Characteristics and outcomes of over 300,000 COVID-19 individuals with history of cancer in the United States and Spain

Roel, Elena; Pistillo, Andrea; Recalde, Martina; Sena, Anthony G; Fernandez-Bertolin, Sergio; Aragon, Maria

Published in:
Cancer Epidemiology, Biomarkers and Prevention

DOI:
10.1158/1055-9965.EPI-21-0266

Publication date:
2021

Document Version
Peer reviewed version

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*] Talita Duarte-Salles[1*]

*Joint senior authorship

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, NJ USA

4. Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
5. NDORMS, University of Oxford, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD, UK
6. College of Medicine and Health, University of Exeter, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU, UK.
7. Faculty of Medicine, Islamic University of Gaza, Palestine
8. Massachusetts General Hospital, Harvard Medical School, USA
9. Medication Safety Research Chair, King Saud University
10. Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, OX3 9DU, UK
11. Data Science to Patient Value Program, Department of Medicine, University of Colorado Anschutz Medical Campus
12. Real-World Evidence, Trial Form Support, Barcelona, Spain
13. Department of Data Science, Dana-Farber Cancer Institute, Boston MA, USA
14. Department of Biostatistics, Harvard TH Chan School of Public Health, Boston MA, USA
15. Faculty of Pharmacy, Cairo University, Cairo, Egypt
16. VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA
17. Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA
18. Department of Biomedical Informatics, Columbia University, New York, NY, US
19. New York-Presbyterian Hospital, 622 W 168 St, PH20 New York, NY 10032 USA

20. Department of Epidemiology, Johns Hopkins School of Public, Baltimore MD, USA

21. Pharmacoepidemiology, Regeneron Pharmaceuticals, NY, US

22. Digital Health China Technologies Co., LTD, Beijing, China

23. Associate Director, Epidemiology, Janssen Research and Development, LLC.

24. Centre for Statistics in Medicine, NDORMS, University of Oxford, OX3 7LD, UK

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

27. School of Medical Sciences, University of Manchester, UK

28. Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN, USA

29. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

30. Division of Population Health and Genomics, University of Dundee, UK

31. University of Southern 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, CA USA

34. Real World Solutions, IQVIA, Cambridge, MA USA

35. Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD

36. Section of Cancer Surveillance, International Agency for Research on Cancer , 150 Cours Albert Thomas, 69008 Lyon, France

37. College of Engineering, The University of Arizona, Tucson, Arizona, USA

38. Fielding School of Public Health, University of California, Los Angeles

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

41. School of Population Health and Global Health, The University of Melbourne

Running title: Characteristics of 300,000 COVID-19 individuals with cancer

Abbreviations: ARDS: Acute Respiratory Distress Syndrome; CDM: Common Data Model; CHARYBDIS: Characterizing Health Associated Risks, and Your Baseline Disease In SARS-COV-2; CI: Confidence Interval; COPD: Chronic Obstructive Pulmonary Disease: COVID-19: Coronavirus disease 2019; CU-AMC-DHC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; EHR: Electronic Health Record; EHR: Electronic Health Record; IRB: Institutional Review Board; OHDSI: Observational Health Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership; NHL: Non-Hodgkin's lymphoma; SARS-CoV-2: Severe Acute respiratory Syndrome Coronavirus-2; SIDIAP: Information System for Research in Primary Care; SMD: Standardized Mean Differences; SNOMED: Systematized Nomenclature of Medicine; STARR-OMOP: Stanford Medicine Research Data Repository; UK: United Kingdom; US: United States; VA-OMOP: United States Department of Veterans Affairs.

Corresponding author

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

Tel: +34935824342

Email: tduarte@idiapjgol.org

Authors' Disclosures

A.G. Sena reports personal fees from Janssen Research & Development, outside the submitted work; S.L. 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 Veteran Research, outside the submitted work; W.U.R. Ahmed reports funding from the NIHR Oxford Biomedical Research Centre (BRC), Aziz Foundation, Wolfson Foundation, and the Royal College Surgeons of England; A. 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. D.R. 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; J. Jonnagaddala reports grants from National Health and Medical Research Council, outside the submitted work; F. Nyberg reports other from AstraZeneca, outside the submitted work; N.H. 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; D. 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 D. Prieto-Alhambra's department and open for external participants. K. Kostka and C.G. Reich report being employees of IQVIA Inc. P.B. 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.

Abstract

Background

We described the demographics, cancer subtypes, comorbidities, and outcomes of patients with a history of cancer and COVID-19. Secondly, 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 healthcare databases from Spain and the US, 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 2017-2018. 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: 5-19% and 1-14% in the *diagnosed* cohort, respectively). Hematological malignancies were also frequent, with non-Hodgkin's lymphoma being among the 5 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 h*ospitalized* 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. Hematological malignancies were frequent.

Impact

This study provides epidemiologic characteristics that can inform clinical care and etiological studies.

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 to 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 etiological 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)(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 sub-study were extracted from the overarching result set on January 29th 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 Figure S1. Eight databases from Spain and the United States (US) were included in this study.

Spanish data came from the SIDIAP database, a primary care database from Catalonia, a northeastern region in Spain.(12) Data from the US included Electronic Health Records (EHR) from the hospital setting: CU-AMC-HDC (Colorado), CUIMC (New York), Optum-

EHR (national),(13) STARR-OMOP (California), 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 2017-2018.

We included all patients (regardless of age) with at least one 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 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 2017-2018.(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 one 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 c*ohorts (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 pre-specified 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 reidentification 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 4-25% in patients *diagnosed*; and 11-40% in patients *hospitalized* (Supplementary Table S4). Additionally, 274,557 patients hospitalized with seasonal influenza in 2017-2018 were identified (lifetime cancer prevalence range: 18-39%).

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

slighty 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 Figure 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 ten 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 7 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 and STARR-OMOP, where NHL was the fifth most common. Leukemia, liver and lung cancer were also within the top ten 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 Figure S3).

Prior comorbidities

In both COVID-19 cohorts, the most common comorbidities were cardiometabolic conditions, which were more frequent in US databases (especially VA-OMOP) than in the Spanish SIDIAP database. For example, in the US the range of hypertension was 52-85% (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 Figure 1. Several comorbidities were more frequent among patients *hospitalized* compared to 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) (Figure 2).

Thirty-day outcomes

In the COVID-19 *diagnosed* cohort, hospitalization in the US databases ranged from 14% to 32% (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 8-41%) was higher than 30% in 3 out of 6 databases (IQVIA-OpenClaims, Optum-EHR, VA-OMOP). Sepsis (6-25%), cardiovascular disease events (7-21%) and AKI (10-16%) 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 US (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 Figure 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 COVID-19 and influenza patients, we did not observe consistent differences across databases (Supplementary Figure S5). The distribution of comorbidities was similar in both groups, with few exceptions (Figure 3A). For example, COPD was more common among patients with influenza in CU-AMC-HDC, Optum-EHR and VA-OMOP (Figure 4A). Aside from CUIMC, outcomes were slighty 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 41% (COVID-19) vs 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 COVID-19 patients compared to influenza patients in Optum-EHR and VA-OMOP: 6 vs 1% and 18% vs 6%, respectively (SMD>0.2) (Figures 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). In both COVID-19 cohorts, the most frequent cancer types were prostate cancer and breast cancer; hematological 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.* Occurence of death ranged from 2% to 14% among those *diagnosed* and from 6% to 26% among those *hospitalized*. When compared to 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 US, the lifetime cancer prevalence is 5%;(20) (data on the lifetime cancer prevalence in Spain is unavailable to our knowledge),(16) which is lower than our findings in COVID-19 patients (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 COVID-19 patients at hospital settings, with cancer prevalences of 6-11% in studies from Europe and the US.(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 US are prostate and breast cancer.(20) These cancer types were also those more frequent in our COVID-19 cohorts. However, hematological 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 US 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 hematological 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 hematological malignancies compared to 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 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 to studies describing cancer patients, 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).(22) In addition, heart disease, chronic kidney disease and type 2 diabetes were meaningfully higher among those *hospitalized* compared to 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 US,(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). This suggests that studies from the beginning of the pandemic, when testing was limited, might have overestimated mortality rates in COVID-19 patients, including those with cancer. For instance, a meta-analysis including studies prior to July 2020, with data over 18,000 cancer patients with COVID-19 (mostly inpatients), reported a pooled case mortality rate of 25.6% (95% CI: 22.0-29.5%), (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 COVID-19 patients are more often male, younger and less likely to have respiratory and cardiovascular diseases than influenza patients.(14) Interestingly, COVID-19 and influenza patients 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 one 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 vs patients without cancer. Additionally, 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 US). 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. While 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 COVID-19 patients 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.

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.

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 programme and EFPIA. This research received partial support from the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), US National Institutes of Health (NIH), US 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 number: HI16C0992]; 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 number CM20/00174). W.U.R. Ahmed reports funding from the NIHR Oxford Biomedical Research Centre (BRC), Aziz Foundation, Wolfson Foundation, and the Royal College 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. Prat-Uribe is supported by Fundacion Alfonso Martin Escudero and the Medical Research Council (grant numbers 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 programme, NIHR, Department of Veterans Affairs or the United States 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.

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. Comment on Williamson et al. (OpenSAFELY): The Table 2 Fallacy in a Study of COVID-19 Mortality Risk Factors. Epidemiology. 2021;32(1):e1–2.

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(1):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 García-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(1):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 M, jdposada. ohdsi-studies/Covid19CharacterizationCharybdis: Charybdis v1.1.1 - Publication Package. 2020 [cited 2020 Dec 3];Zenodo. 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. JNCI 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;395:1919–1926.

28. Jutzeler CR, Bourguignon L, Weis C V., 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]. World Heal. Organ. 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.

Tables

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	-	-	-	6.3	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	NA	NA	NA	NA	NA	NA	NA	NA	-	-	2.7	2.1	2.1	3.5	-	4.0

embolism events																
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.

Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; SIDIAP: Information System for Research in Primary Care; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs

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																
Rank	SIDIAP n=8,854		CU-AMC-HDC n=806		CUIMC n=1,433		HealthVerity n=4,857		IQVIA-Open Claims n=315,523		Optum-EHR n=22,996		STARR-OMOP n=821		VA-OMOP 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	5.2 (4.0-6.4)	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)	Thryoid	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

Rank	SIDIAP n=2,610		CU-AMC-HDC n=265		CUIMC n=561		HealthVerity n=797		IQVIA-Open Claims n=105,931		Optum-EHR n=5,806		STARR-OMOP n=244		VA-OMOP 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 records.

Abbreviations: COVID-19: Coronavirus disease 2019; SIDIAP: Information System for Research in Primary Care; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs; NHL: Non-Hodgkin's lymphoma; LOCP: Lip, oral cavity and pharynx

Figure legends

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).

Abbreviations: COVID-19: Coronavirus disease 2019; SIDIAP: Information System for Research in Primary Care; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs

Figure 2. Standardized mean differences 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 a |SMD| of 0.1. SMD calculated for comorbidities meeting the minimum count required (5 individuals) in each database and cohort.

Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center;

STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs; SMD: Standardized Mean Difference

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.
Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs

Figure 4. Standardized mean differences 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.

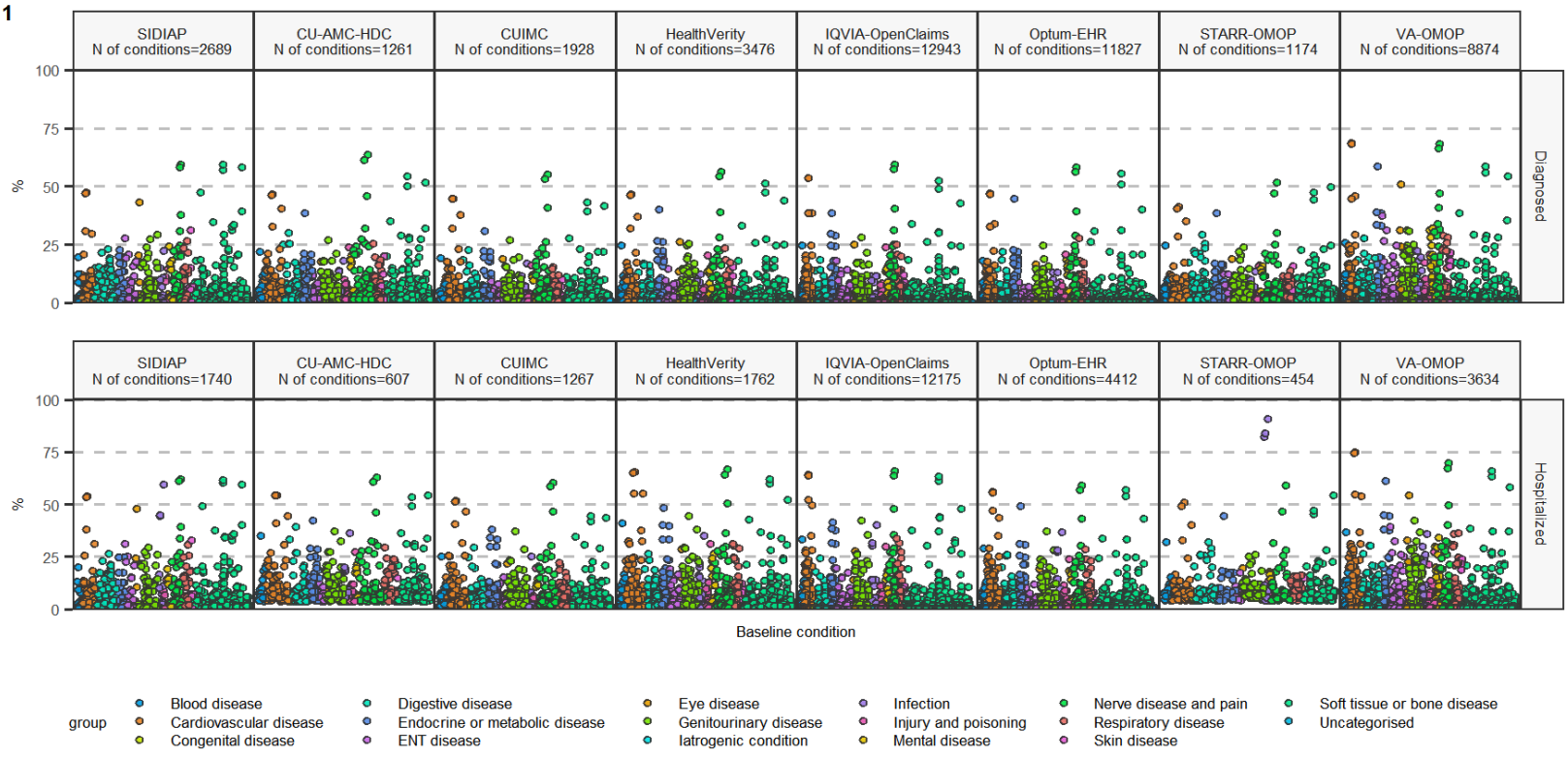
Notes: Comorbidities and outcomes ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Black-dotted lines indicate a |SMD| of 0.1. SMD calculated for comorbidities and outcomes meeting the minimum count required (5 individuals) in each database and cohort. Outcomes not shown due to data not available: all outcomes in CU-AMC-HDC, occurence of death in CUIMC and IQVIA-OpenClaims, intensive services in CUIMC.
Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; CU-AMC-HDC: Colorado University

Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: Stanford Medicine Research Data Repository; VA-OMOP: Department of Veterans Affairs; SMD: Standardized Mean Difference

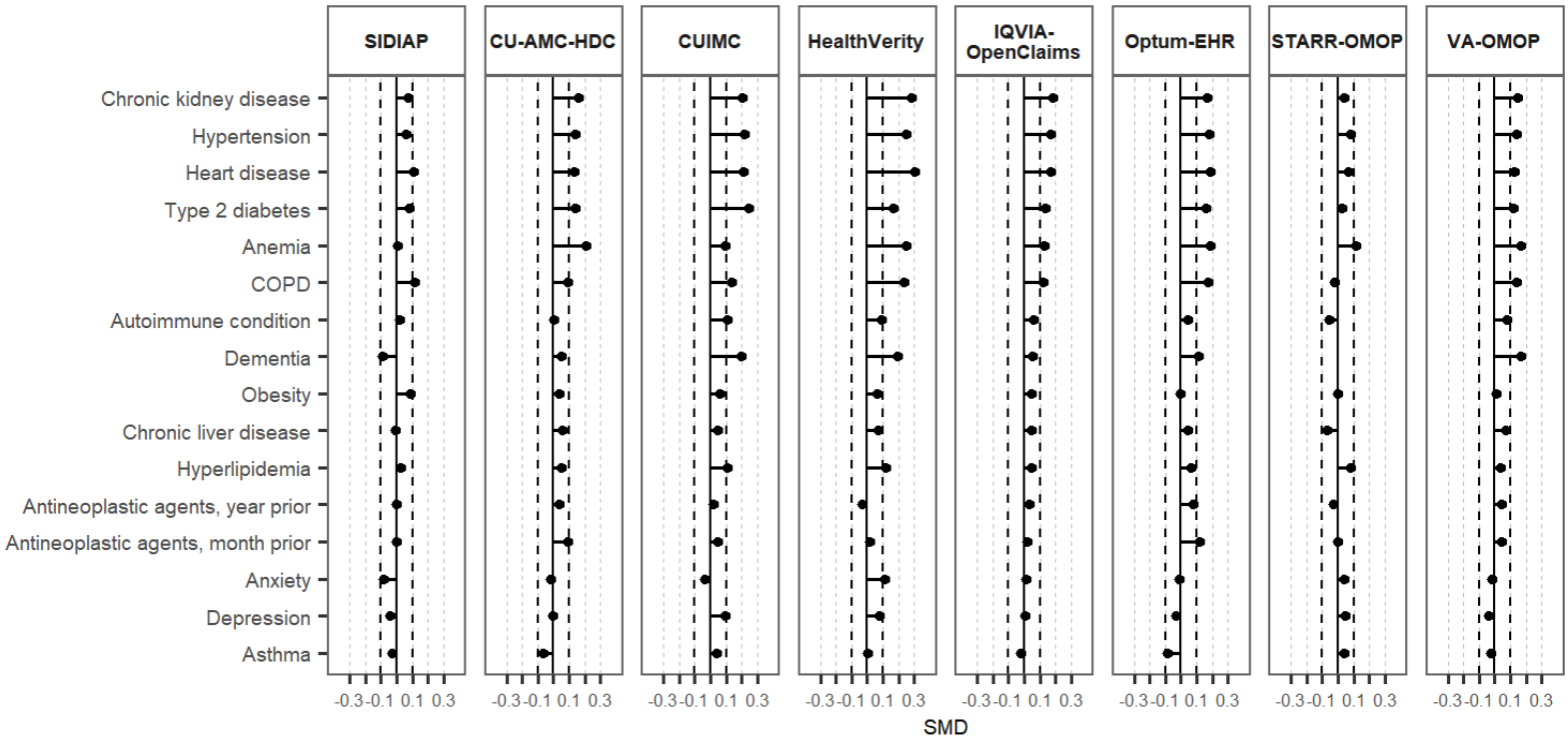
Authors' contributions

E. Roel, A Prats-Uribe, P.B. Ryan, D. Prieto-Alhambra, K. Kostka and T. Duarte-Salles conceived and designed the study. M. Aragón, C. Blacketer, W. Carter, F. DeFalco, S.L. DuVall, T. Falconer, G. Hripcsak, K.E. Lynch, M.E. Matheny, K.Natarajan, A. Ostropolets, J.D. Posada, C.G. Reich, L.M. Schilling, N.H. Shah, M. Spotniz, and T. Duarte-Salles coordinated data contributions at their respective sites. A. Pistillo, A.G. Sena, S. Fernández-Bertolín, J. D. Posada, K. Kostka and T. Duarte-Salles analyzed the data. E. Roel and A. Pistillo produced the figures and tables. E. Roel, M. Recalde, A. Golozar, L. Hester, D.R. Morales, F. Nyberg, I. Soerjomataram, K. Kostka and T. Duarte-Salles interpreted the data. E. Roel and T. Duarte-Salles searched the literature and wrote the first draft with insightful contributions from M. Recalde, D. Puente, A. Golozar, L. Hester, D.R. Morales, F. Nyberg, A Prats-Uribe, L.M. Schilling, I. Soerjomataram, V. Subbian, and K. Kostka. All authors contributed to the revision of the first draft, reviewed and approved the final version of the manuscript.

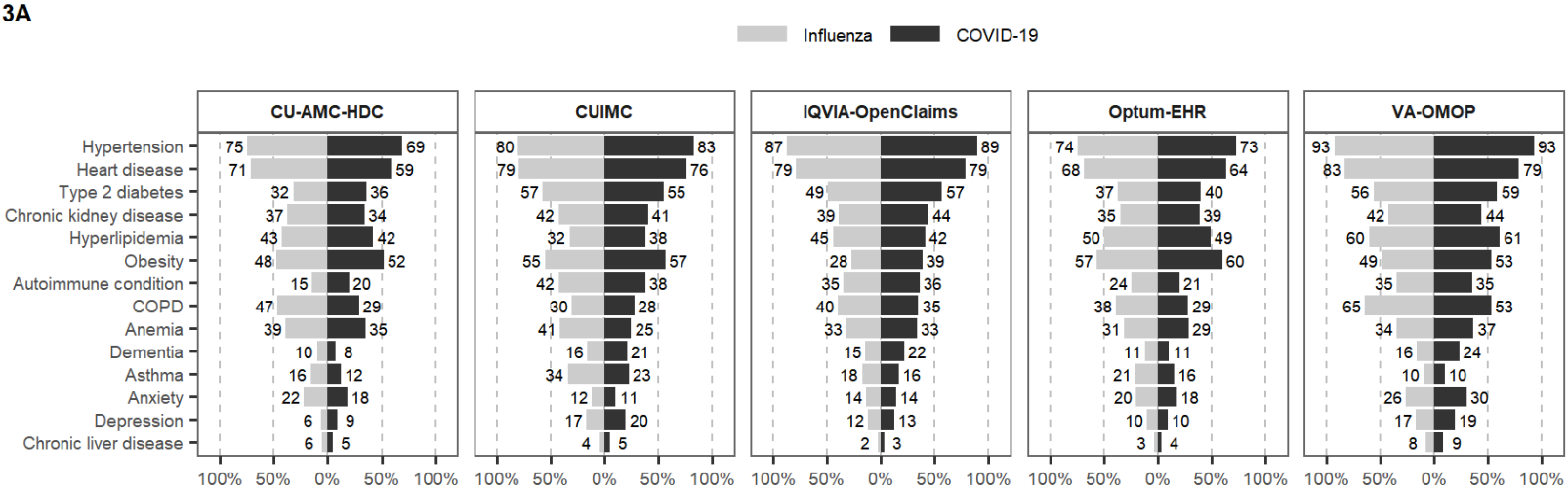
1



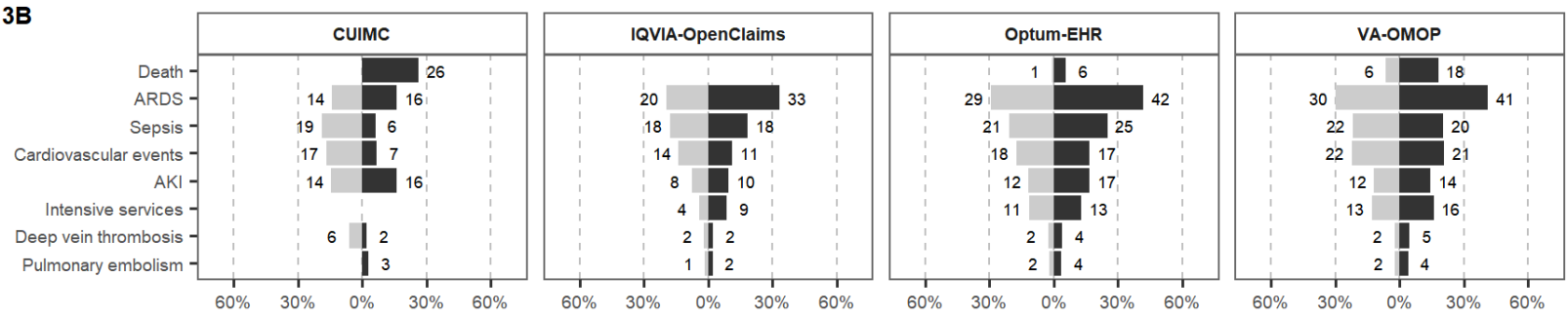
2



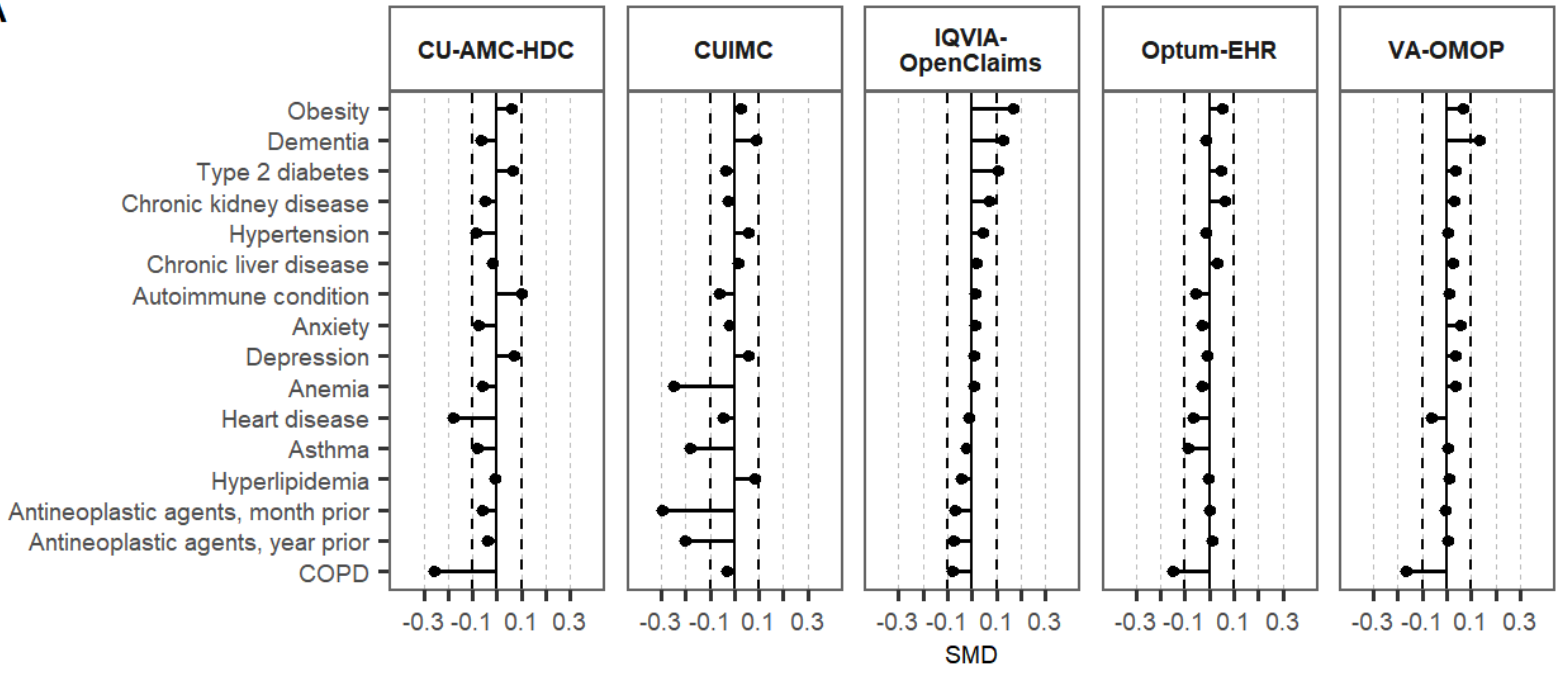
3A



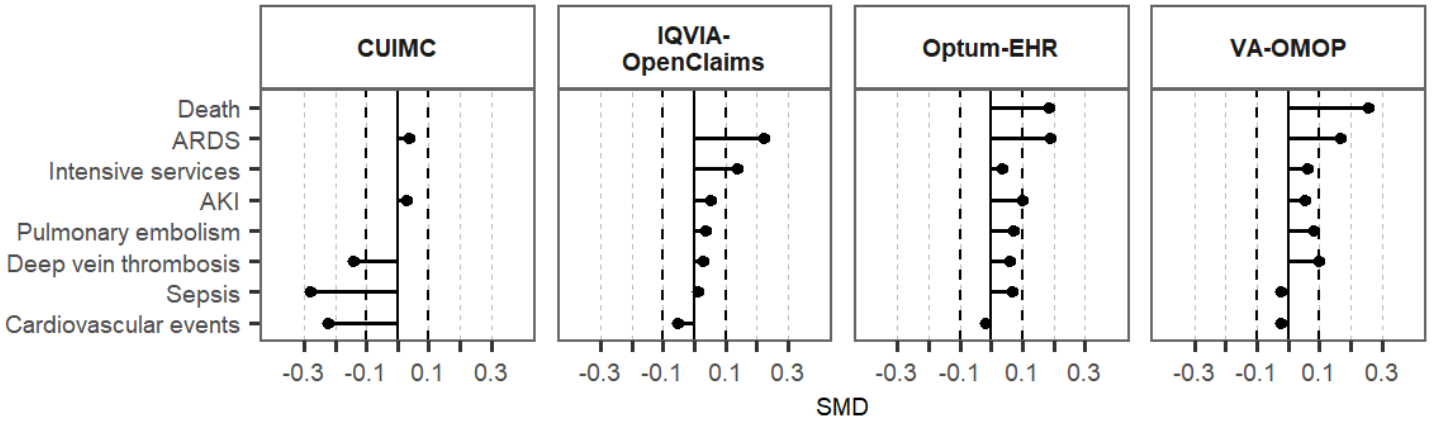
3B



4A



4B



Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Characteristics and outcomes of over 300,000 COVID-19 individuals with 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.