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

Elena Roel1,2, Andrea Pistillo1, Martina Recalde1,2, Anthony G. Sena3,4, Sergio Fernández-Bertolín1, Maria Aragón1, Diana Puente1,2, Waheed-Ul-Rahman Ahmed5,6, Heba Alghoul7, Osaïd Alser8, Thamir M. Alshammar9, Carlos Areia10, Clai Blacketer4, William Carter1, Paula Casajuast1,2, Aedín C. Culhane11,12, Dalia Dawoud13, Frank DeFalco14, Scott L. DuVall15,16, Thomas Falconer17,18, Asieh Golozar19,20, Mengchu Gong21, Laura Hester22, George Hripcsak18,19, Eng Hooi Tan24, Hokyun Jeon25, Jitendra Jonnagaddala26, Lana Y.H. Lai27, Kristine E. Lynch15,17, Michael E. Matheny28,29, Daniel R. Morales30,31, Karthik Natarajan18,19, Daniel R. Morales30,31, Karthik Natarajan18,19, Fredrik Nyberg32, Anna Ostropolets18, José D. Posada13, Albert Prats-Urbié14, Christian G. Reich15, Donna R. Rivera13, Lisa M. Schilling1, Isabelle Soerjomataram16, Karishma Shah5, Nathan H. Shah33, Yang Shen22, Matthew Spotnitz18, Vignesh Subbian7, Marc A. Suchard58, Annalisa Trama39, Lin Zhang40,41, Ying Zhang22, Patrick B. Ryan3,18, Karthik Natarajan18,19, Fredrik Nyberg32, Anna Ostropolets18, Jos
Introduction

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

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

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

Materials and Methods

Study design and setting

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

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

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

Study participants

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

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

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).
diagnoses or test results. Similarly, patients hospitalized with seasonal influenza were those who had a hospitalization episode and a influenza clinical diagnosis or positive test result for influenza in 2017 to 2018 (14). The criteria to define patients with cancer history and COVID-19 and influenza cases can be found in Supplementary Table S2.

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

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

### Patient characteristics and outcomes

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

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

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

### Analysis

Analysis was performed through a federated analysis approach (15). Following a prespecified analysis plan, an analytical code for the whole CHARYBDIS study was developed and run locally in each site (code available at zenodo.org) (17). Individual-level data remained within each database, and 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/Covid19Characterization/Charybdis/).

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

To compare characteristics between study cohorts, we calculated standardized mean differences (SMD). SMD are independent of sample sizes and can be used to compare the prevalence of dichotomous variables between two groups. An SMD $>$ 0.1 indicates a meaningful difference in the prevalence of a given condition (18, 19).

As this study was designed as a detailed descriptive study, statistical modelling was out of scope in the developed analytical packages. Therefore, differences across the groups compared should not be interpreted as causal effects.

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

### Results

#### Lifetime cancer prevalence

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

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

#### Demographics

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

#### Cancer types

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

In the hospitalized cohort, prostate cancer was the most frequent cancer in all databases (equally with NHL in CU-AMC-HDC, 6.4%); aside from Optum-EHR (second most frequent). NHL was among the three most frequent cancers in all databases aside from SIDIAP (fifth...
Table 1. Demographics, comorbidities, and outcomes among patients with a history of cancer diagnosed and hospitalized with COVID-19.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>SIDIAP</th>
<th>CU-AMC-HDC</th>
<th>CUIMC</th>
<th>HealthVerity</th>
<th>IQVIA-OpenClaims</th>
<th>STARR-OMOP</th>
<th>VA-OMOP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 8,854</td>
<td>n = 806</td>
<td>n = 1,433</td>
<td>n = 4,857</td>
<td>n = 315,523</td>
<td>n = 22,996</td>
<td>n = 107,60</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>53.9</td>
<td>53.0</td>
<td>54.7</td>
<td>53.7</td>
<td>55.1</td>
<td>58.6</td>
<td>47.0</td>
</tr>
<tr>
<td>Male</td>
<td>46.1</td>
<td>47.0</td>
<td>45.3</td>
<td>46.3</td>
<td>44.9</td>
<td>41.4</td>
<td>53.0</td>
</tr>
<tr>
<td><strong>Race, African American</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Antineoplastic and immunomodulating agents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>4.7</td>
<td>15.4</td>
<td>20.5</td>
<td>9.9</td>
<td>17.2</td>
<td>20.3</td>
<td>14.3</td>
</tr>
<tr>
<td>COPD</td>
<td>33.7</td>
<td>23.7</td>
<td>19.8</td>
<td>17.3</td>
<td>26.8</td>
<td>18.4</td>
<td>11.9</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>17.8</td>
<td>26.9</td>
<td>38.0</td>
<td>31.7</td>
<td>47.5</td>
<td>29.9</td>
<td>22.0</td>
</tr>
<tr>
<td>Hypertension</td>
<td>22.8</td>
<td>38.7</td>
<td>30.8</td>
<td>40.1</td>
<td>38.4</td>
<td>44.7</td>
<td>38.7</td>
</tr>
<tr>
<td>Obesity</td>
<td>418</td>
<td>49.3</td>
<td>52.8</td>
<td>21.8</td>
<td>35.5</td>
<td>60.4</td>
<td>40.0</td>
</tr>
<tr>
<td>Heart disease</td>
<td>34.3</td>
<td>50.0</td>
<td>62.8</td>
<td>39.9</td>
<td>68.1</td>
<td>51.1</td>
<td>40.8</td>
</tr>
<tr>
<td>Hypertension</td>
<td>37.2</td>
<td>60.2</td>
<td>70.0</td>
<td>58.1</td>
<td>80.8</td>
<td>61.3</td>
<td>52.4</td>
</tr>
<tr>
<td>Anemia</td>
<td>24.2</td>
<td>19.1</td>
<td>12.2</td>
<td>11.5</td>
<td>13.5</td>
<td>18.6</td>
<td>16.6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>10.7</td>
<td>5.8</td>
<td>11.0</td>
<td>7.1</td>
<td>18.3</td>
<td>6.3</td>
<td>14.3</td>
</tr>
<tr>
<td>Depression</td>
<td>7.2</td>
<td>9.2</td>
<td>14.8</td>
<td>2.7</td>
<td>12.3</td>
<td>10.9</td>
<td>10.4</td>
</tr>
<tr>
<td>Anemia</td>
<td>19.6</td>
<td>22.1</td>
<td>19.3</td>
<td>24.8</td>
<td>24.8</td>
<td>38.0</td>
<td>24.7</td>
</tr>
<tr>
<td>Autoimmune condition</td>
<td>10.4</td>
<td>9.9</td>
<td>30.8</td>
<td>13.0</td>
<td>32.0</td>
<td>18.3</td>
<td>14.6</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>19.4</td>
<td>23.8</td>
<td>27.4</td>
<td>38.4</td>
<td>37.7</td>
<td>27.9</td>
<td>18.0</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>19</td>
<td>3.2</td>
<td>36.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.0</td>
<td>2.4</td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>14.4</td>
<td>10.4</td>
<td>17</td>
<td>1.7</td>
<td>7.6</td>
<td>21.4</td>
<td>26.2</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>24.9</td>
<td>13.5</td>
<td>32.1</td>
<td>24.1</td>
<td>27.0</td>
<td>27.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Intensive services requirement</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>AIDS during hospitalization</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Cardiovascular disease events</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Pulmonary embolism events</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Acute kidney injury during hospitalization</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Sepsis during hospitalization</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Notes: — indicates data not available or below the minimum cell count required (5 individuals); NA indicates not applicable.
Patients with a history of cancer diagnosed with COVID-19

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breast 14.2 (13.5-14.9)</td>
<td>Breast 7.1 (5.9-9.1)</td>
<td>Prostate 8.7 (5.5-6.5)</td>
<td>Lung 6.6 (4.9-7.3)</td>
<td>Prostate 6.4 (6.3-6.5)</td>
<td>Breast 6.1 (6.4-6.7)</td>
<td>Prostate 5.0 (4.7-5.3)</td>
<td>Lung 7.7 (5.9-9.5)</td>
</tr>
<tr>
<td>2</td>
<td>Breast 10.4 (9.8-11.0)</td>
<td>Prostate 6.8 (5.4-8.6)</td>
<td>NHL 5.4 (4.2-6.6)</td>
<td>Breast 11.5 (10.6-12.4)</td>
<td>Breast 6.2 (6.3-6.3)</td>
<td>Prostate 5.0 (6.1-6.3)</td>
<td>Prostate 10.2 (3.0-12.3)</td>
<td>Lung 3.9 (3.5-4.3)</td>
</tr>
<tr>
<td>3</td>
<td>Prostate 9.4 (8.8-10.0)</td>
<td>NHL 5.0 (4.5-6.5)</td>
<td>Breast 3.2 (4.6-6.4)</td>
<td>NHL 4.6 (4.0-5.2)</td>
<td>NHL 3.1 (3.0-3.2)</td>
<td>NHL 3.0 (3.0-3.1)</td>
<td>Lung 4.2 (3.5-4.9)</td>
<td>NHL 3.8 (3.4-4.2)</td>
</tr>
<tr>
<td>4</td>
<td>Bladder 6.4 (5.9-6.9)</td>
<td>Lung 3.6 (2.5-5.5)</td>
<td>Leukemia 4.4 (3.3-5.5)</td>
<td>Colorectal 3.9 (3.4-4.4)</td>
<td>Colorectal 2.5 (2.4-2.6)</td>
<td>NHL 2.7 (2.5-2.9)</td>
<td>NHL 5.8 (4.2-7.4)</td>
<td>NHL 2.9 (2.6-3.2)</td>
</tr>
<tr>
<td>5</td>
<td>NHL 3.0 (2.6-3.4)</td>
<td>Leukemia 3.3 (2.1-4.5)</td>
<td>Thyroid 3.5 (3.0-4.0)</td>
<td>Lung 2.5 (2.4-2.6)</td>
<td>NHL 2.9 (2.6-3.2)</td>
<td>NHL 2.7 (2.5-2.9)</td>
<td>NHL 5.8 (4.2-7.4)</td>
<td>NHL 2.9 (2.6-3.2)</td>
</tr>
<tr>
<td>6</td>
<td>Melanoma 2.9 (2.6-3.2)</td>
<td>Melanoma 3.0 (1.8-4.2)</td>
<td>Lung 3.1 (2.2-4.0)</td>
<td>Leukemia 2.8 (2.3-3.8)</td>
<td>Leukemia 2.1 (2.0-2.4)</td>
<td>Leukemia 2.2 (2.0-2.4)</td>
<td>Thyroid 5.2 (3.7-6.7)</td>
<td>Leukemia 2.6 (2.3-2.9)</td>
</tr>
<tr>
<td>7</td>
<td>Leukemia 2.7 (2.4-3.0)</td>
<td>Colorectal 2.9 (1.7-4.3)</td>
<td>Multiple myeloma 3.1 (2.4-4.0)</td>
<td>Lung 2.7 (2.3-3.2)</td>
<td>Bladder 1.5 (1.5-1.5)</td>
<td>Lung 2.1 (1.9-2.3)</td>
<td>LOCP 4.9 (3.4-6.4)</td>
<td>Colorectal 2.6 (2.3-2.9)</td>
</tr>
<tr>
<td>8</td>
<td>Uterus 2.5 (2.2-2.8)</td>
<td>Multiple myeloma 2.5 (1.4-3.6)</td>
<td>Colorectal 2.7 (1.9-3.5)</td>
<td>Bladder 2.5 (2.3-2.9)</td>
<td>Liver 1.4 (1.4-1.4)</td>
<td>Colorctal 1.9 (1.7-2.3)</td>
<td>Leukemia 3.8 (2.5-3.3)</td>
<td>Kidney 2.5 (2.2-2.8)</td>
</tr>
<tr>
<td>9</td>
<td>Kidney 2.4 (2.1-2.7)</td>
<td>Bladder 2.2 (1.3-3.2)</td>
<td>Uterus 2.0 (0.9-2.7)</td>
<td>Multiple myeloma 2.1 (1.7-2.3)</td>
<td>LOCP 1.3 (0.9-1.3)</td>
<td>Thyroid 1.5 (0.9-1.7)</td>
<td>Colorectal 3.7 (2.4-5.0)</td>
<td>Liver 1.8 (1.5-2.3)</td>
</tr>
<tr>
<td>10</td>
<td>LOCP 1.4 (1.2-1.6)</td>
<td>Thyroid 2.2 (1.5-3.2)</td>
<td>Kidney 1.8 (1.1-2.5)</td>
<td>Kidney 2.1 (1.7-2.5)</td>
<td>Kidney 1.3 (1.3-1.3)</td>
<td>Bladder 1.3 (1.2-1.4)</td>
<td>Bladder 3.2 (2.0-4.4)</td>
<td>Larynx 1.3 (1.1-1.5)</td>
</tr>
</tbody>
</table>

Notes: — indicates data not available. A single individual can have multiple cancer types recorded. Abbreviation: LOCP: Lip, oral cavity, and pharynx.
Leukemia, liver and lung cancer were also within the top 10 in the majority of databases. We did not observe meaningful differences (i.e., |SMD| > 0.1) when comparing cancer types between the diagnosed and the hospitalized cohorts (Supplementary Fig. S3).

**Prior comorbidities**

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

**Thirty-day outcomes**

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

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

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

**Discussion**

In this multinational cohort study, we described the characteristics of 366,050 patients with a history of cancer and COVID-19, including outcomes rarely reported in this population (e.g., deep vein thrombosis, pulmonary embolism, or acute kidney injury).
Figure 2.
SMDs of selected baseline comorbidities between patients with cancer diagnosed and hospitalized with COVID-19. SMD < 0 indicates that the prevalence was greater in patients diagnosed, SMD > 0 indicates that the prevalence was greater in patients hospitalized. NOTES: Comorbidities ordered according to SMD descending values in the largest database (IQVIA-OpenClaims). Black-dotted lines indicate an |SMD| of 0.1. SMD calculated for comorbidities meeting the minimum count required (5 individuals) in each database and cohort.

Figure 3.
Baseline comorbidities (A) and 30-day outcomes (B) among patients with history of cancer hospitalized with COVID-19 and with seasonal influenza. NOTES: Comorbidities and outcomes ordered according to descending values in the largest database (IQVIA-OpenClaims). Comorbidities and outcomes are shown if meeting the minimum count required (5 individuals) in each database and cohort. Outcomes not shown due to data not available: all outcomes in CU-AMC-HDC, occurrence of death in CUIMC (influenza cohort) and IQVIA-OpenClaims, intensive services in CUIMC.
In both COVID-19 cohorts, the most frequent cancer types were prostate cancer and breast cancer; hematologic malignancies were also frequent. The proportion of patients that had received anti-cancer therapies the year or the month prior was similar in both cohorts. Comorbidities were common in both cohorts but were higher among those hospitalized. Occurrence of death ranged from 2% to 14% among those diagnosed and from 6% to 26% among those hospitalized. When compared with patients with cancer history hospitalized with seasonal influenza, patients hospitalized with COVID-19 had a similar distribution of age and comorbidities but had more severe outcomes.

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

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

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

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

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

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

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

However, this study also has limitations. First, we were not able to provide detailed cancer information, such as year of cancer diagnosis, nor identify patients with active cancer treatment; although we had information on the use of antineoplastic agents during the year and month prior to the index date. Second, by including patients with a clinical COVID-19 diagnosis we might have incurred some false positives. However, we used a broad COVID-19 definition to reduce selection bias due to testing restrictions during the first months of the pandemic (32), as well as (hypothetical) differential patterns in testing between patients with cancer versus patients without cancer.

In addition, we did not have information on socioeconomic status, ethnicity, nor race in most databases. We also lacked information on the cause of death and reported instead all-cause death. Third, the overlap between the diagnosed and hospitalized COVID-19 cohorts might have masked some differences in the prevalence of comorbidities between cohorts. Moreover, some patients might be included in more than one database (e.g., in a hospital-based and claims-based database from the United States). Unfortunately, we were unable to determine the degree of overlap across data sources because patient-level data was not shared for confidentiality purposes. Fourth, the differences found in the COVID-19/seasonal influenza comparison may have been influenced by temporal changes in clinical practice standards and coding. Further, the influenza vaccine likely contributed to the low frequency of adverse events among influenza patients. Fifth, the use of routinely collected data could have led to an underestimation of the lifetime cancer prevalence, cancer types, comorbidities, and outcomes due to incomplete reporting. Finally, our findings were heterogeneous across data sources. Heterogeneity is a known phenomenon when using real-world data that reflects the existence of different coding practices, observation period, healthcare settings, and populations. Although the interpretation of heterogeneous results is challenging, these also provide valuable insights into the particularities of each setting. Yet, despite this heterogeneity, we found consistent patterns when comparing characteristics across cohorts, which lends cre-
idence to our results.

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

Authors’ Disclosures

A.G. Sena reports employment with Janssen R&D and Johnson & Johnson stock ownership. C. Blacketer reports other support from Janssen Research & Development outside the submitted work. S.L. DuVall reports grants from Genomic
Characteristics of 300,000 COVID-19 Individuals with Cancer

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

Authors’ Contributions

E. Roel: Conceptualization, visualization, writing—original draft, writing—review and editing. A. Pistillo: Formal analysis, visualization, writing—review and editing. M. McRade: Writing—review and editing. A.G. Sena: Data curation, formal analysis, writing—review and editing. S. Fernandez-Bertolini: Formal analysis, writing—review and editing. M. Aragón: Data curation, writing—review and editing. D. Puente: Writing—review and editing. W.U.R. Ahmed: Writing—review and editing. H. Alghoul: Writing—review and editing. O. Alser: Writing—review and editing. T.M. Alshammar: Writing—review and editing. C. Areia: Writing—review and editing. M. Gong: Writing—review and editing. A. Pistillo: Writing—review and editing. M. Aragón: Data curation, writing—review and editing. W. Carter: Data curation, writing—review and editing. P. Casajust: Writing—review and editing. A.C. Culhane: Writing—review and editing. D. Davoud: Writing—review and editing. F. DeFalco: Data curation, writing—review and editing. S.L. DuVall: Data curation, writing—review and editing. T. Falconer: Data curation, writing—review and editing. A. Golozor: Writing—review and editing. L. Hester: Writing—review and editing. G. Hripcsak: Data curation, writing—review and editing. E.H. Tan: Writing—review and editing. H. Jeon: Writing—review and editing. J. Jonnagaddala: Writing—review and editing. L.Y.H. Lai: Writing—review and editing. K.E. Lynch: Data curation, writing—review and editing. M. Matheny: Data curation, writing—review and editing. D.R. Morales: Writing—review and editing. K. Natarajan: Data curation, writing—review and editing. F. Nyberg: Writing—review and editing. A. Ostropole: Data curation, writing—review and editing. J.D. Posada: Data curation, formal analysis, writing—review and editing. A. Prats-Urbi: Conceptualization, formal analysis, writing—review and editing. C.G. Reich: Data curation, writing—review and editing. D.R. Rivera: Writing—review and editing. L.M. Schilling: Data curation, writing—review and editing. I. Soerjomataram: Writing—review and editing. Y. Shah: Writing—review and editing. N.I. Shah: Data curation, writing—review and editing. Y. Shen: Writing—review and editing. M. Spotniz: Data curation, writing—review and editing. M.A. Suchard: Writing—review and editing. Y. Zhang: Writing—review and editing. P.B. Ryan: Conceptualization, writing—review and editing. D. Prieto-Alhambra: Conceptualization, writing—review and editing. K. Kostka: Conceptualization, data curation, formal analysis, writing—review and editing. T. Duarte-Salles: Conceptualization, data curation, formal analysis, supervision, writing—review and editing.

Acknowledgments

We would like to acknowledge the patients who suffered from or died of this devastating disease, and their families and carers. We would also like to thank the health care professionals involved in the management of COVID-19 during these challenging times, from primary care to intensive care units. This project has received support from the European Health Data and Evidence Network (EHEDEN) project. EHEDEN has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (IU) under grant agreement no. 806968. The IU receives support from the European Union’s Horizon 2020 research and innovation program and EFPIA. This research received partial support from the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), U.S. National Institutes of Health (NIH), U.S. Department of Veterans Affairs, Janssen Research & Development, and IQVIA. This work was also supported by the Bio Industrial Strategic Technology Development Program (20001234) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no.: HI16C0092). Bill & Melinda Gates Foundation (INV-016201), and the Dirección General de Recerca i Innovació en Salut from the Department of Health of the Generalitat de Catalunya. E. Roel is supported by Instituto de Salud Carlos III (grant no. CM20/00174). W.U.R. Ahmed and others reports funding from the NIHR Oxford Biomedical Research Centre (BRC), Arzif Foundation, Wollon Foundation, and the Royal College of Surgeons of England. A.G. Sena and L. Hester are employed by Janssen Research and Development, LLC. J. Jonnagaddala reports grants from the National Health and Medical Research Council (APP1192469). M. Gong and Y. Zhang report grants from National Key Research & Development Program of China (Project No. 2018YFC0116901). V. Subbian reports funding from the National Science Foundation, Agency for Healthcare Research & Quality, National Institute of Health, and the State of Arizona. A. Prats-Urbi is supported by Fundación Alfonso Martin Escudero support from the Medical Research Council (grant nos. MR/K501256/1, MR/N013468/1). E. Roel is supported by Instituto de Salud Carlos III (grant no. CM20/00174). W.U.R. Ahmed and others reports 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 Clinicalician Scientist Award program, NIHR, Department of Veterans Affairs, or the U.S. Government, NHS, or the Department of Health, England. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

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

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

References

Roel et al.


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

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


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-21-0266

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2021/07/15/1055-9965.EPI-21-0266.DC1

Cited articles
This article cites 26 articles, 3 of which you can access for free at:
http://cebp.aacrjournals.org/content/30/10/1884.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/30/10/1884.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.