Received: 12 July 2021 | Revised: 31 August 2021 | Accepted: 20 September 2021

DOI: 10.1002/ijc.33846

CANCER EPIDEMIOLOGY

IJC INTERNATIONAL JOURNAL of CANCER

Cancer and the risk of coronavirus disease 2019 diagnosis, hospitalisation and death: A population-based multistate cohort study including 4 618 377 adults in Catalonia, Spain

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

[1]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

[2]Universitat Autònoma de Barcelona, Barcelona, Spain

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

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

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

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

Abstract

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

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

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

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

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

KEYWORDS

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

What's new?

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

1 | INTRODUCTION

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

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

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

2 | MATERIALS AND METHODS

2.1 | Study design, setting and data sources

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

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

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

2.2 | Study participants

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

2.3 | Multistate framework

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

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

2.4 | Variables

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

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

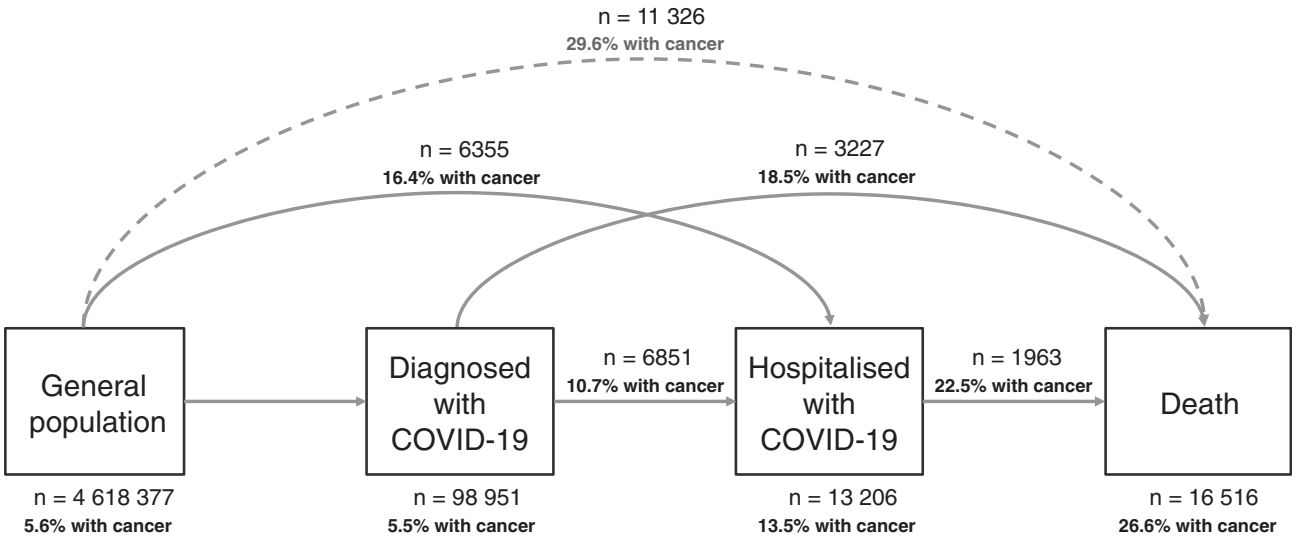

FIGURE 1 Overview of the multistate model used in our study

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

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

2.5 | Statistical analyses

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

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

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

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

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

3 | RESULTS

3.1 | Population included

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

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

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

TABLE 1 (Continued)

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

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

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

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

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

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

Abbreviations: CI, cumulative incidence; IRQ, interquartile range.
[a]Years since cancer diagnosis to the index date (1 March 2020).

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

3.2 | Occurrence of COVID-19 outcomes

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

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

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

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

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

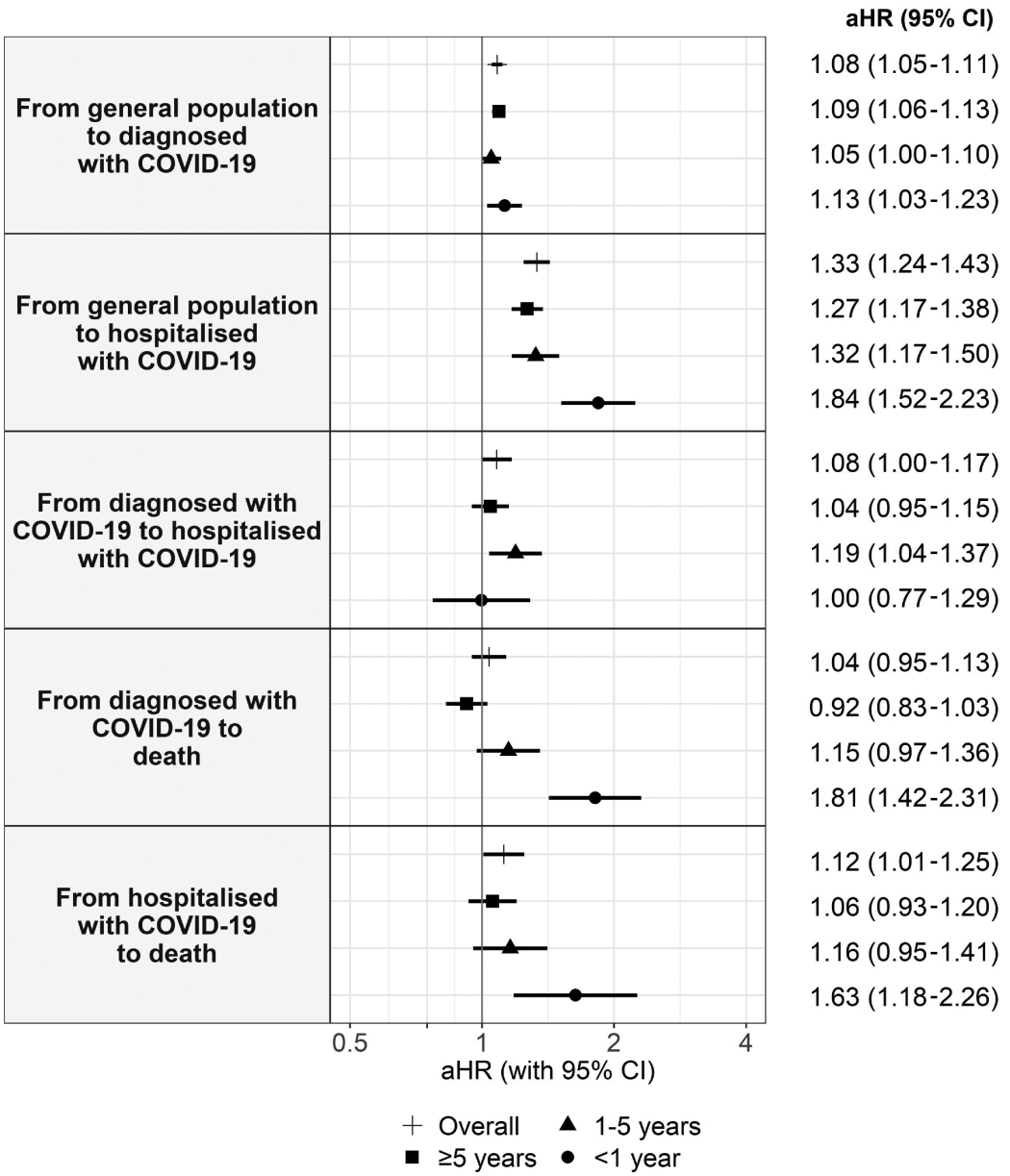

FIGURE 2 Adjusted hazard ratios of COVID-19 outcomes in patients with cancer compared to patients without cancer, overall and by years since cancer diagnosis. Models are adjusted for age, sex, the MEDEA deprivation index, smoking status and comorbidities (autoimmune conditions, chronic kidney disease,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

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

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

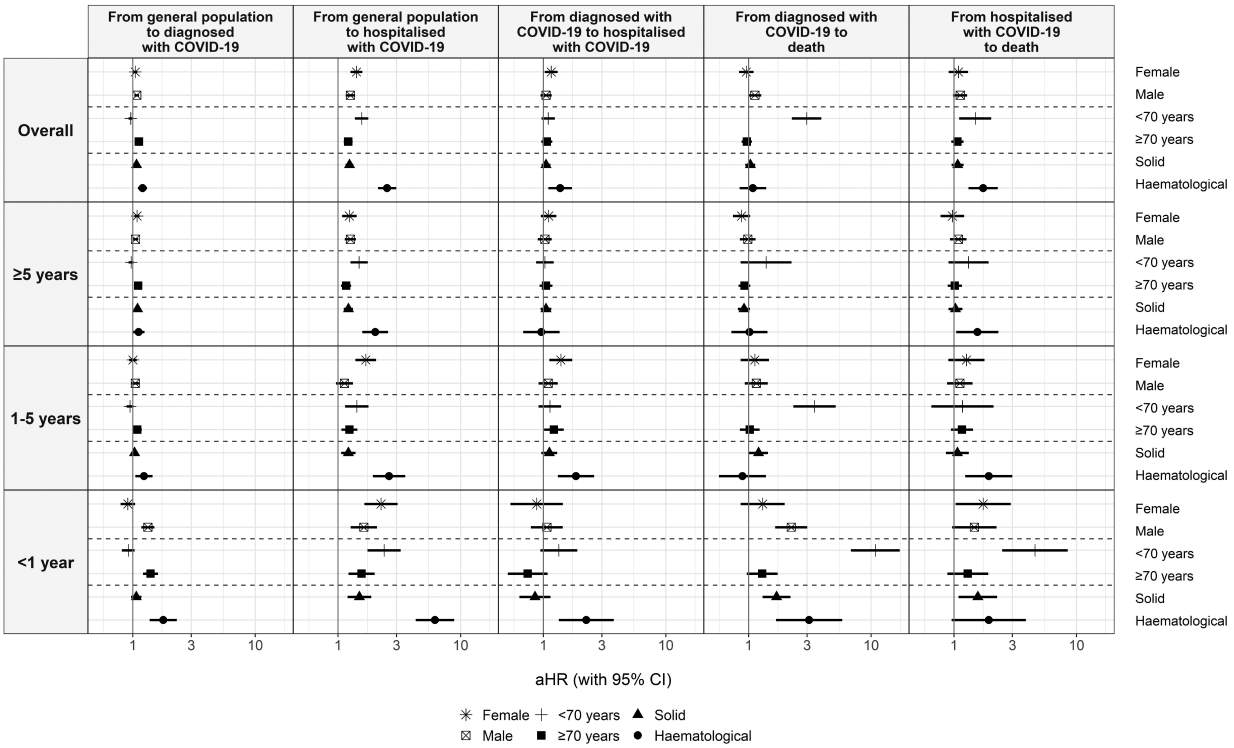

FIGURE 3 Adjusted hazard ratios of COVID-19 outcomes in patients with cancer (overall and by years since cancer diagnosis) compared to patients without cancer, stratified by sex, age and cancer type (solid or haematological). Models are adjusted for sex (excepting models stratified by sex), age, the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

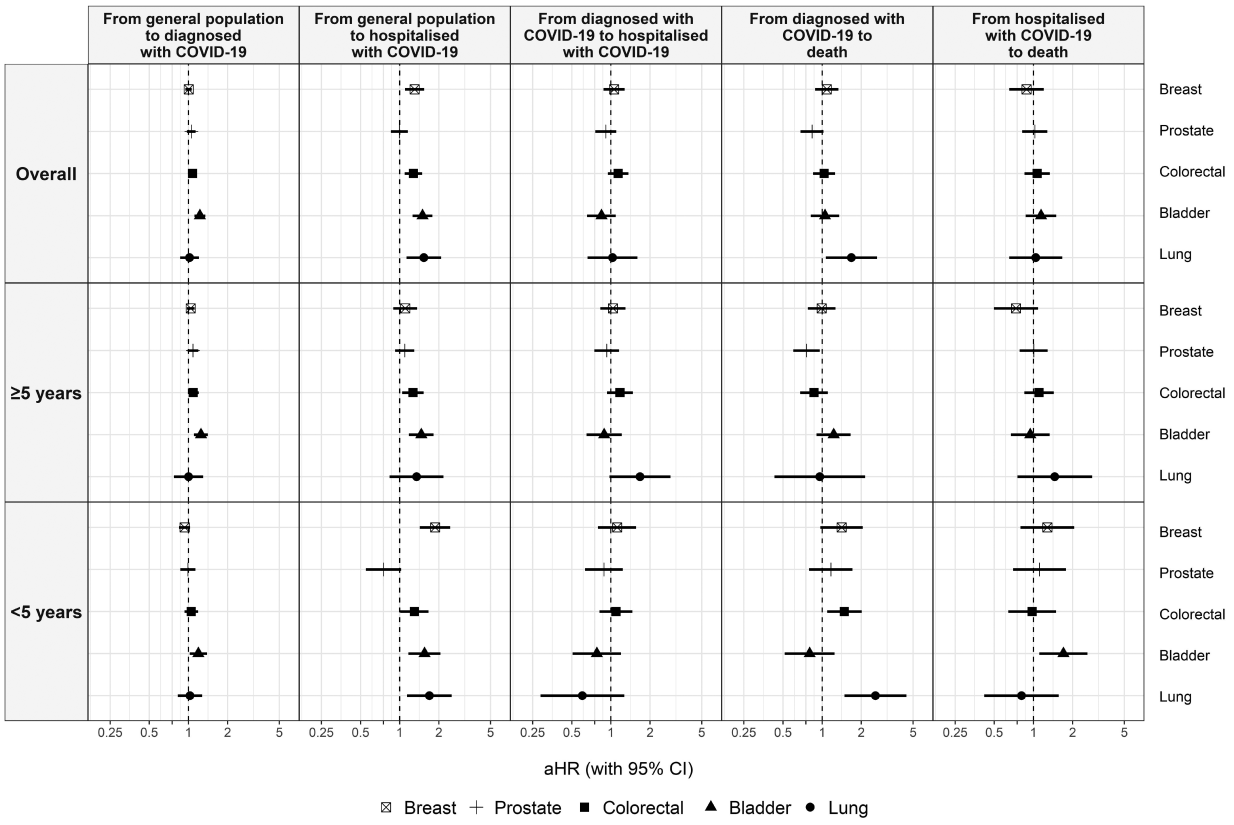

FIGURE 4 Adjusted hazard ratios of COVID-19 outcomes in patients with cancer (overall and by years since the cancer diagnosis) compared to patients without cancer, stratified by solid cancer type. Models for specific cancer types include patients without cancer and patients with the cancer type of interest;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

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

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

3.4 | Sensitivity analysis

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

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

4 | DISCUSSION

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

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

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

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

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

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

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

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

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

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

ACKNOWLEDGEMENTS

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

CONFLICT OF INTEREST

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

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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

TRANSPARENCY STATEMENT

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

ORCID

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

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. https://doi.org/10.3322/caac.21660
2. Kumar D, Humar A. Respiratory viral infections in transplant and oncology patients. *Infect Dis Clin North Am*. 2010;24:395-412.
3. Liang W, Guan W, Chen R, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *Lancet Oncol*. 2020;21(3):335-337.
4. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708-1720.
5. Zhang Y, Han H, Tian Y, et al. Impact of cancer on mortality and severity of corona virus disease 2019: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. 2020;99(44):e23005.
6. Wang Q, Berger NA, Xu R. Analyses of risk, racial disparity, and outcomes among US patients with cancer and COVID-19 infection. *JAMA Oncol*. 2020;7(2):220-227.
7. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584:430-436.
8. Kuderer NM, Choueiri TK, Shah DP, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet*. 2020;6736:1-13.
9. Lee LYW, Cazier JB, Starkey T, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol*. 2020;21:1309-1316.
10. Angelis V, Tippu Z, Joshi K, Reis S, et al. Defining the true impact of coronavirus disease 2019 in the at-risk population of patients with cancer. *Eur J Cancer*. 2020;136:99-106.
11. Reilev M, Kristensen KB, Pottegård A, et al. Characteristics and predictors of hospitalization and death in the first 11 122 cases with a positive RT-PCR test for SARS-CoV-2 in Denmark: a nationwide cohort. *Int J Epidemiol*. 2020;49:1468-1481.
12. Brar G, Pinheiro LC, Shusterman M, et al. COVID-19 severity and outcomes in patients with cancer: a matched cohort study. *J Clin Oncol*. 2020;38:3914-3924.
13. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11:1-12.
14. Prieto-Alhambra D, Balló E, Coma E, et al. Filling the gaps in the characterization of the clinical management of COVID-19: 30-day hospital admission and fatality rates in a cohort of 118 150 cases diagnosed in outpatient settings in Spain. *Int J Epidemiol*. 2020;49:1930-1939.
15. Bolíbar B, Fina Avilés F, Morros R, et al. Base de datos SIDIAP: La historia clínica informatizada de Atención Primaria como fuente de información para la investigación epidemiológica. *Med Clin (Barc)*. 2012;138:617-621.
16. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-578.
17. Burn E, Tebé C, Fernandez-Bertolin S, et al. The natural history of symptomatic COVID-19 during the first wave in Catalonia. *Nat Commun*. 2021;12:1-12.
18. Putter H, Fiocco M, Gekus RB. Tutorial in biostatistics: competing risk and multi-state models. *Stat Med*. 2007;26:2389-2430.
19. Recalde M, Manzano-Salgado CB, Díaz Y, et al. Validation of cancer diagnoses in electronic health records: results from the information system for research in primary care (SIDIAP) in Northeast Spain. *Clin Epidemiol*. 2019;11:1015-1024.
20. Domínguez-Berjón MF, Borrell C, Cano-Serral G, et al. Construcción de un índice de privación a partir de datos censales en grandes ciudades españolas (Proyecto MEDEA). *Gac Sanit*. 2008;22:179-187.
21. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA*. 2020;323:2052-2059.
22. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
23. Harrell FE. *Regression Modeling Strategies*. New York, NY: Springer; 2015.
24. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput*. 2009;38:1228-1234.

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

SUPPORTING INFORMATION

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

How to cite this article: Roel E, Pistillo A, Recalde M, et al. Cancer and the risk of coronavirus disease 2019 diagnosis, hospitalisation and death: A population-based multistate cohort study including 4 618 377 adults in Catalonia, Spain. *Int. J. Cancer*. 2022;150(5):782-794. doi:10.1002/ijc.33846