.int/activities/tracking-SARS-CoV-2-variants

45. Centro de Coordinación de Alertas y Emergencias Sanitarias. *14 de Febrero de 2022. Actualización de La Situación Epidemiológica de Las Variantes de SARS-CoV-2 En España*. Vol 2022.; 2022.

46. Campbell F, Archer B, Laurenson-Schafer H, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*. 2021;26(24):1-6. doi:10.2807/1560-7917.ES.2021.26.24.2100509

47. Funk T, Pharris A, Spiteri G, et al. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance*. 2021;26(16):2100348. doi:10.2807/1560-7917.ES.2021.26.16.2100348

48. Nyberg T, Ferguson NM, Nash SG, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *The Lancet*. 2022;399(10332):1303-1312. doi:10.1016/s0140-6736(22)00462-7

49. Zammit C, Liddicoat H, Moonsie I, Makker H. Obesity and respiratory diseases. *Int J Gen Med*. 2010;3(4):335. doi:10.2147/IJGM.S11926

50. European Centre for Disease Prevention and Control. Data on SARS-CoV-2 variants in the EU/EEA. Accessed January 17, 2023. https://www.ecdc.europa.eu/en/publications-data/data-virus-variants-covid-19-eueea

51. World Health Organization. Coronavirus disease (COVID-19): How is it transmitted? Published December 23, 2021. Accessed September 8, 2022. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted

52. Novelli G, Biancolella M, Mehrian-Shai R, et al. COVID-19 update: the first 6 months of the pandemic. *Hum Genomics*. 2020;14(1). doi:10.1186/S40246-020-00298-W

53. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA*. 2020;323(18):1843-1844. doi:10.1001/jama.2020.3786

54. Byrne AW, McEvoy D, Collins AB, et al. Inferred duration of infectious period of SARS-CoV-2: Rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*. 2020;10(8):e039856. doi:10.1136/bmjopen-2020-039856

55. Gullón P, Díez J, Cainzos-Achirica M, Franco M, Bilal U. Social inequities in cardiovascular risk factors in women and men by autonomous regions in Spain. *Gac Sanit*. 2021;35(4):326-332. doi:10.1016/j.gaceta.2020.04.014

56. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497-506. doi:10.1016/S0140-6736(20)30183-5

57. Guan W jie, Ni Z yi, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*. 2020;382(18):1708-1720. doi:10.1056/NEJMoa2002032

58. McIntosh K. *COVID-19: Clinical Features*. UpToDate. (Ted. W Post, ed.). UpToDate; 2022. Accessed September 8, 2022. https://www.uptodate.com/contents/covid-19-clinical-features

59. Sah P, Fitzpatrick MC, Zimmer CF, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proc Natl Acad Sci U S A*. 2021;118(34). doi:10.1073/pnas.2109229118

60. Oran DP, Topol EJ. The Proportion of SARS-CoV-2 Infections That Are Asymptomatic: A Systematic Review. *Ann Intern Med*. 2021;174(5):655-662. doi:10.7326/M20-6976

61. UK Health Security Agency (UKHSA). *SARS-CoV-2 Variants of Concern and Variants under Investigation in England- Technical Briefing 31*.; 2021. Accessed April 20, 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1040076/Techni cal_Briefing_31.pdf

62. European Centre for Disease Prevention and Control. Clinical characteristics of COVID-19. doi:10.22730/jmls.2020.17.2.33

63. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. 2020;323(13):1239-1242. doi:10.1001/JAMA.2020.2648

64. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(24):759-765. doi:10.15585/mmwr.mm6924e2

65. Equipo COVID-19. RENAVE. CNE. CNM (ISCIII). *Informe Nº 59. Situación de COVID-19 En España. Casos Diagnosticados a Partir 10 de Mayo. Informe COVID-19. 29 de Diciembre de 2020*.; 2020. Accessed April 14, 2022. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/Informe%20COVID-19.%20N%C2%BA%2059_29%20de%20diciembre%20de%202020.pdf

66. World Health Organization. Estimating mortality from COVID-19. Published 2020. Accessed April 14, 2022. https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19

67. Mallapaty S. How deadly is the coronavirus? Scientists are close to an answer. *Nature*. 2020;582(7813):467-468. doi:10.1038/d41586-020-01738-2

68. Equipo COVID-19. RENAVE. CNE. CNM (ISCIII). *Informe Sobre La Situación de COVID-19 En España. Informe COVID-19 Nº 30. 11 de Mayo de 2020*.; 2020. Accessed April 12, 2022. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/INFORMES/Informes%20COVID-19/Informe%20COVID-19.%20N%C2%BA%2059_29%20de%20diciembre%20de%202020.pdf

69. Cao Y, Hiyoshi A, Montgomery S. COVID-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data. *BMJ Open*. 2020;10(11):e043560. doi:10.1136/bmjopen-2020-043560

70. Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, et al. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. *BMJ*. 2020;371:m4509. doi:10.1136/bmj.m4509

71. Logue JK, Franko NM, McCulloch DJ, et al. Sequelae in Adults at 6 Months after COVID-19 Infection. *JAMA Netw Open*. 2021;4(2):e210830-e210830. doi:10.1001/jamanetworkopen.2021.0830

72. Hossain MA, Hossain KMA, Saunders K, et al. Prevalence of Long COVID symptoms in Bangladesh: A prospective Inception Cohort Study of COVID-19 survivors. *BMJ Glob Health*. 2021;6(12):e006838. doi:10.1136/bmjgh-2021-006838

73. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. Kretzschmar MEE, ed. *PLoS Med*. 2021;18(9):e1003773. doi:10.1371/journal.pmed.1003773

74. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27(4):601-615. doi:10.1038/s41591-021-01283-z

75. Greenhalgh T, Sivan M, Delaney B, Evans R, Milne R. Long covid—an update for primary care. *BMJ*. 2022;378:e072117. doi:10.1136/bmj-2022-072117

76. World Health Organization. *A Clinical Case Definition of Post COVID-19 Condition by a Delphi Consensus*.; 2020. https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

77. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*. 2021;594. doi:10.1038/s41586-021-03553-9

78. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med*. Published online 2022. doi:10.1038/s41591-022-01689-3

79. Bowe B, Xie Y, Xu E, Al-Aly Z. Kidney Outcomes in Long COVID. *Journal of the American Society of Nephrology*. 2021;32(11):2851-2862. doi:10.1681/ASN.2021060734

80. Xie Y, Xu E, Al-Aly Z. Risks of mental health outcomes in people with covid-19: cohort study. *BMJ*. 2022;376:e068993. doi:10.1136/bmj-2021-068993

81. Struyf T, Deeks JJ, Dinnes J, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database of Systematic Reviews*. 2021;2021(3). doi:10.1002/14651858.CD013665.pub2

82. S. Pavia C, M. Plummer M. The evolution of rapid antigen detection systems and their application for COVID-19 and other serious respiratory infectious diseases. *Journal of Microbiology, Immunology and Infection*. 2021;54(5):776-786. doi:10.1016/j.jmii.2021.06.003

83. Holborow A, Asad H, Porter L, et al. The clinical sensitivity of a single SARS-CoV-2 upper respiratory tract RT-PCR test for diagnosing COVID-19 using convalescent antibody as a comparator. *Clinical Medicine*. 2020;20(6):e209-e211. doi:10.7861/clinmed.2020-0555

84. Kortela E, Kirjavainen V, Ahava MJ, et al. Real-life clinical sensitivity of SARS-CoV-2 RT-PCR test in symptomatic patients. *PLoS One*. 2021;16(5 May). doi:10.1371/journal.pone.0251661

85. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med*. 2020;173(4):262-268. doi:10.7326/M20-1495

86. Wikramaratna PS, Paton RS, Ghafari M, Lourenço J. Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Eurosurveillance*. 2020;25(50):1. doi:10.2807/1560-7917.ES.2020.25.50.2000568

87. Sub-direcció General de Vigilància i Resposta a Emergències de Salut Pública. *Procediment d'actuació Enfront de Casos d'infecció Pel Nou Coronavirus SARS-CoV-2 En La Fase de Desconfinament. Indicadors de Seguiment. Actualitzat 29.06.2020.*; 2020. http://canalsalut.gencat.cat/coronavirus

88. Ministerio de Sanidad. *Estrategia de Detección Precoz, Vigilancia y Control de COVID-19.*; 2021. Accessed February 2, 2022. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/COVID19_Estrategia_vigilancia_y_control_e_indicadores.pdf

89. Brihn A, Chang J, OYong K, et al. Diagnostic Performance of an Antigen Test with RT-PCR for the Detection of SARS-CoV-2 in a Hospital Setting — Los Angeles County, California, June–August 2020. *MMWR Recommendations and Reports*. 2021;70(19):702-706. doi:10.15585/mmwr.mm7019a3

90. European Centre for Disease Prevention and Control. *Considerations for the Use of Antibody Tests for SARS-COV-2 - First Update*. ECDC; 2022. Accessed September 7, 2022. https://www.ecdc.europa.eu/sites/default/files/documents/Considerations-for-the-use-of-antibody-tests-for-SARS-CoV2-first-update.pdf

91. Ares-Blanco S, Astier-Peña MP, Gómez-Bravo R, Fernández-García M, Bueno-Ortiz JM. El papel de la atención primaria en la pandemia COVID-19: Una mirada hacia Europa. *Aten Primaria*. 2021;53(8):102134. doi:10.1016/j.aprim.2021.102134

92. Lamb YN. Nirmatrelvir Plus Ritonavir: First Approval. *Drugs*. 2022;82(5):585-591. doi:10.1007/S40265-022-01692-5

93. Cohen P, Gebo K. *COVID-19: Management of Adults with Acute Illness in the Outpatient Setting*. UpToDate. (Ted. W Post, ed.). UpToDate; 2022. Accessed January 15, 2023. https://www.uptodate.com/contents/covid-19-management-of-adults-with-acute-illness-in-the-outpatient-setting

94. Prats-Uribe A, Sena AG, Lai LYH, et al. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. *BMJ*. 2021;373:n1038. doi:10.1136/bmj.n1038

95. Axfors C, Schmitt AM, Janiaud P, et al. Mortality outcomes with hydroxychloroquine and chloroquine in COVID-19 from an international collaborative meta-analysis of randomized trials. *Nat Commun*. 2021;12(1):2349. doi:10.1038/s41467-021-22446-z

96. Ted. W Post, ed. *COVID-19: Management in Hospitalized Adults - UpToDate*. UpToDate. UpToDate Accessed January 15, 2023. https://www.uptodate.com/contents/covid-19-management-in-hospitalized-adults?topicRef=139068&source=related_link

97. Bellino S. COVID-19 treatments approved in the European Union and clinical recommendations for the management of non-hospitalized and hospitalized patients. *Ann Med*. 2022;54(1):2856-2860. doi:10.1080/07853890.2022.2133162

98. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized with

COVID-19 in the New York City Area. *JAMA*. 2020;323(20):2052-2059. doi:10.1001/jama.2020.6775

99. Centers for Disease Control and Prevention. Underlying Medical Conditions Associated with Higher Risk for Severe COVID-19: Information for Healthcare Professionals. Centers for Disease Control and Prevention. Published 2022. Accessed April 12, 2022. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html

100. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020;369:m1966. doi:10.1136/bmj.m1966

101. Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*. 2020;369:m1985. doi:10.1136/bmj.m1985

102. 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(1):777. doi:10.1038/s41467-021-21100-y

103. Peckham H, de Gruijter NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat Commun*. 2020;11(1):6317. doi:10.1038/s41467-020-19741-6

104. Mahase E. Covid-19: Why are age and obesity risk factors for serious disease? *BMJ*. 2020;371:m4130. doi:10.1136/bmj.m4130

105. Qian J, Zhao L, Ye RZ, Li XJ, Liu YL. Age-dependent Gender Differences in COVID-19 in Mainland China: Comparative Study. *Clinical Infectious Diseases*. 2020;71(9):2488-2494. doi:10.1093/cid/ciaa683

106. Lai CC, Wang JH, Hsueh PR. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. *International Journal of Infectious Diseases*. 2020;101:314-322. doi:10.1016/j.ijid.2020.10.011

107. Gupta S, Hayek SS, Wang W, et al. Factors Associated with Death in Critically Ill Patients with Coronavirus Disease 2019 in the US. *JAMA Intern Med*. 2020;180(11):1436-1447. doi:10.1001/jamainternmed.2020.3596

108. Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed

Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):458-464. doi:10.15585/mmwr.mm6915e3

109. Kragholm K, Andersen MP, Gerds TA, et al. Association between Male Sex and Outcomes of Coronavirus Disease 2019 (COVID-19)-A Danish Nationwide, Register-based Study. *Clinical Infectious Diseases*. 2021;73(11):E4025-E4030. doi:10.1093/cid/ciaa924

110. Viveiros A, Rasmuson J, Vu J, et al. Sex differences in COVID-19: candidate pathways, genetics of ACE2, and sex hormones. *American Journal of Physiology-Heart and Circulatory Physiology*. 2021;320(1):H296-H304. doi:10.1152/ajpheart.00755.2020

111. Reilev M, Kristensen KB, Pottegård A, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: A nationwide cohort. *Int J Epidemiol*. 2020;49(5):1468-1481. doi:10.1093/ije/dyaa140

112. Politi J, Martín-Sánchez M, Mercuriali L, et al. Epidemiological characteristics and outcomes of COVID-19 cases: mortality inequalities by socio-economic status, Barcelona, Spain, 24 February to 4 May 2020. *Eurosurveillance*. 2021;26(20):2001138. doi:10.2807/1560-7917.ES.2021.26.20.2001138

113. World Health Organization. Obesity. Accessed June 16, 2022. https://www.who.int/health-topics/obesity#tab=tab_1

114. Nuttall FQ. Body Mass Index. *Nutr Today*. 2015;50(3):117-128. doi:10.1097/NT.0000000000000092

115. Popkin BM, Du S, Green WD, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. *Obesity Reviews*. 2020;21(11):1-17. doi:10.1111/obr.13128

116. Recalde M, Pistillo A, Fernandez-Bertolin S, et al. Body Mass Index and Risk of COVID-19 Diagnosis, Hospitalization, and Death: A Cohort Study of 2 524 926 Catalans. *J Clin Endocrinol Metab*. 2021;106(12):E5030-E5042. doi:10.1210/clinem/dgab546

117. Gao M, Piernas C, Astbury NM, et al. Associations between body-mass index and COVID-19 severity in 6·9 million people in England: a prospective, community-based, cohort study. *Lancet Diabetes Endocrinol*. 2021;9(6):350-359. doi:10.1016/S2213-8587(21)00089-9

118. NCI. Definition of cancer - NCI Dictionary of Cancer Terms. Published 2015. Accessed June 16, 2022. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/cancer

119. Kuderer NM, Choueiri TK, Shah DP, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet*. 2020;6736(20):1-13. doi:10.1016/S0140-6736(20)31187-9

120. Lee LYW, Cazier JB, Starkey T, et al. COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. *Lancet*. 2020;395:1919–1926. doi:10.1016/S0140-6736(20)31173-9

121. Wang Q, Berger NA, Xu R. When hematologic malignancies meet COVID-19 in the United States: Infections, death and disparities. *Blood Rev*. 2021;47(Advance online publication):100775. doi:10.1016/j.blre.2020.100775

122. Wang Q, Berger NA, Xu R. Analyses of Risk, Racial Disparity, and Outcomes Among US Patients With Cancer and COVID-19 Infection. *JAMA Oncol*. 2020;44106:1-8. doi:10.1001/jamaoncol.2020.6178

123. Rolston KVI. Infections in Cancer Patients with Solid Tumors: A Review. *Infect Dis Ther*. 2017;6(1):69-83. doi:10.1007/s40121-017-0146-1

124. Safdar A, Armstrong D. Infections in patients with hematologic neoplasms and hematopoietic stem cell transplantation: Neutropenia, humoral, and splenic defects. *Clinical Infectious Diseases*. 2011;53(8):798-806. doi:10.1093/cid/cir492

125. Niedzwiedz CL, O'Donnell CA, Jani BD, et al. Ethnic and socioeconomic differences in SARS-CoV-2 infection: Prospective cohort study using UK Biobank. *BMC Med*. 2020;18(1). doi:10.1186/s12916-020-01640-8

126. Departament de Salut Generalitat de Catalunya. COVID Data. Published 2021. Accessed April 14, 2022. https://dadescovid.cat/?lang=eng

127. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. *Build Back Fairer: The Covid-19 Marmot Review. The Pandemic, Socioeconomic and Health Inequalities in England.*; 2020. https://www.health.org.uk/publications/build-back-fairer-the-covid-19-marmot-review

128. Khanijahani A, Iezadi S, Gholipour K, Azami-Aghdash S, Naghibi D. A systematic review of racial/ethnic and socioeconomic disparities in

COVID-19. *Int J Equity Health*. 2021;20(1):248. doi:10.1186/s12939-021-01582-4

129. Jaljaa A, Caminada S, Tosti ME, et al. Risk of SARS-CoV-2 infection in migrants and ethnic minorities compared with the general population in the European WHO region during the first year of the pandemic: a systematic review. *BMC Public Health*. 2022;22(1):143. doi:10.1186/s12889-021-12466-1

130. Vandentorren S, Smaïli S, Chatignoux E, et al. The effect of social deprivation on the dynamic of SARS-CoV-2 infection in France: a population-based analysis. *Lancet Public Health*. 2022;7(3):e240-e249. doi:10.1016/S2468-2667(22)00007-X

131. Mackey K, Ayers CK, Kondo KK, et al. Racial and ethnic disparities in covid-19-related infections, hospitalizations, and deaths a systematic review. *Ann Intern Med*. 2021;174(3):362-373. doi:10.7326/M20-6306

132. Guijarro C, Pérez-Fernández E, González-Piñeiro B, et al. Differential risk for COVID-19 in the first wave of the disease among Spaniards and migrants from different areas of the world living in Spain. *Rev Clin Esp*. 2021;221(5):264-273. doi:10.1016/j.rce.2020.10.006

133. Selden TM, Berdahl TA. COVID-19 And Racial/Ethnic Disparities In Health Risk, Employment, And Household Composition. *Health Aff*. 2020;39(9):1624-1632. doi:10.1377/hlthaff.2020.00897

134. Adler NE, Newman K. Socioeconomic Disparities In Health: Pathways And Policies. *Health Aff*. 2002;21(2):60-76. doi:10.1377/hlthaff.21.2.60

135. Office for National Statistics. Coronavirus (COVID-19) case rates by socio-demographic characteristics, England. Accessed April 13, 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19caseratesbysociodemographiccharacteristicsengland/1september2020to25july2021

136. Batty GD, Gaye B, Gale CR, Hamer M, Lassale C. Explaining Ethnic Differentials in COVID-19 Mortality: A Cohort Study. *Am J Epidemiol*. 2022;191(2):275-281. doi:10.1093/aje/kwab237

137. Mahajan S, Caraballo C, Lu Y, et al. Trends in Differences in Health Status and Health Care Access and Affordability by Race and Ethnicity in the United States, 1999-2018. *JAMA - Journal of the American Medical Association*. 2021;326(7):637-648. doi:10.1001/jama.2021.9907

138. Gullón P, Díez J, Cainzos-Achirica M, Franco M, Bilal U. Social inequities in cardiovascular risk factors in women and men by autonomous regions in Spain. *Gac Sanit*. 2021;35(4):326-332. doi:10.1016/j.gaceta.2020.04.014

139. Espelt A, Kunst AE, Palència L, Gnavi R, Borrell C. Twenty years of socio-economic inequalities in type 2 diabetes mellitus prevalence in Spain, 1987-2006. *Eur J Public Health*. 2012;22(6):765-771. doi:10.1093/eurpub/ckr158

140. Espelt A, Borrell C, Roskam AJ, et al. Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century. *Diabetologia*. 2008;51(11):1971-1979. doi:10.1007/s00125-008-1146-1

141. Ho FK, Gray SR, Welsh P, et al. Ethnic differences in cardiovascular risk: examining differential exposure and susceptibility to risk factors. *BMC Med*. 2022;20(1):149. doi:10.1186/s12916-022-02337-w

142. Loring B, Robertson A. *Obesity and Inequities: Guidance for Addressing Inequities in Overweight and Obesity*.; 2014. Accessed May 19, 2022. http://www.euro.who.int/pubrequest

143. Dolgin E. The tangled history of mRNA vaccines. *Nature*. 2021;597(7876):318-324. doi:10.1038/d41586-021-02483-w

144. Mendonça SA, Lorincz R, Boucher P, Curiel DT. Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic. *NPJ Vaccines*. 2021;6(1):97. doi:10.1038/s41541-021-00356-x

145. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*. 2020;383(27):2603-2615. doi:10.1056/nejmoa2034577

146. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine*. 2021;384(5):403-416. doi:10.1056/nejmoa2035389

147. Voysey M, Clemens SAC, Madhi SA, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*. 2020;397(10269):2021. doi:10.1016/S0140-6736(20)32661-1

148. Sadoff J, Gray G, Vandebosch A, et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *New England Journal of Medicine*. 2021;384(23):2187-2201. doi:10.1056/nejmoa2101544

149. Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *New England Journal of Medicine*. 2021;384(15):1412-1423. doi:10.1056/nejmoa2101765

150. Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ*. 2021;373:n1088. doi:10.1136/bmj.n1088

151. Cabezas C, Coma E, Mora-Fernandez N, et al. Associations of BNT162b2 vaccination with SARS-CoV-2 infection and hospital admission and death with covid-19 in nursing homes and healthcare workers in Catalonia: prospective cohort study. *BMJ*. 2021;374:n1868. doi:10.1136/bmj.n1868

152. Tenforde MW, Olson SM, Self WH, et al. Effectiveness of Pfizer-BioNTech and Moderna Vaccines Against COVID-19 Among Hospitalized Adults Aged ≥65 Years — United States, January–March 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(18):674-679. doi:10.15585/mmwr.mm7018e1

153. Shrotri M, Krutikov M, Palmer T, et al. Vaccine effectiveness of the first dose of ChAdOx1 nCoV-19 and BNT162b2 against SARS-CoV-2 infection in residents of long-term care facilities in England (VIVALDI): a prospective cohort study. *Lancet Infect Dis*. 2021;21(11):1529-1538. doi:10.1016/S1473-3099(21)00289-9

154. Nordström P, Ballin M, Nordström A. Risk of infection, hospitalisation, and death up to 9 months after a second dose of COVID-19 vaccine: a retrospective, total population cohort study in Sweden. *The Lancet*. 2022;399(10327):814-823. doi:10.1016/S0140-6736(22)00089-7

155. Ferdinands JM, Rao S, Dixon BE, et al. Waning of vaccine effectiveness against moderate and severe covid-19 among adults in the US from the VISION network: test negative, case-control study. *BMJ*. Published online October 3, 2022:e072141. doi:10.1136/bmj-2022-072141

156. Gram MA, Emborg HD, Schelde AB, et al. Vaccine effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the Alpha, Delta, or Omicron SARS-CoV-2 variant: A nationwide Danish cohort study. *PLoS Med*. 2022;19(9). doi:10.1371/JOURNAL.PMED.1003992

157. Thompson MG, Natarajan K, Irving SA, et al. Effectiveness of a Third Dose of mRNA Vaccines Against COVID-19-Associated Emergency Department and Urgent Care Encounters and Hospitalizations Among

Adults During Periods of Delta and Omicron Variant Predominance - VISION Network, 10 States, August 2021–January 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71(4):139-145. doi:10.15585/MMWR.MM7104E3

158. European Medicines Agency. *Heterologous Primary and Booster COVID-19 Vaccination: Evidence Based Regulatory Considerations*.; 2021. Accessed January 2, 2023. https://www.ema.europa.eu/en/documents/report/heterologous-primary-booster-covid-19-vaccination-evidence-based-regulatory-considerations_en.pdf

159. European Medicines Agency. ECDC and EMA update recommendations on additional booster doses of mRNA COVID-19 vaccines. Published July 11, 2022. Accessed January 2, 2023. https://www.ema.europa.eu/en/news/ecdc-ema-update-recommendations-additional-booster-doses-mrna-covid-19-vaccines

160. Barda N, Dagan N, Cohen C, et al. Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: an observational study. *The Lancet*. 2021;398(10316):2093-2100. doi:10.1016/S0140-6736(21)02249-2

161. Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology*. Vol 6.; 2016.

162. Warkentin T, Cuker A. *COVID-19: Vaccine-Induced Immune Thrombotic Thrombocytopenia (VITT)*. UpToDate. (Ted. W Post, ed.). UpToDate; 2022. Accessed December 28, 2022. https://www.uptodate.com/contents/covid-19-vaccine-induced-immune-thrombotic-thrombocytopenia-vitt

163. Shoaibi A, Rao GA, Voss EA, et al. Phenotype Algorithms for the Identification and Characterization of Vaccine-Induced Thrombotic Thrombocytopenia in Real World Data: A Multinational Network Cohort Study. *Drug Saf*. 2022;45(6):685-698. doi:10.1007/s40264-022-01187-y

164. Buoninfante A, Andeweg A, Baker AT, et al. Understanding thrombosis with thrombocytopenia syndrome after COVID-19 vaccination. *NPJ Vaccines*. 2022;7(1). doi:10.1038/S41541-022-00569-8

165. Dix C, McFadyen J, Huang A, Chunilal S, Chen V, Tran H. Understanding vaccine-induced thrombotic thrombocytopenia (VITT). *Intern Med J*. Published online April 21, 2022. doi:10.1111/IMJ.15783

166. See I, Su JR, Lale A, et al. US Case Reports of Cerebral Venous Sinus Thrombosis With Thrombocytopenia After Ad26.COV2.S Vaccination, March 2 to April 21, 2021. *JAMA*. 2021;325(24):2448. doi:10.1001/jama.2021.7517

167. Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. *New England Journal of Medicine*. 2021;384(22):2092-2101. doi:10.1056/NEJMoa2104840

168. Wise J. Covid-19: European countries suspend use of Oxford-AstraZeneca vaccine after reports of blood clots. *BMJ*. 2021;372:n699. doi:10.1136/bmj.n699

169. De Benito E. La vacuna de AstraZeneca se utilizará en personas de 60 a 69 años | Sociedad | EL PAÍS. *El País*.

170. Patone M, Mei XW, Handunnetthi L, et al. Risks of myocarditis, pericarditis, and cardiac arrhythmias associated with COVID-19 vaccination or SARS-CoV-2 infection. *Nat Med*. Published online December 14, 2021:1-13. doi:10.1038/s41591-021-01630-0

171. Edwards KM, Orenstein WA. *COVID-19: Vaccines*. UpToDate. (Ted. W. Post, ed.). UpToDate; 2022. Accessed January 15, 2023. https://www.uptodate.com/contents/covid-19-vaccines

172. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus J V. Pandemic publishing poses a new COVID-19 challenge. *Nat Hum Behav*. 2020;4(7):666-669. doi:10.1038/s41562-020-0911-0

173. Fox MP, Murray EJ, Lesko CR, Sealy-Jefferson S. On the Need to Revitalize Descriptive Epidemiology. *Am J Epidemiol*. Published online March 22, 2022. doi:10.1093/aje/kwac056

174. Westreich D, Edwards JK, van Smeden M. Comment on Williamson et al. (OpenSAFELY): The Table 2 Fallacy in a Study of COVID-19 Mortality Risk Factors. *Epidemiology*. 2021;32(1):e1-e2. doi:10.1097/EDE.0000000000001259

175. Tennant P, Murray EJ. The Quest for Timely Insights into COVID-19 Should not Come at the Cost of Scientific Rigor. *Epidemiology*. 2021;32(1):e2. doi:10.1097/EDE.0000000000001258

176. Williamson EJ, Walker AJ, Bhaskaran K, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature*. Published online 2020. doi:10.1038/s41586-020-2521-4

177. Westreich D, Greenland S. The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol*. 2013;177(4):292-298. doi:10.1093/aje/kws412

178. Williamson E, Bhaskaran K, Walker A, Smeeth L, Goldacre B. The Authors Respond. *Epidemiology*. Published online 2020:E2-E3. doi:10.1097/EDE.0000000000001260

179. Prats-Uribe A, Xie J, Prieto-Alhambra D, Petersen I. Smoking and covid-19 infection and related mortality: A prospective cohort analysis of uk biobank data. *Clin Epidemiol*. 2021;13:357-365. doi:10.2147/CLEP.S300597

180. Holmberg MJ, Andersen LW. Collider Bias. *JAMA*. 2022;327(13):1282. doi:10.1001/jama.2022.1820

181. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11(1):5749. doi:10.1038/s41467-020-19478-2

182. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48. doi:10.1097/00001648-199901000-00008

183. Lee LYW, Cazier JB, Starkey T, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol*. 2020;21(10):1309-1316. doi:10.1016/S1470-2045(20)30442-3

184. Angelis V, Tippu Z, Joshi K, et al. Defining the true impact of coronavirus disease 2019 in the at-risk population of patients with cancer. *Eur J Cancer*. 2020;136:99-106. doi:10.1016/j.ejca.2020.06.027

185. Brar G, Pinheiro LC, Shusterman M, et al. COVID-19 Severity and Outcomes in Patients With Cancer: A Matched Cohort Study. *Journal of Clinical Oncology*. 2020;38(33):3914-3924. doi:10.1200/jco.20.01580

186. Robinson AG, Gyawali B, Evans G. COVID-19 and cancer: do we really know what we think we know? *Nat Rev Clin Oncol*. 2020;17(7):386-388. doi:10.1038/s41571-020-0394-y

187. Khalatbari-Soltani S, Cumming RC, Delpierre C, Kelly-Irving M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J Epidemiol Community Health (1978)*. 2020;74(8):620-623. doi:10.1136/jech-2020-214297

188. Dorn A van, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *The Lancet*. 2020;395(10232):1243-1244. doi:10.1016/s0140-6736(20)30893-x

189. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891. doi:10.1001/jama.2020.6548

190. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? *BMJ*. 2020;369:m1548. doi:10.1136/bmj.m1548

191. Curtis HJ, Inglesby P, Morton CE, et al. Trends and clinical characteristics of COVID-19 vaccine recipients: A federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. *British Journal of General Practice*. 2022;72(714):E51-E62. doi:10.3399/BJGP.2021.0376

192. Barry V, Dasgupta S, Weller DL, et al. Patterns in COVID-19 Vaccination Coverage, by Social Vulnerability and Urbanicity — United States, December 14, 2020–May 1, 2021. *MMWR Recommendations and Reports*. 2021;70(22):818-824. doi:10.15585/mmwr.mm7022e1

193. Dades COVID VACUNACIÓ EDAT I SEXE. Accessed September 7, 2022. https://dadescovid.cat/vacunacio/cobertura_edat_sexe

194. Pottegård A, Lund LC, Karlstad Ø, et al. Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study. *BMJ*. 2021;373:n1114. doi:10.1136/bmj.n1114

195. Simpson CR, Shi T, Vasileiou E, et al. First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland. *Nat Med*. 2021;27(7):1290-1297. doi:10.1038/s41591-021-01408-4

196. Food and Drug Administration. Real-World Evidence, FDA. FDA Website. Published 2020. Accessed June 10, 2022. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

197. Bérard A. Pharmacoepidemiology Research-Real-World Evidence for Decision Making. *Front Pharmacol*. 2021;12. doi:10.3389/FPHAR.2021.723427

198. Toh S. Analytic and Data Sharing Options in Real-World Multidatabase Studies of Comparative Effectiveness and Safety of Medical Products. *Clin Pharmacol Ther*. 2020;107(4):834-842. doi:10.1002/cpt.1754

199. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine*. 2016;375(23):2293-2297. doi:10.1056/nejmsb1609216

200. Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *The Lancet*. 2019;394(10211):1816-1826. doi:10.1016/S0140-6736(19)32317-7

201. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*. 2016;113(27):7329-7336. doi:10.1073/pnas.1510502113

202. 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(1). doi:10.1186/S12916-020-01877-3

203. Williams RD, Markus AF, Yang C, et al. Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network. *BMC Med Res Methodol*. 2022;22(1):35. doi:10.1186/s12874-022-01505-z

204. Reynolds MW, Bourke A, Dreyer NA. Considerations when evaluating real-world data quality in the context of fitness for purpose. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1316-1318. doi:10.1002/pds.5010

205. Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med*. 2014;275(6):551-561. doi:10.1111/joim.12159

206. Gini R, Sturkenboom MCJ, Sultana J, et al. Different Strategies to Execute Multi-Database Studies for Medicines Surveillance in Real-World Setting: A Reflection on the European Model. *Clin Pharmacol Ther*. 2020;108(2):228-235. doi:10.1002/cpt.1833

207. Kent S, Burn E, Dawoud D, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *Pharmacoeconomics*. 2021;39(3):275-285. doi:10.1007/s40273-020-00981-9

208. About the Food and Drug Administration (FDA) Sentinel Initiative | Sentinel Initiative. Accessed January 19, 2023. https://www.sentinelinitiative.org/about

209. Home - The National Patient-Centered Clinical Research Network. Accessed January 19, 2023. https://pcornet.org/

210. Thurin NH, Pajouheshnia R, Roberto G, et al. From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding. *Clin Pharmacol Ther*. 2022;111(1):321-331. doi:10.1002/CPT.2476

211. OHDSI. OMOP Common Data Model. Accessed June 10, 2022. https://ohdsi.github.io/CommonDataModel/cdm53.html#Clinical_Data_Tables

212. Recalde M, Rodriguez C, Burn E, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). *Int J Epidemiol*. Published online 2022. doi:10.1093/ije/dyac068

213. Kostka K, Duarte-Salles T, Prats-Uribe A, et al. Unraveling COVID-19: A Large-Scale Characterization of 4.5 Million COVID-19 Cases Using CHARYBDIS. *Clin Epidemiol*. 2022;Volume 14:369-384. doi:10.2147/CLEP.S323292

214. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-578. doi:10.3233/978-1-61499-564-7-574

215. Burn E, Fernández-bertolín S, Voss EA, Blacketer C. Establishing and characterising large COVID-19 cohorts after mapping the Information System for Research in Primary Care in Catalonia to the OMOP Common Data Model. *medRxiv*. Published online 2021:2021.11.23.21266734. doi:10.1101/2021.11.23.21266734

216. Observational Health Data Sciences and Informatics. *The Book of OHDSI.*; 2019. https://ohdsi.github.io/TheBookOfOhdsi/TheBookOfOhdsi.pdf

217. National Library of Medicine. RxNorm. Accessed June 10, 2022. https://www.nlm.nih.gov/research/umls/rxnorm/index.html

218. Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. [Constructing a deprivation index based on census data in large Spanish cities (the MEDEA project)]. *Gac Sanit*. 2008;22(3):179-187. doi:10.1157/13123961

219. UK Health Security Agency. *Technical Summary. UK Health Security Agency Data Series on Deaths in People with COVID-19*. Accessed February 24, 2023. https://www.gov.uk/government/publications/phe-data-series-on-deaths-in-people-with-covid-19-technical-summary?utm_source=dac7e433-2d30-42a4-9622-

ba39a93af8d7&utm_medium=email&utm_campaign=govuk-notifications&utm_content=immediate

220. Burn E, Li X, Kostka K, et al. Background rates of five thrombosis with thrombocytopenia syndromes of special interest for <scp>COVID</scp> -19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. *Pharmacoepidemiol Drug Saf*. 2022;31(5):495-510. doi:10.1002/pds.5419

221. Li X, Ostropolets A, Makadia R, et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. *BMJ*. 2021;373:n1435. doi:10.1136/bmj.n1435

222. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput*. 2009;38(6):1228-1234. doi:10.1080/03610910902859574

223. Putter H, Fiocco M, Gekus RB. Tutorial in biostatistics: Competing risk and multi-state models. *Stat Med*. 2007;26(11):2389-2430. doi:10.1002/sim.2712

224. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*. 2019;69(5):363-385. doi:10.3322/caac.21565

225. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. Published online February 4, 2021:caac.21660. doi:10.3322/caac.21660

226. Burn E, You SC, Sena AG, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. *Nat Commun*. 2020;11(1)(1):5009. doi:10.1038/s41467-020-18849-z

227. Ludwig M, Jacob J, Basedow F, Andersohn F, Walker J. Clinical outcomes and characteristics of patients hospitalized for Influenza or COVID-19 in Germany. *International Journal of Infectious Diseases*. 2021;103:316-322. doi:10.1016/j.ijid.2020.11.204

228. Khoury E, Nevitt S, William ;, Madsen R, Davies G, Palmieri C. Differences in Outcomes and Factors Associated With Mortality Among Patients With SARS-CoV-2 Infection and Cancer Compared With Those Without Cancer A Systematic Review and Meta-analysis + Supplemental

content. *JAMA Netw Open*. 2022;5(5):2210880. doi:10.1001/jamanetworkopen.2022.10880

229. Pagano L, Salmanton-García J, Marchesi F, et al. COVID-19 infection in adult patients with hematological malignancies: a European Hematology Association Survey (EPICOVIDEHA). *J Hematol Oncol*. 2021;14(1). doi:10.1186/S13045-021-01177-0

230. Elkrief A, Wu JT, Jani C, et al. Learning through a Pandemic: The Current State of Knowledge on COVID-19 and Cancer. *Cancer Discov*. 2022;12(2):303-330. doi:10.1158/2159-8290.CD-21-1368

231. Uzzo R, Kutikov A, Geynisman DM. *COVID-19: Considerations in Patients with Cancer*. UpToDatw. (Ted. W Post, ed.). UpToDate; 2022. Accessed January 8, 2023. https://www.uptodate.com/contents/covid-19-considerations-in-patients-with-cancer

232. Luxenburger H, Thimme R. Capturing adaptive immunity against SARS-CoV-2 in patients with cancer. *Nat Cancer*. 2021;2(12):1298-1300. doi:10.1038/s43018-021-00298-2

233. Fendler A, Au L, Shepherd STC, et al. Functional antibody and T cell immunity following SARS-CoV-2 infection, including by variants of concern, in patients with cancer: the CAPTURE study. *Nat Cancer*. 2021;2(12):1321-1337. doi:10.1038/s43018-021-00275-9

234. Thakkar A, Pradhan K, Jindal S, et al. Patterns of seroconversion for SARS-CoV-2 IgG in patients with malignant disease and association with anticancer therapy. *Nat Cancer*. 2021;2(4):392-399. doi:10.1038/s43018-021-00191-y

235. Yekedüz E, Utkan G, Ürün Y. A systematic review and meta-analysis: the effect of active cancer treatment on severity of COVID-19. *Eur J Cancer*. 2020;141:92-104. doi:10.1016/J.EJCA.2020.09.028

236. de Azambuja E, Brandão M, Wildiers H, et al. Impact of solid cancer on in-hospital mortality overall and among different subgroups of patients with COVID-19: a nationwide, population-based analysis. *ESMO Open*. 2020;5(5):e000947. doi:10.1136/esmoopen-2020-000947

237. Mangone L, Gioia F, Mancuso P, et al. Cumulative COVID-19 incidence, mortality and prognosis in cancer survivors: A population-based study in Reggio Emilia, Northern Italy. *Int J Cancer*. 2021;149(4):820-826. doi:10.1002/ijc.33601

238. Várnai C, Palles C, Arnold R, et al. Mortality Among Adults With Cancer Undergoing Chemotherapy or Immunotherapy and Infected With

COVID-19. *JAMA Netw Open*. 2022;5(2):e220130. doi:10.1001/jamanetworkopen.2022.0130

239. Marí-Dell'olmo M, Gotsens M, Pasarín MI, et al. Socioeconomic inequalities in COVID-19 in a European urban area: Two waves, two patterns. *Int J Environ Res Public Health*. 2021;18(3):1-12. doi:10.3390/ijerph18031256

240. Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. *Immunity*. 2020;52(5):737-741. doi:10.1016/j.immuni.2020.04.012

241. Malmusi D, Pasarín MI, Marí-Dell'Olmo M, et al. Multi-level policy responses to tackle socioeconomic inequalities in the incidence of COVID-19 in a European urban area. *Int J Equity Health*. 2022;21(1):28. doi:10.1186/s12939-022-01628-1

242. Guo Y, Kaniuka AR, Gao J, Sims OT. An Epidemiologic Analysis of Associations between County-Level Per Capita Income, Unemployment Rate, and COVID-19 Vaccination Rates in the United States. *Int J Environ Res Public Health*. 2022;19(3):1755. doi:10.3390/ijerph19031755

243. DiRago N v., Li M, Tom T, et al. COVID-19 Vaccine Rollouts and the Reproduction of Urban Spatial Inequality: Disparities Within Large US Cities in March and April 2021 by Racial/Ethnic and Socioeconomic Composition. *J Urban Health*. 2022;99(2):191-207. doi:10.1007/S11524-021-00589-0

244. Schnake-Mahl AS, O'Leary G, Mullachery PH, et al. Higher COVID-19 Vaccination And Narrower Disparities In US Cities With Paid Sick Leave Compared To Those Without. *Health Aff*. 2022;41(11):1565-1574. doi:10.1377/hlthaff.2022.00779

245. Hughes MM, Wang A, Grossman MK, et al. County-Level COVID-19 Vaccination Coverage and Social Vulnerability — United States, December 14, 2020-March 1, 2021. *MMWR Surveillance Summaries*. 2021;70(12):431-436. doi:10.15585/mmwr.mm7012e1

246. Nafilyan V, Dolby T, Razieh C, et al. Sociodemographic inequality in COVID-19 vaccination coverage among elderly adults in England: a national linked data study. *BMJ Open*. 2021;11(7):e053402. doi:10.1136/bmjopen-2021-053402

247. Perry M, Akbari A, Cottrell S, et al. Inequalities in coverage of COVID-19 vaccination: A population register based cross-sectional study in Wales, UK. *Vaccine*. 2021;39(42):6256-6261. doi:10.1016/j.vaccine.2021.09.019

248. Mody A, Bradley C, Redkar S, et al. Quantifying inequities in COVID-19 vaccine distribution over time by social vulnerability, race and ethnicity, and location: A population-level analysis in St. Louis and Kansas City, Missouri. Low N, ed. *PLoS Med*. 2022;19(8):e1004048. doi:10.1371/journal.pmed.1004048

249. Li X, Burn E, Duarte-Salles T, et al. Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different covid-19 vaccines: international network cohort study from five European countries and the US. *BMJ*. Published online October 26, 2022:e071594. doi:10.1136/bmj-2022-071594

250. Hippisley-Cox J, Patone M, Mei XW, et al. Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: self-controlled case series study. *BMJ*. 2021;374:n1931. doi:10.1136/bmj.n1931

251. Kerr S, Joy M, Torabi F, et al. First dose ChAdOx1 and BNT162b2 COVID-19 vaccinations and cerebral venous sinus thrombosis: A pooled self-controlled case series study of 11.6 million individuals in England, Scotland, and Wales. *PLoS Med*. 2022;19(2). doi:10.1371/journal.pmed.1003927

252. Whiteley WN, Ip S, Cooper JA, et al. Association of COVID-19 vaccines ChAdOx1 and BNT162b2 with major venous, arterial, or thrombocytopenic events: A population-based cohort study of 46 million adults in England. *PLoS Med*. 2022;19(2). doi:10.1371/journal.pmed.1003926

253. Torabi F, Bedston S, Lowthian E, et al. Risk of thrombocytopenic, haemorrhagic and thromboembolic disorders following COVID-19 vaccination and positive test: a self-controlled case series analysis in Wales. *Scientific Reports 2022 12:1*. 2022;12(1):1-13. doi:10.1038/S41598-022-20118-6

254. Burn E, Li X, Delmestri A, et al. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in the United Kingdom. *Nat Commun*. 2022;13(1):7167. doi:10.1038/S41467-022-34668-W

255. Dag Berild J, Bergstad Larsen V, Myrup Thiesson E, et al. Analysis of Thromboembolic and Thrombocytopenic Events After the AZD1222, BNT162b2, and MRNA-1273 COVID-19 Vaccines in 3 Nordic Countries. *JAMA Netw Open*. Published online 2022:E2217375. doi:10.1001/jamanetworkopen.2022.17375

256. Garbe E, Andersohn F, Bronder E, et al. Drug-induced immune thrombocytopaenia: results from the Berlin Case-Control Surveillance Study. *Eur J Clin Pharmacol*. 2012;68(5):821-832. doi:10.1007/S00228-011-1184-3

257. Geier DA, Geier MR. A case-control study of serious autoimmune adverse events following hepatitis B immunization. *Autoimmunity*. 2005;38(4):295-301. doi:10.1080/08916930500144484

258. France EK, Glanz J, Xu S, et al. Risk of Immune Thrombocytopenic Purpura After Measles-Mumps-Rubella Immunization in Children. *Pediatrics*. 2008;121(3):e687-e692. doi:10.1542/peds.2007-1578

259. Bertuola F, Morando C, Menniti-Ippolito F, et al. Association between drug and vaccine use and acute immune thrombocytopenia in childhood: a case-control study in Italy. *Drug Saf*. 2010;33(1):65-72. doi:10.2165/11530350-000000000-00000

260. Burn E, Roel E, Pistillo A, et al. Thrombosis and thrombocytopenia after vaccination against and infection with SARS-CoV-2 in Catalonia, Spain. *Nat Commun*. 2022;13(1):7169. doi:10.1038/s41467-022-34669-9

261. Ashrani AA, Crusan DJ, Petterson T, Bailey K, Heit JA. Age- and Sex-Specific Incidence of Cerebral Venous Sinus Thrombosis Associated With Ad26.COV2.S COVID-19 Vaccination. *JAMA Intern Med*. 2022;182(1):80. doi:10.1001/jamainternmed.2021.6352

262. Barda N, Dagan N, Ben-Shlomo Y, et al. Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting. *New England Journal of Medicine*. 2021;385(12):1078-1090. doi:10.1056/nejmoa2110475

263. Wells BJ, Nowacki AS, Chagin K, Kattan MW. Strategies for Handling Missing Data in Electronic Health Record Derived Data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2013;1(3):7. doi:10.13063/2327-9214.1035

264. Coleman N, Halas G, Peeler W, Casaclang N, Williamson T, Katz A. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract*. 2015;16(1):11. doi:10.1186/s12875-015-0223-z

265. Recalde M, Manzano-Salgado CB, 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;11:1015-1024. doi:10.2147/CLEP.S225568

266. 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. *Revista Española de Cardiología (English Edition)*. 2012;65(1):29-37. doi:10.1016/j.rec.2011.07.016

267. 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;Volume 11:217-228. doi:10.2147/CLEP.S186590

268. Li X, Lai LY, Ostropolets A, et al. Bias, Precision and Timeliness of Historical (Background) Rate Comparison Methods for Vaccine Safety Monitoring: An Empirical Multi-Database Analysis. *Front Pharmacol*. 2021;12. doi:10.3389/fphar.2021.773875

269. Embrett M, Sim SM, Caldwell HAT, et al. Barriers to and strategies to address COVID-19 testing hesitancy: a rapid scoping review. *BMC Public Health*. 2022;22(1):750. doi:10.1186/s12889-022-13127-7

270. 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. *Endocrinol Diabetes Nutr*. 2017;64:15-22. doi:10.1016/j.endonu.2016.07.002

271. 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. *Int J Cancer*. 2022;150(5):782-794. doi:10.1002/ijc.33846

272. Yang Y, Dugué PA, Lynch BM, et al. Trajectories of body mass index in adulthood and all-cause and cause-specific mortality in the Melbourne Collaborative Cohort Study. *BMJ Open*. 2019;9(8):e030078. doi:10.1136/bmjopen-2019-030078

273. Fan B, Yang Y, Dayimu A, et al. Body Mass Index Trajectories During Young Adulthood and Incident Hypertension: A Longitudinal Cohort in Chinese Population. *J Am Heart Assoc*. 2019;8(8). doi:10.1161/JAHA.119.011937

274. Heymsfield SB, Peterson CM, Thomas DM, Heo M, Schuna JM. Why are there race/ethnic differences in adult body mass index-adiposity relationships? A quantitative critical review. *Obesity Reviews*. 2016;17(3):262-275. doi:10.1111/obr.12358

275. Gonzalez MC, Correia MITD, Heymsfield SB. A requiem for BMI in the clinical setting. *Curr Opin Clin Nutr Metab Care*. 2017;20(5):314-321. doi:10.1097/MCO.0000000000000395

276. Fillmore NR, La J, Szalat RE, et al. Prevalence and Outcome of COVID-19 Infection in Cancer Patients: A National Veterans Affairs Study. *JNCI: Journal of the National Cancer Institute*. 2021;113(6):691-698. doi:10.1093/jnci/djaa159

277. Chen R, Ryan P, Natarajan K, et al. Treatment Patterns for Chronic Comorbid Conditions in Patients With Cancer Using a Large-Scale Observational Data Network. *JCO Clin Cancer Inform*. 2020;4(4):171-183. doi:10.1200/CCI.19.00107

278. Pedersen A, Mikkelsen E, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017;Volume 9:157-166. doi:10.2147/CLEP.S129785

279. Diez Roux A v., Kiefe CI, Jacobs DR, et al. Area Characteristics and Individual-Level Socioeconomic Position Indicators in Three Population-Based Epidemiologic Studies. *Ann Epidemiol*. 2001;11(6):395-405. doi:10.1016/S1047-2797(01)00221-6

280. Exeter DJ, Boyle PJ, Norman P. Deprivation (im)mobility and cause-specific premature mortality in Scotland. *Soc Sci Med*. 2011;72(3):389-397. doi:10.1016/j.socscimed.2010.10.009

281. Duque I, Domínguez-Berjón MF, Cebrecos A, et al. Índice de privación en España por sección censal en 2011. *Gac Sanit*. 2021;35(2):113-122. doi:10.1016/j.gaceta.2019.10.008

282. Larsen VB, Grøsland M, Telle K, Magnusson K. Utilization of health care services before and after media attention about fatal side effects of the AstraZeneca vaccine: a nation-wide register-based event study. *BMC Health Serv Res*. 2021;21(1):1229. doi:10.1186/s12913-021-07233-2

283. Kyriacou DN, Lewis RJ. Confounding by Indication in Clinical Research. *JAMA*. 2016;316(17):1818. doi:10.1001/jama.2016.16435

284. Mathieu E, Ritchie H, Ortiz-Ospina E, et al. Coronavirus Pandemic (COVID-19). *Our World in Data*. 2020;5(7):947-953. doi:10.1038/S41562-021-01122-8

285. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics. Accessed January 2, 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsoc

ialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/1december2022

286. Office for National Statistics. *Self-Reported Long COVID and Labour Market Outcomes, UK: 2022* -. Accessed January 2, 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/selfreportedlongcovidandlabourmarketoutcomesuk2022/selfreportedlongcovidandlabourmarketoutcomesuk2022

287. Huang L, Yao Q, Gu X, et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study. *The Lancet*. 2021;398(10302):747-758. doi:10.1016/S0140-6736(21)01755-4

288. Groff D, Sun A, Ssentongo AE, et al. Short-term and Long-term Rates of Postacute Sequelae of SARS-CoV-2 Infection. *JAMA Netw Open*. 2021;4(10):e2128568. doi:10.1001/jamanetworkopen.2021.28568

289. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine*. 2021;38:101019. doi:10.1016/j.eclinm.2021.101019

290. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep*. 2021;11(1). doi:10.1038/S41598-021-95565-8

291. Fernández-de-las-Peñas C, Rodríguez-Jiménez J, Cancela-Cilleruelo I, et al. Post–COVID-19 Symptoms 2 Years After SARS-CoV-2 Infection Among Hospitalized vs Nonhospitalized Patients. *JAMA Netw Open*. 2022;5(11):e2242106. doi:10.1001/jamanetworkopen.2022.42106

292. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. Kretzschmar MEE, ed. *PLoS Med*. 2021;18(9):e1003773. doi:10.1371/journal.pmed.1003773

293. Antonelli M, Pujol JC, Spector TD, Ourselin S, Steves CJ. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. *The Lancet*. 2022;399(10343):2263-2264. doi:10.1016/S0140-6736(22)00941-2

294. Nehme M, Vetter P, Chappuis F, et al. Prevalence of Post-COVID Condition 12 Weeks After Omicron Infection Compared With Negative Controls and Association With Vaccination Status. *Clinical Infectious Diseases*. Published online December 15, 2022. doi:10.1093/cid/ciac947

295. Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med*. 2022;28(7):1461-1467. doi:10.1038/s41591-022-01840-0

296. Antonelli M, Penfold RS, Merino J, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis*. 2022;22(1):43-55. doi:10.1016/S1473-3099(21)00460-6

297. Corti C, Curigliano G. Commentary: SARS-CoV-2 vaccines and cancer patients. *Annals of Oncology*. 2021;32(4):569-571. doi:10.1016/j.annonc.2020.12.019

298. Grupo de Trabajo Técnico de Vacunación COVID-19 de la Ponencia de Programa y Registro de Vacunaciones. *Actualización 7 - Estrategia de Vacunación Frente a COVID- 19 En España.*; 2021. Accessed July 1, 2021. https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/docs/COVID-19_Actualizacion7_EstrategiaVacunacion.pdf

299. Song Q, Bates B, Shao YR, et al. Risk and Outcome of Breakthrough COVID-19 Infections in Vaccinated Patients With Cancer: Real-World Evidence From the National COVID Cohort Collaborative. *Journal of Clinical Oncology*. 2022;40(13):1414-1427. doi:10.1200/JCO.21.02419

300. Wu JTY, La J, Branch-Elliman W, et al. Association of COVID-19 Vaccination With SARS-CoV-2 Infection in Patients With Cancer. *JAMA Oncol*. 2022;8(2):281. doi:10.1001/jamaoncol.2021.5771

301. Lee LYW, Starkey T, Ionescu MC, et al. Vaccine effectiveness against COVID-19 breakthrough infections in patients with cancer (UKCCEP): a population-based test-negative case-control study. *Lancet Oncol*. 2022;23(6):748-757. doi:10.1016/S1470-2045(22)00202-9

302. Lee LYW, Ionescu MC, Starkey T, et al. COVID-19: Third dose booster vaccine effectiveness against breakthrough coronavirus infection, hospitalisations and death in patients with cancer: A population-based study. *Eur J Cancer*. 2022;175:1-10. doi:10.1016/J.EJCA.2022.06.038

303. Murthy NC, Zell E, Fast HE, et al. Disparities in First Dose COVID-19 Vaccination Coverage among Children 5–11 Years of Age, United States. *Emerg Infect Dis*. 2022;28(5):986-989. doi:10.3201/eid2805.220166

304. Murthy BP, Zell E, Saelee R, et al. COVID-19 Vaccination Coverage Among Adolescents Aged 12–17 Years — United States, December 14, 2020–July 31, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(35):1206-1213. doi:10.15585/mmwr.mm7035e1

305. Santibanez TA, Zhou T, Black CL, et al. Sociodemographic Variation in Early Uptake of COVID-19 Vaccine and Parental Intent and Attitudes Toward Vaccination of Children Aged 6 Months–4 Years — United States, July 1–29, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71(46):1479-1484. doi:10.15585/mmwr.mm7146a3

306. Ministerio de Sanidad. *GIV COVID-19 Gestión Integral de La Vacunación COVID-19: Informe de Actividad*.; 2022. Accessed January 13, 2023. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Informe_GIV_comunicacion_20221111.pdf

307. Lu PJ, Srivastav A, Vashist K, et al. COVID-19 Booster Dose Vaccination Coverage and Factors Associated with Booster Vaccination among Adults, United States, March 2022. *Emerg Infect Dis*. 2023;29(1):133-140. doi:10.3201/eid2901.221151

308. European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU). Accessed January 3, 2023. https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu

309. European Medicines Agency. *DARWIN EU® Data Partners Onboarded in Phase I*.; 2022. Accessed January 20, 2023. https://www.ema.europa.eu/en/documents/other/darwin-eu-data-partners-onboarded-phase-i_en.pdf

310. Liu F, Demosthenes P. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology 2022 22:1*. 2022;22(1):1-10. doi:10.1186/S12874-022-01768-6

311. US Food & Drug Administration. *Framework for FDA's Real-World Evidence Program*.; 2018. Accessed January 3, 2023. https://www.fda.gov/media/120060/download

312. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *J Biomed Inform*. 2017;73:14-29. doi:10.1016/j.jbi.2017.07.012

313. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of

electronic health records: a systematic review. *Journal of the American Medical Informatics Association*. 2019;26(4):364-379. doi:10.1093/jamia/ocy173

314. Wu CY, Chang CK, Robson D, et al. Evaluation of Smoking Status Identification Using Electronic Health Records and Open-Text Information in a Large Mental Health Case Register. Gorlova OY, ed. *PLoS One*. 2013;8(9):e74262. doi:10.1371/journal.pone.0074262

315. López-Úbeda P, Martín-Noguerol T, Juluru K, Luna A. Natural Language Processing in Radiology: Update on Clinical Applications. *Journal of the American College of Radiology*. 2022;19(11):1271-1285. doi:10.1016/j.jacr.2022.06.016

316. Martinez D, Pitson G, MacKinlay A, Cavedon L. Cross-hospital portability of information extraction of cancer staging information. *Artif Intell Med*. 2014;62(1):11-21. doi:10.1016/j.artmed.2014.06.002

317. Joined up data in government: the future of data linking methods - GOV.UK. Accessed January 12, 2023. https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/joined-up-data-in-government-the-future-of-data-linkage-methods#foreword-from-professor-sir-ian-diamond

318. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol*. 2019;Volume 11:563-591. doi:10.2147/CLEP.S179083

319. Ministry of Housing Communities & Local Government. *Shielding Clinically Vulnerable People from COVID-19: Guidance for Councils, LRFs and Other Local Delivery Partners*.; 2020. Accessed January 15, 2023. www.gov.uk/government/publications/guidance-on-shielding-and-protecting-

320. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. *Soc Sci Med*. 2020;263:113356. doi:10.1016/j.socscimed.2020.113356

321. Lee SK, Sun J, Jang S, Connelly S. Misinformation of COVID-19 vaccines and vaccine hesitancy. *Scientific Reports 2022 12:1*. 2022;12(1):1-11. doi:10.1038/s41598-022-17430-6

322. World Health Organization. Coronavirus disease (COVID-19): Similarities and differences between COVID-19 and Influenza. Accessed January 13, 2023. https://www.who.int/news-room/questions-and-

answers/item/coronavirus-disease-covid-19-similarities-and-differences-with-influenza

323. Pierri F, Perry BL, DeVerna MR, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Sci Rep*. 2022;12(1):5966. doi:10.1038/s41598-022-10070-w

324. Nikoloski Z, Alqunaibet AM, Alfawaz RA, et al. Covid-19 and non-communicable diseases: evidence from a systematic literature review. *BMC Public Health*. 2021;21(1). doi:10.1186/S12889-021-11116-W

325. World Health Organization. *The Impact of the COVID-19 Pandemic on Noncommunicable Disease Resources and Services : Results of a Rapid Assessment*.; 2020. Accessed January 11, 2023. https://www.who.int/publications/i/item/9789240010291

326. World Health Organization. Noncommunicable diseases. Published September 16, 2022. Accessed January 11, 2023. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

327. World Health Organization Regional Office for Europe. *Action Plan for the Prevention and Control of Noncommunicable Diseases in the WHO European Region*.; 2016. Accessed January 11, 2023. http://www.euro.who.int/pubrequest

328. Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1223-1249. doi:10.1016/S0140-6736(20)30752-2

329. Reitsma MB, Fullman N, Ng M, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: A systematic analysis from the global burden of disease study 2015. *The Lancet*. 2017;389(10082):1885-1906. doi:10.1016/S0140-6736(17)30819-X

330. World Health Organization. *Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020*.; 2013.

331. World Health Organization. *COVID-19 and the Social Determinants of Health and Health Equity: Evidence Brief*.; 2021. Accessed January 2, 2023. https://www.who.int/publications/i/item/9789240038387

332. Hood CM, Gennuso KP, Swain GR, Catlin BB. County Health Rankings: Relationships between Determinant Factors and Health Outcomes. *Am J Prev Med*. 2016;50(2):129-135. doi:10.1016/j.amepre.2015.08.024

333. Wilkinson R, Marmot M, World Health Organization. Regional Office for Europe. *Social Determinants of Health: The Solid Facts*. 2nd ed. World Health Organization. Regional Office for Europe; 2003. Accessed January 15, 2023. https://apps.who.int/iris/handle/10665/326568

Dr Talita Duarte Salles, Senior Epidemiologist at Fundació Institut Universitari per a la recerca a l'Atenció Primaria de Salut Jordi Gol i Gurina (IDIAPJGol) and,

Dr Daniel Prieto Alhambra, Professor at the Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford

## Certify,

That the doctoral Thesis entitled "Timely insights into the COVID-19 pandemic using real-world data" presented by Elena Roel 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 and Public Health.

In witness thereof, the following document was signed in Barcelona, February 2023.

Talita Duarte Salles

Daniel Prieto Alhambra

# Timely insights into the COVID-19 pandemic using real-world data

PhD thesis

Doctoral program in Methodology of Biomedical Research and Public Health

Author: Elena Roel Herranz

The cover figure was created with the assistance of DALL·E 2.

Universitat Autònoma de Barcelona
Department of Paediatrics, Obstetrics and Gynaecology, and Preventive Medicine and Public Health
Barcelona, March 2023