Characteristics and outcomes of 118,155 COVID-19 individuals with a 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], María 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 Rivera[35], Lisa M. Schilling[11], Isabelle Soerjomataram[36], Karishma Shah[5], Nigam 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 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

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

Running head: Characterization of 118,155 COVID-19 individuals with a history of cancer

Key Words: COVID-19; SARS-CoV-2; cancer; characterization; hospital admission, mortality

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; HAI: Healthcare-associated infection; 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.

Funding

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, 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. WURA reports funding from the NIHR Oxford Biomedical Research Centre (BRC), Aziz Foundation, Wolfson Foundation, and the Royal College Surgeons of England. LH, AGS are employed by Janssen Research and Development, LLC. JJ reports grants from the National Health and Medical Research Council (APP1192469). MG & YZ reports grants from National Key Research & Development Program of China (Project No. 2018YFC0116901). VS reports funding from the National Science Foundation, Agency for Health Care Research & Quality, National Institute of Health, and the State of Arizona. AP-U is supported by Fundacion Alfonso Martin Escudero and the Medical Research Council (grant numbers MR/K501256/1, MR/N013468/1). 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.

Competing interest statement

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 Centre (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. 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. Jonnagaddala reports grants from National Health and Medical Research Council, 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. 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.

Disclaimer

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

Abstract

Purpose

We aimed to describe the demographics, cancer subtypes, comorbidities and outcomes of patients with a history of cancer with COVID-19 from March to June 2020. 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 118,155 patients with a cancer history in the COVID-19 *diagnosed* and 41,939 in the COVID-19 *hospitalized* cohorts. The most frequent cancer subtypes were prostate and breast cancer (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 more frequently aged above 65 years and had multiple comorbidities. Occurrence of death ranged from 8% to 14% and from 18% to 26% in the *diagnosed* and h*ospitalized* COVID-19 cohorts,

respectively. Patients hospitalized with influenza (n=242,960) had a similar distribution of cancer subtypes, sex, age and comorbidities but lower occurrence of adverse events.

Conclusion

Patients with a history of cancer and COVID-19 have advanced age, multiple comorbidities, and a high occurence of COVID-19-related events. Additionaly, hematological malignancies were frequent in these patients.This observational study provides epidemiologic characteristics that can inform clinical care and future etiological studies.

Background

Shortly after the emergence of the novel 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 (HAI).[3] In addition, patients with cancer are often older and with 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, were conducted across different countries and geographic regions with different incidence of infection at varying time points and used different definitions for cancer (e.g., active cancer, solid 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 essential to inform public health strategies and clinical care, as well as to provide the groundwork for future etiological studies. 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 conducted a multi-national study aiming to describe 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.

Methods

Study design, setting and data sources

This 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 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 October, 9th, 2020.

From the databases that contributed to CHARYBDIS, we included those reporting data 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 and 30-days outcomes affecting 10% of the study population with a confidence interval (CI) width of +/- 5%. The selection process is depicted in Appendix Figure A1. Up to eight databases from two countries (Spain and the United States (US)) were included in this study. Spanish data came

from SIDIAP, a primary care database linked to hospital admissions.[12] Data from the US included Electronic Health Records (EHR) from the hospital setting: CU-AMC-HDC, CUIMC, Optum-EHR,[13] and VA-OMOP (93% male, mostly veterans); as well as claims data: HealthVerity and IQVIA-OpenClaims. All databases reported patients with COVID-19 identified from March to June 2020, apart from CU-AMC-HDC that included data to August 2020. For a description of the included data sources, see Appendix Table A1.

Study participants

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

To comprehensively capture baseline characteristics, we only included patients 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 the index date. Patients *diagnosed* with COVID-19 were those having a clinical diagnosis and/or a positive test for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) 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 test results or diagnoses. 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 a history of cancer and COVID-19 and influenza cases can be found in Appendix Table A2.

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. 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

More than 15,000 medical conditions from up to one year prior to index date were identified in each cohort 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 Appendix Table A2). 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 Spain and the US, according to data from the International Agency for Research on Cancer.[16] The SNOMED codes used to identify each cancer type can be found in Appendix Table A3.

We report here demographics (sex and age), antineoplastic and immunomodulating treatment received in the month and year prior to the index date and a selection of key comorbidities (based on their relevance to the COVID-19 field and on their prevalence in the cohorts).

The 30-day outcomes of interest in the COVID-19 *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.

Data characterization and analysis

Analysis was performed through a federated analysis approach.[15] Following a pre-specified analysis plan, a common analytical code for the whole CHARYBDIS study was developed by the OHDSI Methods library and run locally in each participating site (code available at zenodo.org).[17] Individual-level data remained within host institutions and only aggregate results were provided to the research team and publicly-shared. The results reported in this manuscript, and additional data, are available for consultation on a regularly updated website as new databases and/or results are added (https://data.ohdsi.org/Covid19CharacterizationCharybdis/).

Herein, we describe the lifetime cancer prevalence (recorded any time prior to the index date) and the number of patients included by database and cohort. Results are calculated and reported by cohort and database. Demographics, cancer types, comorbidities, and outcomes are reported as proportions (calculated as the number of persons within a given category, divided by the total number of persons), along with 95%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.

To compare characteristics between study cohorts, we calculated standardized mean differences (SMD), with an SMD >|0.1| indicating a meaningful difference in the prevalence of a given condition.[18] 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, the databases included in this study identified 906,691 patients *diagnosed* with COVID-19 and 196,018 patients *hospitalized* with COVID-19. The lifetime cancer prevalence ranged from 4% (HealthVerity) to 25% (STARR-OMOP) in all patients *diagnosed* with COVID-19; whereas it ranged from 11% (HealthVerity) to 40%(VA-OMOP) in all patients *hospitalized* with COVID-19 (Appendix Table A4). In addition, 242,960 patients hospitalized with seasonal influenza in 2017-2018 were identified. Among those, the lifetime cancer prevalence ranged from 21% (Optum-EHR) to 39% (VA-OMOP).

We included 118,155 patients *diagnosed* (Spain: 8,854; US: 109,301) and 41,939 patients *hospitalized* (Spain: 2,610; US: 39,329) with COVID-19 and history of cancer and 62,077 patients hospitalized (all from the US) with seasonal influenza and history of cancer.

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 Appendix Table A5). In the *diagnosed* cohort, patients were more commonly female (range 53% to 55%), aside from STARR-OMOP (47%) and VA-OMOP (7.1%). In contrast, in the *hospitalized cohort*, male

predominated in all databases (range 50% to 60%, VA-OMOP: 96%). Overall, patients were mainly aged above 65 years in both COVID-19 *diagnosed* and *hospitalized* cohorts and patients *hospitalized* were consistently older than those *diagnosed* (Appendix Figure A2). In the *diagnosed* cohort, the proportion of patients having received antineoplastic and immunomodulating agents the month and the year prior to index date ranged from 5% (HealthVerity) to 23% (STARR-OMOP) and from 14% (SIDIAP) to 35% (CU-AMC-DHC), respectively. Similarly, in the *hospitalized* cohort it ranged from 6% (HealthVerity) to 28% (CU-AMC-DHC) and from 14% (SIDIAP) to 38% (CU-AMC-DHC), respectively.

Frequency of cancer types

For both COVID-19 cohorts, the frequency of the 26 selected cancer types can be consulted in Appendix Table A6. The top ten cancer types by frequency are reported in Table 2. In the COVID-19 *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: 7.1%; 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 also common (among the ten most frequent in at least 7 out of 8 databases).

In the COVID-19 *hospitalized* cohort, prostate cancer was the most frequent cancer in all databases (ex aequo with NHL in CU-AMC-HDC, 6.4%), aside from Optum-EHR (breast cancer, 6.4%), in which it was the 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 (Appendix Figure A3).

Prior comorbidities

In both COVID-19 cohorts, the most common comorbidities were cardiometabolic conditions, which were more frequent in the US databases (especially VA-OMOP) than in the Spanish SIDIAP database. For example, hypertension ranged in the US from 52% to 87% (Spain: 32%) among *diagnosed* and from 58% to 93% (Spain: 33%) among *hospitalized* patients. Other common conditions include chronic obstructive pulmonary disease (COPD) (range in *diagnosed*: 12% to 43% (Spain: 34%), range in *hospitalized*: 11% to 53% (Spain: 42%)); and chronic kidney disease (range in *diagnosed*: 18% to 34% (Spain: 19%), range in *hospitalized*: 20% to 44% (Spain: 24%)). The prevalence of the full range of prior conditions summarized is shown in Figure 2. Several comorbidities were more frequent among patients *hospitalized* compared to patients *diagnosed* (SMD>0.1): heart disease (all databases except STARR-OMOP) and chronic kidney disease, hypertension and type 2 diabetes (all except SIDIAP and STARR-OMOP) (Figure 3).

Thirty-day outcomes

In the COVID-19 *diagnosed* cohort, hospitalization rates in the US databases ranged from 14% to 40% (Spain: 25%) and occurrence of death from 8% to 10% (Spain: 14%).

In the COVID-19 *hospitalized* cohort, outcomes were heterogeneous across databases, occurring less frequently in STARR-OMOP and more frequently in VA-OMOP. ARDS ranged from 8% to 41%, although it was higher than 30% in 3 out of 6 databases with data available (IQVIA-OpenClaims, Optum-EHR, VA-OMOP). Sepsis was also common, ranging from 6% to 25%. Cardiovascular disease events and AKI ranged from 7% to 21% and from

10% to 16%, respectively. Thromboembolic events were less frequent, with deep vein thrombosis ranging from 2% to 5% and pulmonary embolism from 2% to 4%. Intensive services requirement ranged from 6% to 16%, whereas occurrence of death ranged from 18% to 26% in the US (Spain: 21%).

Comparison of patients with a history of cancer hospitalized with COVID-19 to those hospitalized with seasonal influenza

The characteristics of patients with a history of cancer hospitalized with seasonal influenza are reported in Appendix Table A8. Overall, these patients were predominantly male and the majority clustered around the ages of 60 to 85 years old (Appendix Figure A5). When comparing the frequency of cancer types between COVID-19 and influenza patients, we did not observe consistent differences across databases (Appendix Figure A4). The distribution of comorbidities was similar in both groups, with few exceptions (Figure 2A). For example, COPD was more common among patients with influenza in 3 databases: CU-AMC-HDC, Optum-EHR and VA-OMOP (Figure 3A). Aside from CUIMC, outcomes were more common in the COVID-19 cohort, especially ARDS, ranging from 16% to 41% in the COVID-19 cohort and from 14% to 30% in the influenza cohort (SMD>0.2 in IQVIA-OpenClaims and Optum EHR, SMD>0.1 in VA-OMOP). Occurrence of death was higher among COVID-19 patients compared to influenza patients in VA-OMOP: 18% vs 6 (SMD>0.2) (Figures 2B and 3B).

Discussion

In this multinational cohort study, we described in detail the characteristics of 118,155 patients with a history of cancer and COVID-19, including outcomes that have been rarely

reported in this population. The lifetime cancer prevalence was higher among those *hospitalized* compared to all those *diagnosed* with COVID-19. In both COVID-19 cohorts of patients with history of cancer, the most frequent cancer type was either prostate or breast cancer; hematological malignancies were also frequent. Comorbidities were common in both cohorts but were higher among those *hospitalized.* Occurence of death ranged from 7% to 14% among those *diagnosed* and from 18% to 26% among those *hospitalized*. When compared to patients with a history of cancer hospitalized with seasonal influenza, patients hospitalized with COVID-19 had a similar distribution of sex, age and comorbidities but had more severe outcomes.

According to recent estimates, the lifetime cancer prevalence in the US is 5%.[19] In Spain, the 5-year cancer prevalence is 1.7% (data on the lifetime cancer prevalence is unavailable to our knowledge).[16] Although our results were heterogeneous, we found that the lifetime cancer prevalence was higher among COVID-19 cases compared to the general population, ranging from 4% to 25% in the *diagnosed* and from 11% to 40% in the *hospitalized* cohort. Even though comparisons are limited due to differences in the definition of cancer, these prevalences are also higher than prior reports on COVID-19 patients at hospital settings. In early studies from China, less than 3% of inpatients with COVID-19 had cancer, whereas studies from Europe and the US have reported a cancer prevalence of 6% to 11%.[20–25] A Danish study, however, found a lifetime cancer prevalence among patients hospitalized with COVID-19 of 17%, which is in line with our results.[6] Further, the lifetime cancer prevalence was consistently higher in patients *hospitalized* compared to those *diagnosed*, which might partially be due to the older age of the former.

The most lifetime-prevalent cancer types in the US population are prostate (in men) and breast (in women) cancer.[19] Two hematological malignancies are also frequent among cancer

survivors: NHL is the fifth/sixth more frequent (men and women, respectively), whereas leukemia is the ninth in men. We found that prostate and breast cancer were the most frequent cancers in COVID-19 patients, which is in line with the general cancer population. These cancer types were similarly common among *diagnosed* patients (where females predominated slightly), whereas in *hospitalized* patients (where males were the majority), prostate cancer was more frequent than breast cancer. On the other hand, hematological malignancies were more common than expected in our cohorts (both COVID-19 and influenza). For example, in the COVID-19 *hospitalized* cohorts, NHL, leukemia and multiple myeloma were among the third, fifth, and tenth more common cancers, respectively. Prior studies have reported an overrepresentation of hematological malignancies in COVID-19 patients compared to the general population.[5,26] This is of particular concern in light of evidence suggesting that patients with hematological malignancies have an increased risk of COVID-19 mortality compared to those with solid cancers.[5] The fact that hematological malignancies were more common than expected in both the *diagnosed* and the *hospitalized* COVID-19 cohorts raises questions on whether patients with these malignancies are more exposed or more vulnerable to SARS-CoV-2 infection, or both. Because of the underlying severity of their disease, patients with hematological malignancies are particularly at risk of HAI [27] and SARS-CoV-2 outbreaks have already been described in hematological units.[28]

As expected, patients with a history of cancer were older and had more comorbidities than overall COVID-19 cases. A meta-analysis comprising 12,149 COVID-19 cases (mostly hospitalized) found hypertension (23%), heart failure (20%) and diabetes (12%) were the most common comorbidities.[29] These numbers are substantially lower than what we have observed. Similarly, a study characterizing COVID-19 inpatients found a prevalence of heart disease of 27% in SIDIAP (vs 42% in this study), 29% in CUIMC (vs 77%) and 48% in VA-OMOP (vs 80%).[14] Compared to studies reporting comorbidities among patients with cancer,

we also found higher prevalences of comorbidities. For example, chronic kidney disease (range 20-44%), diabetes (23-59%) 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).[23] In addition, we observed that 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.[30] The role of 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 given the high comorbidity burden.

According to the World Health Organization, in June 2020 the case-fatality ratio among confirmed COVID-19 cases was 11.4% in Spain and 5.0% in the US,[31] which is lower than the all-cause mortality observed both in the *diagnosed* (Spain: 14%, US: 8-10%) and *hospitalized* (Spain: 21%, US: 18-26%) cohorts. A meta-analysis including 18,650 patients with cancer and COVID-19 (mostly from the hospital setting) reported a pooled case mortality rate of 25.6% (95% CI: 22.0-29.5%),[32] in line with our findings in the *hospitalized* cohort. Undoubtedly, increased age and underlying comorbidities play a substantial role in COVID-19 related mortality among these patients.

Finally, we have 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, we have observed that COVID-19 and influenza patients

with a history of cancer have a similar sex and age distribution and are of comparable health status. Despite this similarity, patients with cancer history and COVID-19 had worse outcomes than those with cancer history and influenza, including mortality, intensive services requirements, ARDS and thromboembolic events, which might be expected given the virulence of the disease.

This study has many major strengths, such as its large size. We have reported more than 10,000 aggregate characteristics from more than 100,000 patients from eight different databases and two countries. The diverse healthcare settings and populations described in this study, together with our multinational approach, increases the generalizability of our findings. Further, as CHARYBDIS is an ongoing study, we expect that more databases from a range of 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 prior to index date, we have comprehensively captured baseline comorbidities, which could explain the higher prevalence of selected comorbidities in our cohorts compared to prior studies. In addition, we have 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 several limitations. First, we were not able to provide detailed cancer information, such as the year of cancer diagnosis or stage of tumour at diagnosis; nor identify patients with active cancer treatment, and we were not able to characterize patients stratified by cancer types due to small sample sizes. Secondly, although we included patients with a clinical COVID-19 diagnosis to reduce selection bias due to testing restrictions during the first months of the pandemic, we cannot exclude that we have incurred some false positives. Additionally, we did not have information on the cause of death and reported all-cause death as an outcome. Third, 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 use of influenza vaccination among high-risk population groups likely contributed to the observed low proportion of adverse events among influenza patients. Finally, we found heterogeneous results across databases, which hinders the interpretation of our findings. Heterogeneity across data sources is a known phenomenon when using real-world data that reflects the existence of different coding practices, observation period, healthcare settings and populations. Indeed, the fact that the percentages of several cancer types were very low in some databases might be explained by differences in coding practices across data sources (i.e. some codes might be more widely used in some databases than others). Similarly, 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. While the interpretation of heterogeneous results is challenging, these also provide valuable insights into the particularities of each each setting, which is the objective of generalizability of real world data for certain populations. Yet, despite this heterogeneity, we found consistent patterns when comparing characteristics across cohorts, which lends credence to our results.

Conclusion

This in-depth characterization revealed that COVID-19 patients with a history of cancer are mostly aged above 65 years old and have a myriad of comorbidities that may explain the high frequency of severe COVID-19-outcomes observed 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.

References

1 Liang W, Guan W, Chen R, *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, *et al.* Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet (London, England)* 2020; **6736**: 1–13.

5 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**: 1309–16.

6 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. DOI:10.1093/ije/dyaa140.

7 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; : JCO.20.01580.

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**. DOI:10.1097/EDE.0000000000001259.

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**. DOI:10.1097/EDE.0000000000001258.

10 Voss EA, Makadia R, Matcho A, *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, *et al.* Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. In: Studies in Health Technology and Informatics. IOS Press, 2015: 574–8.

12 García-Gil M del M, Hermosilla E, Prieto-Alhambra D, *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* 2011; **19**: 135–45.

13 Datta S, Posada J, Olson G, *et al.* A new paradigm for accelerating clinical data science at Stanford Medicine. *arXiv* 2020; published online March 17. http://arxiv.org/abs/2003.10534 (accessed Aug 20, 2020).

14 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**. DOI:10.1038/s41467-020-18849-z.

15 Observational Health Data Sciences and Informatics. The Book of OHDSI. 2019. https://ohdsi.github.io/TheBookOfOhdsi/ (accessed Aug 20, 2020).

16 World Health Organization. Cancer Today. Int. Agency Res. 2018. https://gco.iarc.fr/today/home (accessed Oct 22, 2020).

17 Sena A, Kostka K, Schuemie M, jdposada. ohdsi-studies/Covid19CharacterizationCharybdis: Charybdis v1.1.1 - Publication Package. 2020; **Zenodo**. DOI:10.5281/zenodo.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 Miller KD, Nogueira L, Mariotto AB, *et al.* Cancer treatment and survivorship

statistics, 2019. *CA Cancer J Clin* 2019; **69**: 363–85.

20 Guan W, Ni Z, Hu Y, *et al*. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; **382**: 1708–20.

21 Zhang L, Zhu F, Xie L, *et al*. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Ann Oncol* 2020; **31**: 894–901.

22 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.

23 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.

24 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**. DOI:10.1136/bmj.m1985.

25 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**. DOI:10.1136/bmj.m1966.

26 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 (London, England)* 2020; **6736**. DOI:10.1016/S0140-6736(20)31173-9.

27 Huoi C, Vanhems P, Nicolle M-C, Michallet M, Bénet T. Incidence of Hospital-Acquired Pneumonia, Bacteraemia and Urinary Tract Infections in Patients with Haematological Malignancies. *PLoS One* 2004; **8**: 58121.

28 Biernat MM, Zińczuk A, Biernat P, *et al.* Nosocomial outbreak of SARS-CoV-2

infection in a haematological unit – High mortality rate in infected patients with haematologic malignancies. *J Clin Virol* 2020; **130**: 104574.

29 Jutzeler CR, Bourguignon L, Weis C V., *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.

30 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. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html (accessed Dec 3, 2020).

31 World Health Organization. Coronavirus disease (COVID-19): situation report, 162. 2020 https://apps.who.int/iris/handle/10665/332970 (accessed Oct 22, 2020).

32 Saini KS, Tagliamento M, Lambertini 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.

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.

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 research was approved by the Columbia University Institutional Review Board as an OHDSI network study. SIDIAP analyses were approved by the Clinical Research Ethics Committee of the IDIAPJGol (project code: 20/070-PCV). Approval for CPRD was provided by the Independent Scientific Advisory Committee (ISAC). For IQVIA-OpenClaims, approval is provided for OHDSI community studies. This is a retrospective database study on de-identified data and it is deemed not human subject research.

Transparency declaration

Lead authors affirm that the manuscript 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.

Contributorship statement

ER, APU, PR, DPA, KK and TDS conceived and designed the study. MA, CB, WC, FD, SLD, TF, GH, KL, MEM, KJ, AO, JP, CR, LMS, NS, MS and TDS coordinated data contributions at their respective sites. AP, AGS, SFB, JDP, KK and TDS analyzed the data. ER and AP produced the figures and tables. ER, MR, AG, LH, DRM, FN, IS, KK and TDS

interpreted the data. ER and TDS searched the literature and wrote the first draft with insightful contributions from MR, DP, AG, LH, DRM, FN, APU, LMS, IS, VS, KK and TDS. All authors contributed to the revision of the first draft, reviewed and approved the final version of the manuscript.

Data Sharing

Open Science is a guiding principle within OHDSI. As such, we provide unfettered access to all open-source analysis tools employed in this study via https://github.com/ohdsi-studies/Covid19CharacterizationCharybdis, as well as all data and results artefacts that do not include patient-level health information via https://data.ohdsi.org/Covid19CharacterizationCharybdis/. Data partners contributing to this study remain custodians of their individual patient-level health information and hold either IRB exemption or approval for participation.

Tables & Figures

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							
Characteristic, in %	**SIDIAP n=8,854**	**CU-AMC-HDC n=806**	**CUIMC n=1433**	**Health Verity n=4857**	**IQVIA-Open Claims n=72,994**	**Optum-EHR n=17,630**	**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=29,628**	**Optum-EHR n=4,451**	**STARR-OMOP n=244**	**VA-OMOP n=3,383**
Tested positive	51.7	-	67.1	61.0	-	37.6	11.3	52.0	76.0	-	89.8	48.9	-	47.2	39.7	32.3
Sex																
Female	53.9	53.0	54.7	53.7	53.2	59.0	47.0	7.1	39.7	46.8	46.3	50.1	46.8	50.5	41.8	3.6
Male	46.1	47.0	45.3	46.3	46.8	41.0	53.0	92.9	60.3	53.2	53.7	49.9	53.2	49.5	58.2	96.4
Antineoplastic and immunomodulating agents																
The month prior	10.9	22.5	12.3	5.4	9.5	6.1	18.8	12.0	10.8	28.3	14.6	5.9	10.6	11.3	18.9	14.2
The year prior	13.6	35.1	24.6	20.3	17.8	16.5	31.9	21.5	13.7	37.7	26.0	18.6	19.8	21.7	29.9	24.1
Comorbidities																
Asthma	4.7	15.4	20.5	9.9	18.5	20.3	14.3	10.8	3.9	12.1	22.8	10.0	17.0	16.4	16.4	9.9
COPD	33.7	23.7	19.8	17.3	28.1	18.1	11.9	43.3	41.5	29.4	28.2	31.4	33.7	28.0	11.1	53.2
Type 2 diabetes	17.8	26.9	38.0	31.7	52.2	30.2	22.0	50.3	22.5	36.2	55.1	43.0	59.4	40.7	23.8	58.7
Hyperlipidemia	22.8	38.7	30.8	40.1	40.0	44.4	38.7	58.8	24.6	42.3	38.0	48.4	43.1	48.7	44.3	61.2
Obesity	41.8	49.3	52.8	21.8	36.0	59.6	40.0	52.7	47.8	52.1	57.0	25.7	38.5	59.9	39.8	53.4
Heart disease	34.3	50.0	62.8	39.9	72.7	51.1	40.8	71.4	41.6	59.2	76.5	60.9	79.6	63.4	45.5	79.1
Hypertension	33.2	60.2	70.0	58.1	83.6	61.2	52.4	86.7	37.3	69.4	83.1	74.4	89.5	73.2	58.2	92.8
Anxiety	24.2	19.1	12.2	13.5	13.8	18.0	16.6	31.1	19.3	18.1	10.7	19.4	13.3	17.5	18.9	30.1
Dementia	10.7	5.8	11.0	7.1	21.4	6.8	1.7	14.3	7.2	7.5	21.2	15.8	21.8	11.1	-	23.7

Depression	7.2	9.2	14.8	2.7	12.0	10.5	10.4	21.1	5.7	9.1	20.0	4.9	11.6	9.2	12.7	19.1
Anemia	19.6	22.1	19.3	24.8	29.0	18.8	24.7	25.8	20.0	35.1	25.0	40.9	35.5	30.0	32.0	36.7
Autoimmune condition	10.4	19.9	30.8	13.0	35.8	18.4	14.6	30.3	11.2	20.0	38.1	17.8	39.0	20.6	11.9	35.4
Chronic kidney disease	19.4	23.8	27.4	18.4	34.1	29.3	18.0	34.1	23.6	34.0	40.8	36.0	43.8	40.8	20.1	44.3
Chronic liver disease	1.9	3.2	3.6	2.2	2.2	2.5	7.3	6.1	1.8	4.9	5.0	4.0	3.0	3.7	4.9	8.8
Outcomes																
Death	14.4	-	10.4	-	-	-	-	7.6	21.4	-	26.2	-	-	-	-	18.1
Hospitalization	24.9	-	34.8	13.5	39.3	24.1	27.0	27.1	NA	NA	NA	NA	NA	NA	NA	NA
Intensive services requirement	-	-	-	-	-	-	-	-	-	-	-	6.3	9.7	13.6	5.7	16.0
ARDS during hospitalization	-	-	-	-	-	-	-	-	-	-	15.9	26.5	31.2	40.2	8.2	41.2
Cardiovascular disease events	-	-	-	-	-	-	-	-	-	-	6.8	10.8	8.7	16.7	8.2	20.8
Deep vein thrombosis events	-	-	-	-	-	-	-	-	-	-	2.1	3.0	2.1	4.4	-	4.8
Pulmonary embolism events	-	-	-	-	-	-	-	-	-	-	2.7	2.1	1.6	3.6	-	4.0
Acute kidney injury during hospitalization	-	-	-	-	-	-	-	-	-	-	16.0	11.4	10.2	16.4	11.9	14.4
Sepsis during hospitalization	-	-	-	-	-	-	-	-	-	-	6.1	17.7	17.1	24.5	9.8	20.5

Note: - 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-OpenClaims n=72,994**		**Optum-EHR n=17,630**		**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	7.1 (6.9-7.3)	Breast	6.7 (6.3-7.1)	Breast	12.3 (10.0-14.6)	Prostate	18.1 (17.4-18.8)
2	Colorrectal	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	6.0 (5.8-6.2)	Prostate	4.9 (4.6-5.2)	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	3.5 (3.4-3.6)	Uterus	3.7 (3.4-4.0)	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)	Colorrectal	3.9 (3.4-4.4)	Lung	2.7 (2.6-2.8)	Lip, oral cavity and pharynx	3.7 (3.4-4.0)	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)	Leukemia	2.6 (2.5-2.7)	NHL	2.6 (2.4-2.8)	NHL	5.8 (4.2-7.4)	Lip, oral cavity and pharynx	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)	Colorrectal	2.6 (2.5-2.7)	Leukemia	2.3 (2.1-2.5)	Thyroid	5.2 (3.7-6.7)	Leukemia	2.6 (2.3-2.9)
7	Leukemia	2.7 (2.4-3.0)	Colorrectal	2.9 (1.7-4.1)	Multiple myeloma	3.1 (2.2-4.0)	Lung	2.7 (2.2-3.2)	Bladder	1.7 (1.6-1.8)	Lung	2.2 (2.0-2.4)	Lip, oral cavity and pharynx	4.9 (3.4-6.4)	Colorrectal	2.6 (2.3-2.9)

8	Uterus	2.5 (2.2-2.8)	Multiple myeloma	2.5 (1.4-3.6)	Colorrectal	2.7 (1.9-3.5)	Bladder	2.5 (2.1-2.9)	Multiple myeloma	1.6 (1.5-1.7)	Colorrectal	1.8 (1.6-2.0)	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)	Liver	1.5 (1.4-1.6)	Thyroid	1.5 (1.3-1.7)	Colorrectal	3.7 (2.4-5.0)	Liver	1.8 (1.5-2.1)
10	Lip, oral cavity and pharynx	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	1.4 (1.3-1.5)	Bladder	1.4 (1.2-1.6)	Bladder	3.2 (2.0-4.4)	Larynx	1.3 (1.1-1.5)

Patients with a history of cancer *hospitalized* with COVID-19

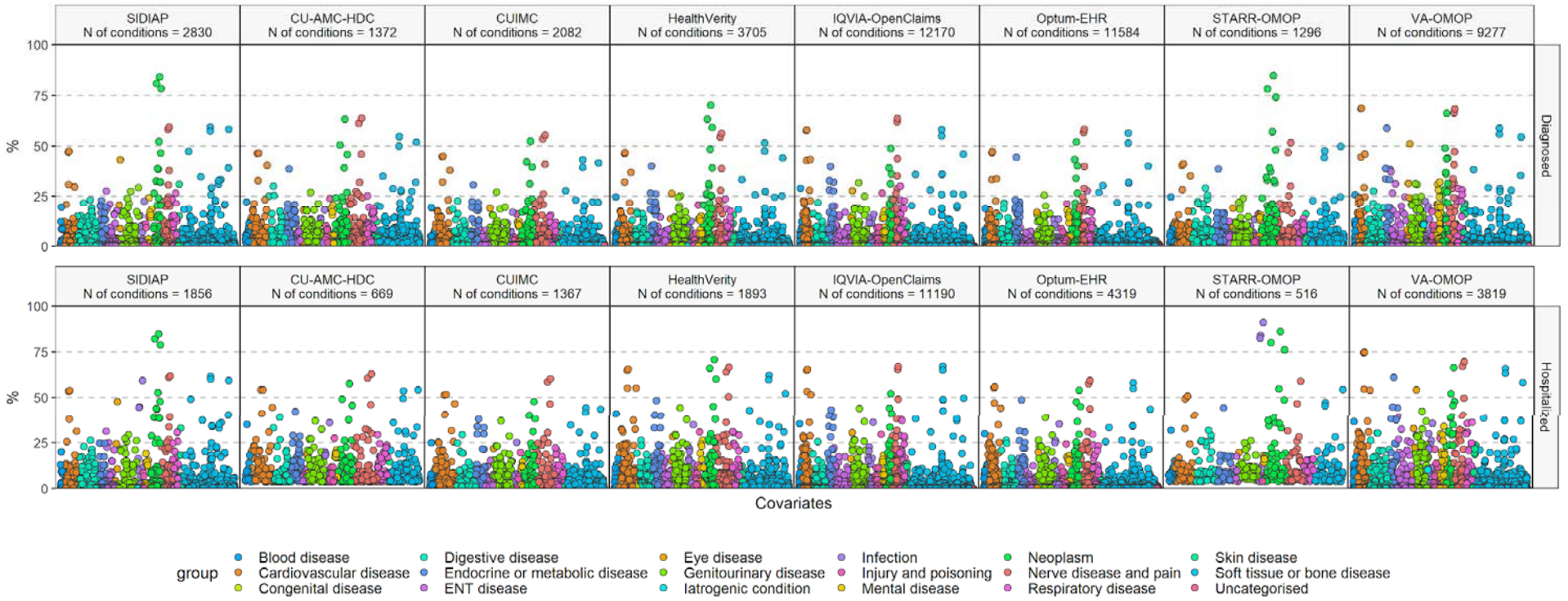
Rank	**SIDIAP n=2,610**		**CU-AMC-HDC n=265**		**CUIMC n=561**		**HealthVerity n=797**		**IQVIA-OpenClaims n=29,628**		**Optum-EHR n=4,451**		**STARR-OMOP n=244**		**VA-OMOP n=3,383**	
	Cancer	**% (95%CI)**	**Cancer**	**% (95%CI)**	**Cancer**	**% (95%CI)**	**Cancer**	**% (95%CI)**	**Cancer**	**% (95%CI)**	**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	8.5 (8.2-8.8)	Breast	6.4 (5.7-7.1)	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	5.4 (5.1-5.7)	Prostate	6.0 (5.3-6.7)	Lip, oral cavity and pharynx	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	4.5 (4.3-4.7)	NHL	4.2 (3.6-4.8)	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	4.0 (3.8-4.2)	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)	Leukemia	3.5 (3.3-3.7)	Lip, oral cavity and pharynx	3.5 (3.0-4.0)	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)	Colorectal	3.0 (2.8-3.2)	Leukemia	3.4 (2.9-3.9)	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	2.2 (2.0-2.4)	Uterus	3.0 (2.5-3.5)	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	2.2 (2.0-2.4)	Liver	2.7 (2.2-3.2)	Leukemia	4.9 (2.2-7.6)	Lip, oral cavity and pharynx	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	2.0 (1.8-2.2)	Colorectal	2.5 (2.0-3.0)	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)	Lip, oral cavity and pharynx	2.5 (1.4-3.6)	Kidney	1.7 (1.6-1.8)	Bladder	2.4 (2.0-2.8)	-		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

Figure 1. Prevalence of baseline conditions among patients with a history of cancer *diagnosed* and *hospitalized* with COVID-19.
Each dot represents one of these conditions with the colour indicating the type of condition.

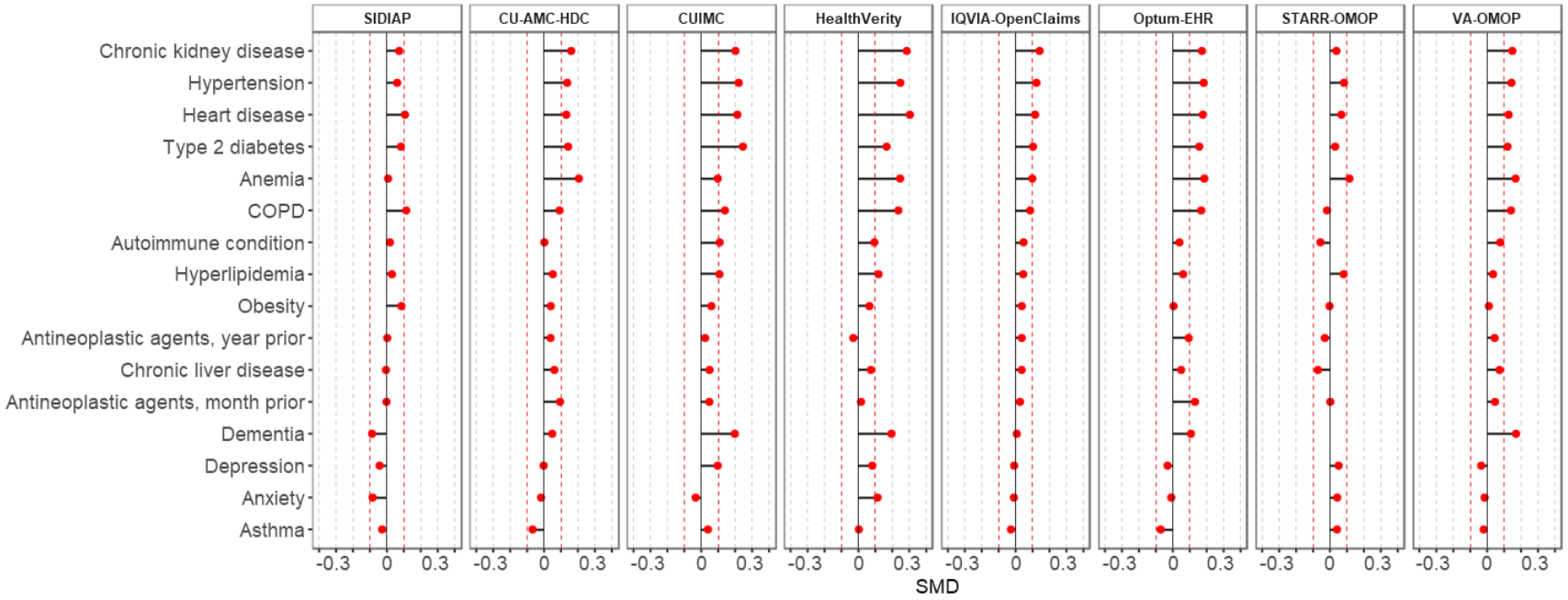


Notes: only conditions meeting the minimum count requirement (5 individuals) are shown.

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 conditions between patients with a history of 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*.

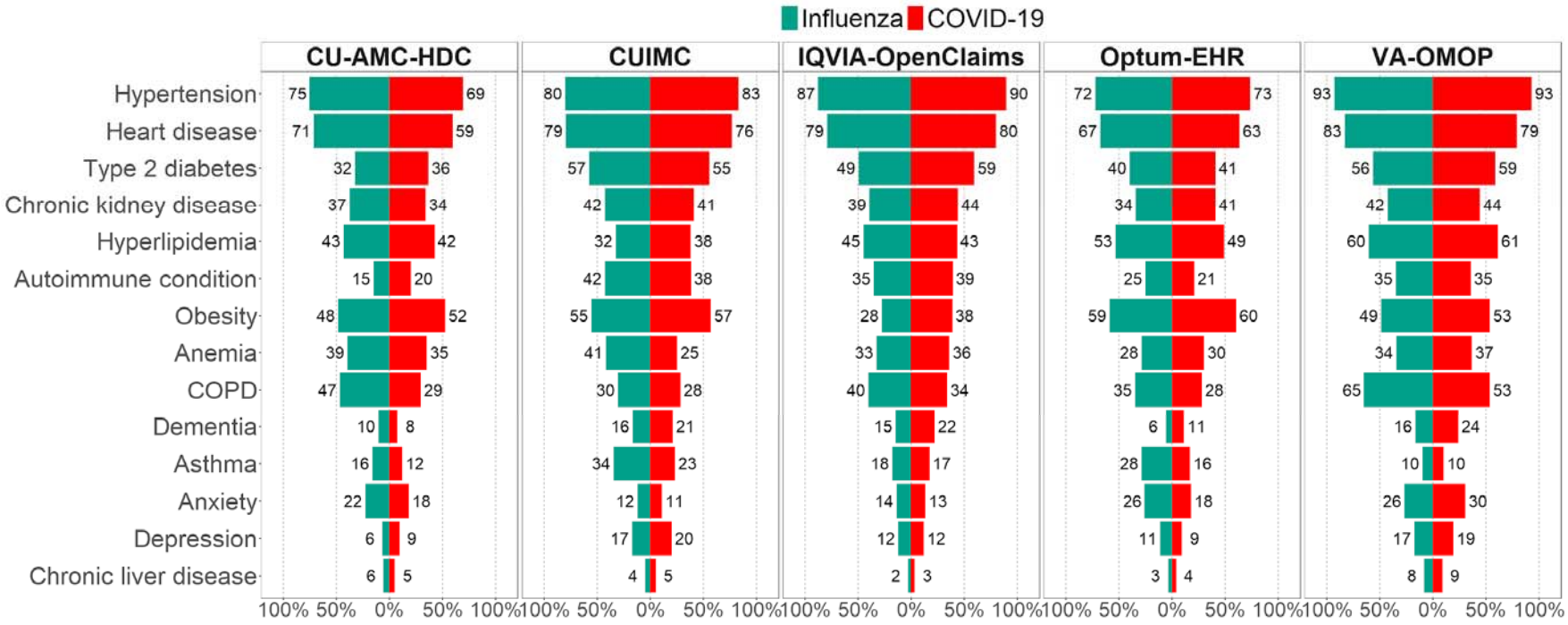


Notes: SMD were calculated for cancer types meeting the minimum count required (5 individuals) in each database and cohort. Cancer types ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Red-dotted lines indicate a SMD of |0.1|.

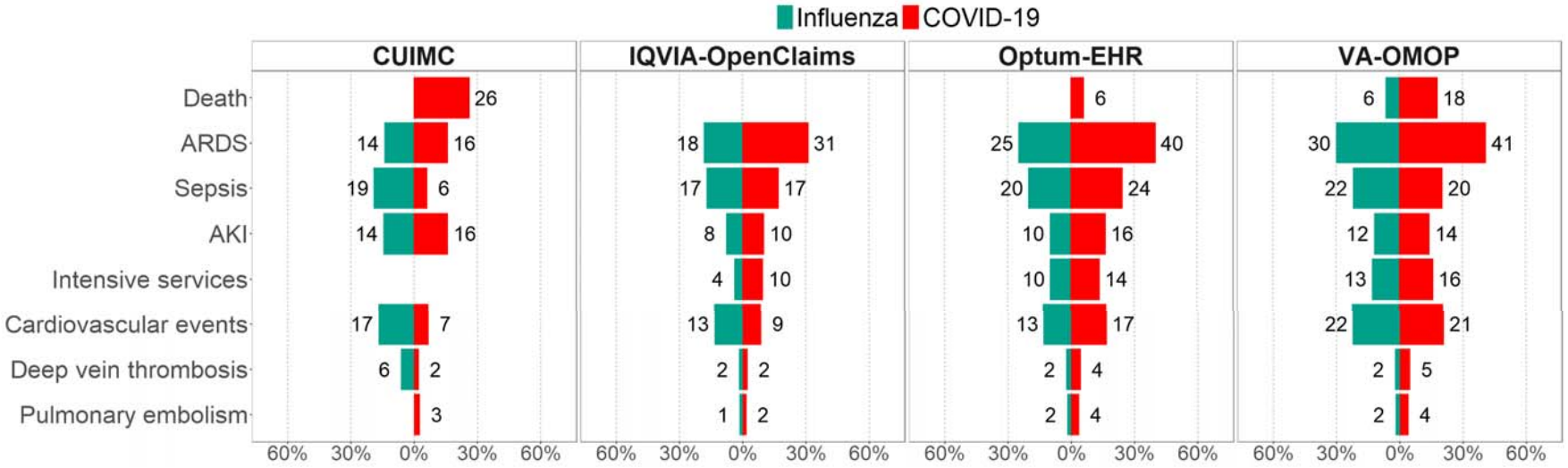
Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: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

Figure 3. Baseline comorbidities and 30-day outcomes among patients with a history of cancer *hospitalized* with COVID-19 and with seasonal influenza

A. Proportion of selected baseline comorbidities, in %



B. Proportion of 30-day outcomes, in %

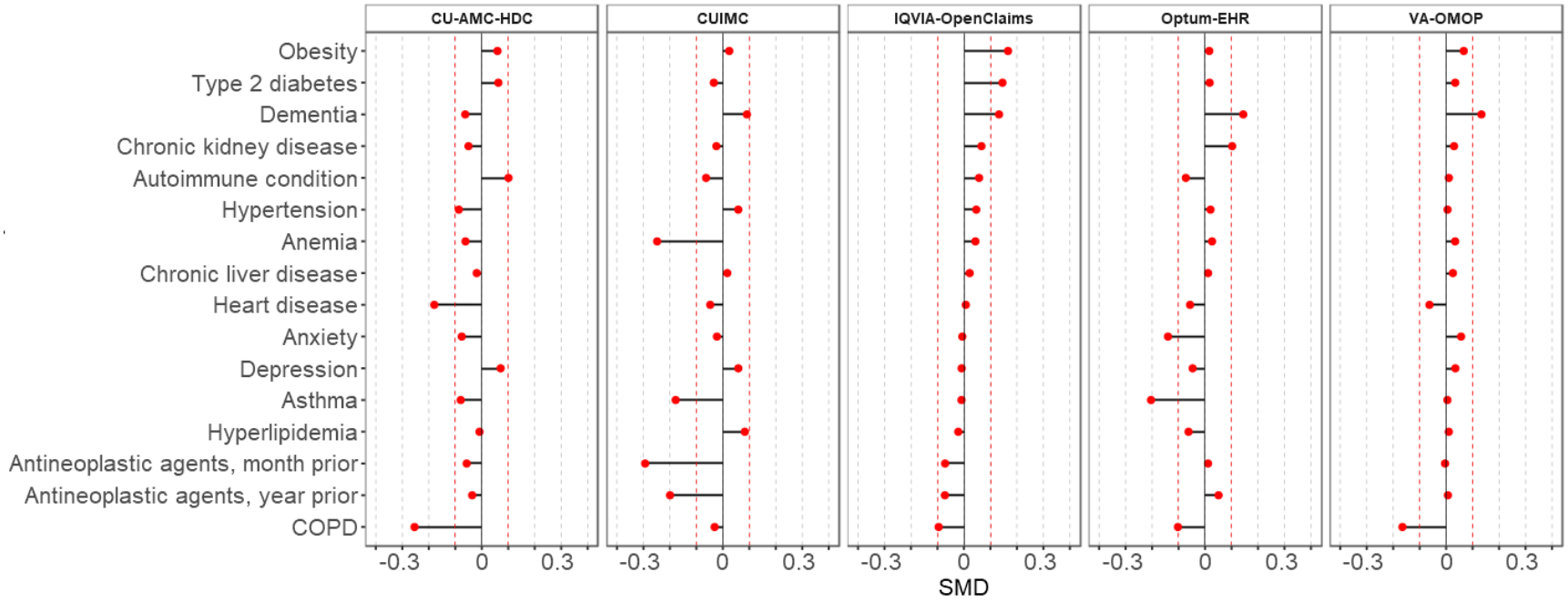


Notes: Comorbidities and outcomes are shown if meeting the minimum count required (5 individuals) in each database and cohort. Comorbidities and outcomes ordered according to descending values in the largest database (IQVIA-OpenClaims).

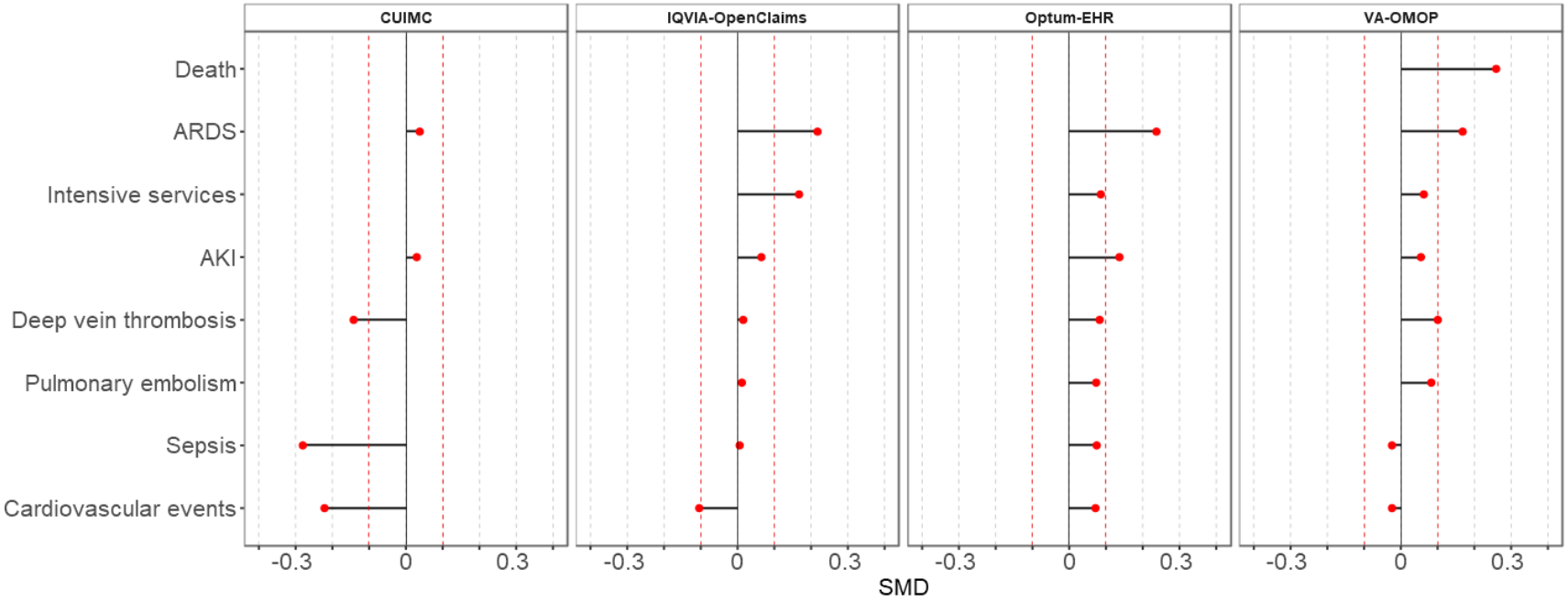
Abbreviations: COVID-19: Coronavirus disease 2019; COPD: Chronic Obstructive Pulmonary Disease; ARDS: Acute Respiratory Distress Syndrome; AKI: Acute Kidney Injury; CU-AMC-HDC: Colorado University Anschutz Medical Campus Health Data Compass; CUIMC: Columbia University Irving Medical Center; STARR-OMOP: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

Figure 4. Standardized mean differences of selected baseline conditions and 30-day outcomes 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.

A. Baseline conditions



B. 30 day-outcomes



Notes: SMD were calculated for conditions and outcomes meeting the minimum count required (5 individuals) in each database and cohort..Comorbidities and outcomes ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Red-dotted lines indicate a SMD of |0.1|.

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 legends

Figure 1. Prevalence of baseline conditions among patients with a history of cancer diagnosed and hospitalized with COVID-19.

Each dot represents one of these conditions with the colour indicating the type of condition.

Notes: only conditions meeting the minimum count requirement (5 individuals) are shown.

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 conditions 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: SMD were calculated for cancer types meeting the minimum count required (5 individuals) in each database and cohort. Cancer types ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Red-dotted lines indicate a SMD of |0.1|.

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 and 30-day outcomes among patients with history of cancer hospitalized with COVID-19 and with seasonal influenza

A. Proportion of selected baseline comorbidities, in %

B. **Proportion of 30-day outcomes, in %**

Notes: Comorbidities and outcomes are shown if meeting the minimum count required (5 individuals) in each database and cohort. Comorbidities and outcomes ordered according to descending values in the largest database (IQVIA-OpenClaims).

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 conditions and 30-day outcomes 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.

A. Baseline conditions

B. 30 day-outcomes

Notes: SMD were calculated for conditions and outcomes meeting the minimum count required (5 individuals) in each database and cohort..Comorbidities and outcomes ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Red-dotted lines indicate a SMD of |0.1|.

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