



BMJ Open Increased risk of COVID-19-related admissions in patients with active solid organ cancer in the West Midlands region of the UK: a retrospective cohort study

Akinfemi Akingboye ¹, Fahad Mahmood,¹ Nabeel Amiruddin,¹ Michael Reay,¹ Peter Nightingale,² Olorunseun O Ogunwobi ³

To cite: Akingboye A, Mahmood F, Amiruddin N, *et al*. Increased risk of COVID-19-related admissions in patients with active solid organ cancer in the West Midlands region of the UK: a retrospective cohort study. *BMJ Open* 2021;**11**:e053352. doi:10.1136/bmjopen-2021-053352

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-053352>).

Received 11 May 2021

Accepted 21 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Russells Hall Hospital, Dudley, UK

²Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

³Biological Sciences, Hunter College CUNY, New York City, New York, USA

Correspondence to

Dr Olorunseun O Ogunwobi; oo158@hunter.cuny.edu

ABSTRACT

Objective Susceptibility of patients with cancer to COVID-19 pneumonitis has been variable. We aim to quantify the risk of hospitalisation in patients with active cancer and use a machine learning algorithm (MLA) and traditional statistics to predict clinical outcomes and mortality.

Design Retrospective cohort study.

Setting A single UK district general hospital.

Participants Data on total hospital admissions between March 2018 and June 2020, all active cancer diagnoses between March 2019 and June 2020 and clinical parameters of COVID-19-positive admissions between March 2020 and June 2020 were collected. 526 COVID-19 admissions without an active cancer diagnosis were compared with 87 COVID-19 admissions with an active cancer diagnosis.

Primary and secondary outcome measures 30-day and 90-day post-COVID-19 survival.

Results In total, 613 patients were enrolled with male to female ratio of 1:6 and median age of 77 years. The estimated infection rate of COVID-19 was 87 of 22 729 (0.4%) in the patients with cancer and 526 of 404 379 (0.1%) in the population without cancer (OR of being hospitalised with COVID-19 if having cancer is 2.942671 (95% CI: 2.344522 to 3.693425); $p < 0.001$). Survival was reduced in patients with cancer with COVID-19 at 90 days. R-Studio software determined the association between cancer status, COVID-19 and 90-day survival against variables using MLA. Multivariate analysis showed increases in age (OR 1.039 (95% CI: 1.020 to 1.057), $p < 0.001$), urea (OR 1.005 (95% CI: 1.002 to 1.007), $p < 0.001$) and C reactive protein (CRP) (OR 1.065 (95% CI: 1.016 to 1.116), $p < 0.008$) are associated with greater 30-day and 90-day mortality. The MLA model examined the contribution of predictive variables for 90-day survival (area under the curve: 0.749); with transplant patients, age, male gender and diabetes mellitus being predictors of greater mortality.

Conclusions Active cancer diagnosis has a threefold increase in risk of hospitalisation with COVID-19. Increased age, urea and CRP predict mortality in patients with cancer. MLA complements traditional statistical analysis in

Strengths and limitations of this study

- The study uses novel analytical methods derived from machine learning to evaluate risk of COVID-19 in patients with cancer from hospitalisation to mortality.
- Statistical and machine learning methods are compared to develop a profile of factors that can worsen outcomes from COVID-19 in patients with cancer.
- The study analyses COVID-19 outcomes in patients with solid organ cancer in a cohort covering a single UK metropolitan region only. No haematological malignancies analysed.
- Patients with COVID-19 and cancer who did not require admission to hospital were not included in this study.

identifying prognostic variables for outcomes of COVID-19 infection in patients with cancer. This study provides proof of concept for MLA in risk prediction for COVID-19 in patients with cancer and should inform a redesign of cancer services to ensure safe delivery of cancer care.

INTRODUCTION

SARS-CoV-2 leads to COVID-19.^{1 2} This highly transmissible disease has led to a global pandemic contributing to significant morbidity and mortality. Increased susceptibility and severity of COVID-19 are attributed to increasing age, smoking status, chronic obstructive pulmonary disease, diabetes mellitus, obesity, cardiovascular disease as well as cancer.³⁻⁶ In addition, the prevalence of all types of active or previous cancer in the UK is reported at 2.5 million cases with an incidence of 1000 newly diagnosed cases each day.⁷ Increased susceptibility to COVID-19 in patients with cancer has been attributed to immune suppression and cancer treatments such as cytotoxic chemotherapy and

immunotherapy.^{8 9} However, it is still not established whether this translates into increased hospitalisation, illness severity or mortality risk. Risk-adjusted models quote a mortality risk of between 25% and 39% in patients with cancer hospitalised with COVID-19.¹⁰ With increasing prevalence of COVID-19 in the UK, the impact of cancer on COVID-19 remains an area of active concern.

In addition, machine learning algorithms (MLAs) have become increasingly applied in healthcare settings due to their prognostic utility.¹¹⁻¹³ They are able to map a large number of observed variables (features) to target outcomes, and through statistical analysis, find relationships without human instruction.^{14 15} This utility has been exploited in cancer research to model risk of susceptibility, survival and recurrence.¹⁵ For example, in breast cancer, algorithms have been developed from detecting breast tumours to determining the prognostic significance of the tumour's morphological features.¹⁴ Moreover, the ability to integrate diverse variables including clinical, biochemical, histopathological, genomic and proteomic data could lead to more reliable predictive models to determine disease outcome.^{16 17} Furthermore, the scalability of MLA distinguishes it from traditional statistical modelling, such as regression analysis, by its ability to perform non-linear modelling using large volume data sets and greater number of variables from registries.^{11 12} MLA models for risk prediction are starting to be validated in large studies.¹⁸ Thus, this nascent technique holds promise for developing better risk assessment and prognostic algorithms to support healthcare delivery and individualised patient care.

This study aims to quantify the risk of hospitalisation in patients with active cancer using specific differences in clinicopathological and biochemical parameters between patients with COVID-19 with cancer and those without cancer through developing an MLA. We seek to identify the most important determinants of high risk of susceptibility and mortality from a diverse range of variables. This will both provide proof of concept for our method as well as inform the recalibration of cancer services to ensure safe care for patients with cancer during the pandemic.

METHODS

A single UK centre retrospective cohort study was conducted. Data on total hospital admissions between March 2018 and June 2020 were obtained from the local information technology department with a record of all hospital admissions before and during the COVID-19 pandemic linking this with the Somerset cancer database to extrapolate the total number of patients with active cancer who were admitted during the study period. Furthermore, all active solid organ cancer diagnoses between March 2019 and June 2020 were obtained from the local cancer network. This was used to determine the total number of patients with active cancer with COVID-19 with the denominator being the total and active cancer population in the Dudley, West Midlands (UK) region.

Biochemical and haematological parameters in the first 48 hours of admission along with 30-day and 90-day post-COVID-19 survival were determined.

Patients below the age of 18 years and those with non-solid organ cancers were excluded. Moreover, patients who attended the emergency department and were not admitted were also excluded. COVID-19 diagnosis was established with a positive reverse transcriptase PCR test from an oropharyngeal swab. Criteria for admission to hospital and critical care were determined by individual clinical assessment and oxygen requirement as well as ventilatory support. Data security was maintained through the REDCap uploading system.

Binary logistic regression analyses with survival status at 30 days as the dependent variable were used to estimate the univariable association with mortality for each explanatory variable. Age-adjusted associations were calculated in a similar way by including age at admission as a continuous variable in each model after checking the assumption of a linear effect of age on the log odds. Both forward and backward stepwise methods were used to determine the final multivariable model. These analyses were performed with SPSS V.25.0.

Patient and public involvement

Patients and the public were not involved in the design and conduct of this study.

MLA: data preprocessing

R-Studio software was used to determine the association between cancer status, COVID-19 and 90-day survival against variables in an MLA. The conduct and reporting of our MLA was done in accordance with best practice guidance.¹¹

Feature selection

The proportion of missing data was calculated for each variable, and variables with less than 40% missing data were included in the analysis. This resulted in 33 variables being included for imputation of missing data, further preprocessing and model development. The decision to limit the proportion of missing data to 40% was an arbitrary one, based on a compromise between a limit high enough to enable the inclusion of as many available variables as possible and low enough to enable the use of more data to predict imputable missing values with the k-nearest neighbours algorithm.

All data within the gender variable were replaced with 'F' (female) and 'M' (male). The documented ethnicities were replaced with the three categories of 'European', 'South Asian' or 'Afro-Caribbean'. The blood pressure information was split into systolic and diastolic pressures. A new dummy variable of 'mean arterial pressure' was derived from the estimate: $(\text{diastolic pressure} + (\text{pulse pressure}/3))$. The pure numerical values from the entered data for oxygen saturation were extracted. For example, 97% would be changed to 97. A dummy variable was created from the difference in time between the date

of onset of symptoms and date of hospital admission. This time interval was recorded in days.

Data partition

The overall data set was partitioned into training and test sets. The training set was used purely for model training and hyperparameter tuning. The test set would only be used for model evaluation against new data. Partitioning was by a random allocation, while ensuring an identical distribution of patients who died at 90 days between both training and test sets. Seventy-five per cent of patients were allocated to the training set, with the remaining going into the test set (online supplemental figure 1).

Imputation of missing values

Missing values were replaced with predicted values using k-nearest neighbours model. This method designated the variable of a missing value as an outcome variable within a predictive model. A prediction of the missing value on the most similar k number of patients was based on their other variables. The value k is a hyperparameter which was set to 10 after comparing the values of 5 and 10 without any difference.

This imputation was performed separately on the training and test data sets in order to minimise overfitting of the final model by having the training data set influence the imputation of values into the test data set.

Additional preprocessing

All numerical variables within the training set were preprocessed for model training to be on comparable scales ranging mainly from 0 to 1. For each such variable, the mean was subtracted from each value before dividing the result by the SD.

The same process was applied to the test set, using the means and SDs from the training set to avoid overfitting.

Model development

The following models were trained using 10-fold cross-validation:

1. Logistic regression.
2. Lasso and elastic-net generalised linear model.
3. K-nearest neighbour.
4. Random forest.
5. Neural network with one hidden layer.
6. Gradient boosted machine.

Hyperparameter tuning during cross-validation was optimised against area under the receiver operating curve as a metric. The random forest model was built with 500 trees.

Model evaluation

Predictions of probabilities of survival to 90 days were made on the test set by each of the five trained models. The known survival outcomes to 90 days and predicted probabilities from each model were used to plot receiver operating curves for model for comparison.

RESULTS

In total, 22 729 patients with active cancer were identified in the Dudley West Midlands region out of a catchment size of 426 658 patients in the region from the local cancer network. Eighty-seven of 22 729 (0.4%) patients with cancer in the Dudley region were admitted with COVID-19 compared with 526 of 404 379 (0.1%) during the study period (HR: OR of being hospitalised with COVID-19 if having cancer is 2.942671 (95% CI: 2.344522 to 3.693425); $p < 0.001$). The types of cancer in our cohort are detailed in figure 1. Thus, the risk of hospital admission on presentation with COVID-19 increased threefold in the presence of an active cancer diagnosis.

Excluding those with incomplete data, the mean age of patients with cancer was 77.8 (SD=12.3) years compared with 70 (SD=17.5) years (t-test; $p < 0.001$). The male:female ratio was similar between the two groups. The majority of patients were of Caucasian ethnicity with similar distribution of diabetes, cardiovascular disease, transplant recipient and smoking status. Moreover, the median white cell count ($p = 0.096$) and C reactive protein (CRP) ($p = 0.115$) were similar between patients with cancer and those without cancer with no statistically significant difference. Thus, both cancer and non-cancer groups affected by COVID-19 had similar baseline characteristics. This is summarised in table 1.

A χ^2 test, comparing patients without cancer not hospitalised with COVID-19 (404 379) yields a p value of $< 2.2 \times 10^{-16}$, implying that there is an association between having cancer and hospitalisation with COVID-19 (table 2). The OR of being hospitalised with COVID-19 if having cancer is 2.942671 (95% CI: 2.344522 to 3.693425).

After training and hyperparameter tuning by 10-fold cross-validation, predictions of probability of 90-day survival were made on the test set data. This is shown in the receiver operating curves plotted for model comparison (figure 2).

The random forest model achieved an area under the receiver operating curve of 0.829 (figure 2). Each variable was evaluated for its relative contribution to enabling classification (figure 3).

Since we accepted variables with up to 40% missing values (online supplemental figure 2), imputation was performed using a separate k-nearest neighbours algorithm, whereby a prediction of a missing value was made based on other available values, having been trained on the other patient data.

Our initial age-adjusted univariate analysis identified age, CRP, urea, creatinine, estimated glomerular filtration rate (GFR), haemoglobin and low initial blood pressure as significantly correlating with mortality risk (online supplemental table 1). A further multivariate analysis of 33 out of 213 clinical variables with $> 60\%$ data completeness showed increased age (HR 0.915 (95% CI: 0.870 to 0.960), $p < 0.001$), urea (HR 1.005 (95% CI: 1.002 to 1.007), $p < 0.001$) and CRP (HR 1.065 (95% CI: 1.016 to 1.116), $p < 0.001$) to be associated with greater risk of 30-day and 90-day mortality (table 3).

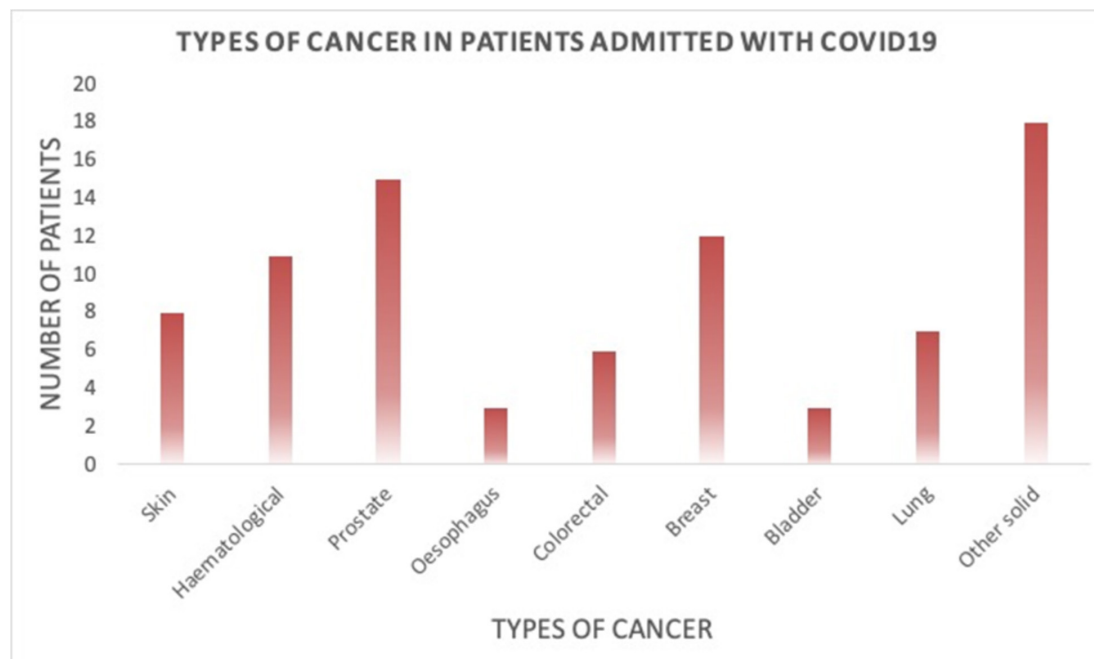


Figure 1 Types of active cancer in our cohort of patients for analysis. Solid organ and skin cancers were grouped together for analysis.

Kaplan-Meier survival analysis revealed reduced overall survival for patients with COVID-19 and cancer (figure 4). However, log-rank analysis did not show significant difference between patients with COVID-19 with cancer and those without cancer (log-rank $p=0.172$).

DISCUSSION

Our study demonstrates that the presence of active cancer increased by threefold the risk of hospitalisation with COVID-19. Moreover, higher CRP and urea are associated with greater mortality at 30 and 90 days post-diagnosis of COVID-19. These findings show that patients with cancer who develop COVID-19 are likely to have a more severe form of the infection that would require supportive care in hospital. It also provides tools for monitoring patient response to treatment with high urea and CRP being poor prognostic markers and a likely consequence of severe COVID-19. This has implications for how we can deliver safe care to patients with cancer in the ongoing pandemic as well as emerging from it given the restrictions on cancer services.

Several studies have reported prevalence and mortality risk of COVID-19 in patients with cancer with a systematic review by Zarifkar *et al* identifying 110 studies covering 10 countries.¹⁹ The pooled prevalence of active cancer in COVID-19-positive hospitalised patients was 2.6% (95% CI: 1.8% to 3.5%) across 37 cohort studies. Furthermore, there was a noticeable difference in the prevalence between western countries (5.6%, 95% CI: 4.5% to 6.7%) and China (1.7%, 95% CI: 1.3% to 2.3%) reflecting the underlying cancer prevalence. In addition, in-hospital mortality of 14.1% (95% CI: 9.1% to 19.8%) for cancer and COVID-19 was derived from 17

retrospective cohort studies covering 904 patients.¹⁹ The mortality rate of 12.6% in a Brazilian cohort was also similarly reported.²⁰ This indicated that patients with COVID-19 with cancer had a fivefold greater risk of death compared with patients without cancer without other comorbidities.^{19 21} However, there was significant heterogeneity between these studies ($I^2=55.9%$, $p<0.01$) with the type of cancer, stage and treatment regimen only specified in eight studies along with incomplete follow-up. Furthermore, Liang *et al* reported a 28% prevalence of lung cancer among hospitalised patients with cancer with COVID-19.⁹ This reflects higher COVID-19 mortality rates in patients with specific cancer including lung and haematological malignancy.^{9 19 20} Further studies have reported 3.5-fold increase in intensive care unit (ICU) admission or need for mechanical ventilation in patients with COVID-19 with cancer.⁹ More recent studies have also examined the impact of COVID-19 on patients with cancer. In an analysis of 306 patients with COVID-19 with cancer, Russell *et al* identified factors including male gender, age greater than 60 years, Asian ethnicity, cancer diagnosis of greater than 2 years, haematological malignancy and a high CRP associated with increased mortality risk.²² A large population-based study by Lee *et al* comparing 23 266 patients with cancer with 1 784 293 patients without cancer identified a 60% increased risk of COVID-19 in patients with cancer, with those on chemotherapy or immunotherapy having a 2.2-fold increased risk of contracting COVID-19.²³ This increased susceptibility could be explained through immune compromise of simply greater exposure through more frequent hospital visits. Even in this large study, subgroup analysis was not performed evaluating the impact of tumour type, stage

Table 1 Comparing characteristics of patients with cancer and those without cancer

	Patients with cancer (n=80)	Patients without cancer (n=276)	P value
Age in years (n=356): mean (SD)	77.8 (12.3)	70.0 (17.5)	<0.001
Sex			0.699
Female	34 (43%)	112 (41%)	
Male	45 (57%)	164 (59%)	
Ethnicity			0.280
Afro-Caribbean	1 (1%)	9 (4%)	
European	70 (95%)	197 (87%)	
South Asian	3 (4%)	50 (9%)	
Smoking			0.176
Current	5 (22%)	18 (30%)	
Ex	9 (39%)	11 (18%)	
Never	9 (39%)	31 (52%)	
Cardiovascular			0.103
Yes	22 (31%)	107 (41%)	
No	50 (69%)	152 (59%)	
Diabetes mellitus			0.885
Yes	22 (29%)	73 (28%)	
No	53 (71%)	187 (72%)	
Transplant patient			0.644
Yes	2 (3%)	5 (2%)	
No	70 (97%)	261 (98%)	
Reason for admission			<0.001
Yes	48 (70%)	227 (90%)	
No	21 (30%)	24 (10%)	
White cell count (n=332): median (lower quartile–upper quartile)	8.8 (5.6–12.7)×10 ⁹ /L	7.2 (5.3–10.6)×10 ⁹ /L	0.096
CRP (n=324): median (lower quartile–upper quartile)	77 (22–135)mg/L	84 (36–157)mg/L	0.115

Values are counts and percentages except where stated. The p values are from Fisher's exact test, except for age (from a t-test), white cell count and CRP (both from Mann-Whitney tests).
CRP, C reactive protein.

and treatment regimens. Furthermore, in a multicentre study, comparing UK (n=468) and European Union (EU) (n=924) patients with cancer with COVID-19, showed a worse mortality rate at 30 days and 6 months independent of age, gender, tumour stage and treatment through a multivariable regression model.²⁴ Moreover, Mehta *et al* showed increased risk of COVID-19 in 218 patients with cancer in New York which was associated with age, comorbidities and elevated lactate dehydrogenase (LDH).²⁵

Table 2 Risk of admission with COVID-19 in patients with cancer

Group	Admitted	Not admitted
Patients without cancer	404 379	526
Patients with cancer	87	22 729

Pearson's X² test with Yates' continuity correction: X-squared=93.641, df=1, p value of <2.2e-16.

Although few studies have indicated no increased risk of severity or mortality from COVID-19 in patients with cancer,^{26 27} the consensus thus far in the literature has coalesced around the idea that patients with cancer in general have a higher risk of susceptibility from severe events and mortality from COVID-19 infection. Using admission risk as a surrogate marker of severity in COVID-19, our results are consistent with the literature showing a threefold higher risk of admission with COVID-19 in the presence of cancer which will likely impact the delivery of care to these particular subgroups of patients. However, the majority of current data are from retrospective cohort studies, using traditional statistical techniques on selected limited variables with a relatively small number of participants. Moreover, for this reason, subgroup analysis has been difficult. Since cancer is a diverse condition from the clinical to genomic spheres, with an equally diverse range of treatments, considering it as a monolithic structure would not let meaningful conclusions to be drawn

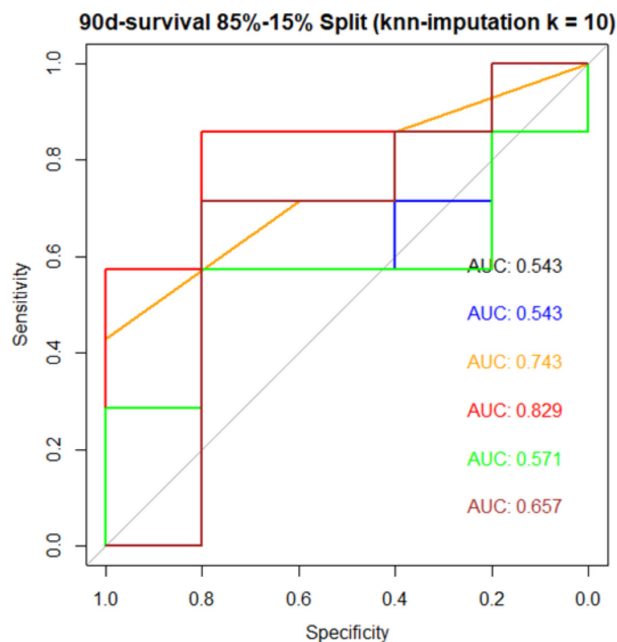


Figure 2 Receiver operating curves for logistic regression (black), generalised linear model (blue), k-nearest neighbours (orange), random forest (red), single hidden layer neural network (green), gradient boosted machine (brown). AUC, area under the curve.

from such analyses. Having a multicentre approach and application of novel big-data analysis techniques such as MLA may enable a more reliable and rapid analysis of data to discover associations in time-critical situations such as delivering healthcare in a global pandemic.

Table 3 Multivariate analysis showing increased age, CRP and urea are associated with the highest 90-day mortality risk in patients with COVID-19

	P value	OR	Lower CI	Upper CI
Age	0.000	1.039	1.020	1.057
CRP	0.001	1.005	1.002	1.007
Urea	0.008	1.065	1.016	1.116

CRP, C reactive protein.

There is likely to be a surge in demand for cancer services as well as predicted poor long-term survival in patients with cancer due to delays in diagnosis and treatment.^{11 28} Over the first national UK lockdown, there was a 84% reduction in urgent cancer referrals which modelling predicted would lead to 181 additional lives lost or 3316 life-years lost with an average presentation delay of 2 months per patient.²⁹ Although having cancer puts a patient at increased risk of hospitalisation with COVID-19, this must be balanced against risks of delayed treatment leading to disease progression to incurable stages.³⁰ Particular cancers where timely intervention is critical such as pancreatic, lung and haematological malignancy should not have delays to treatment, whereas others including prostate and non-melanoma skin cancer treatment may be safely delayed in selected patients.³⁰ Several strategies including: delays to surgery or chemotherapy, switching to oral or monotherapy treatment regimens, strict infection control protocols, online consultation, use of hypofractionated radiotherapy and provision of

Variable Importance for Random Forest Model

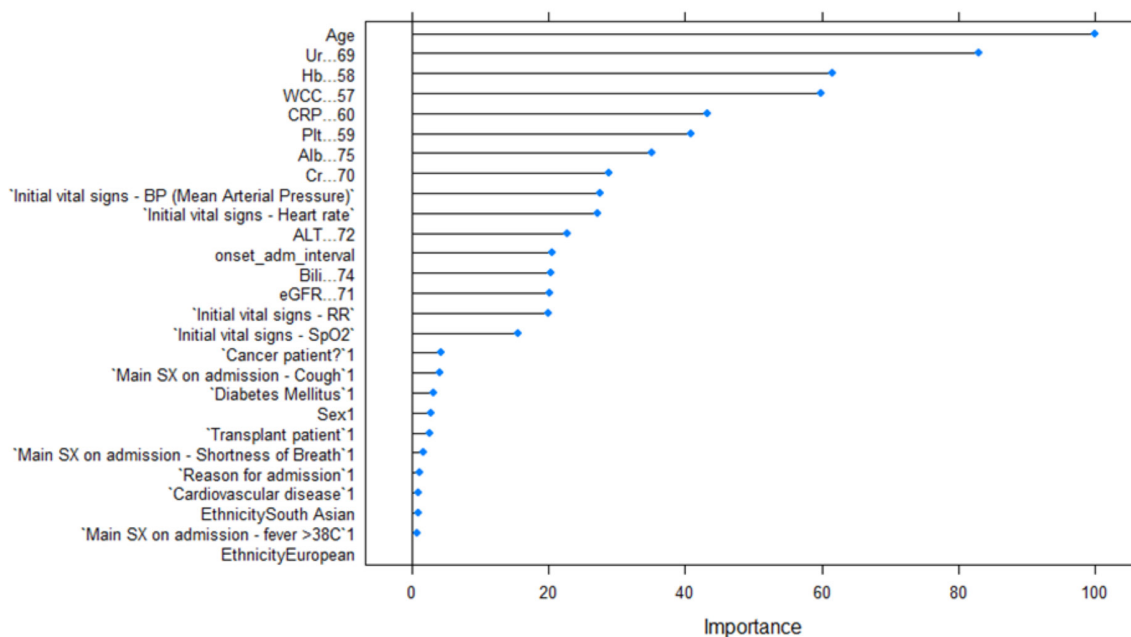


Figure 3 Relative importance of each variable in the machine learning algorithm in determining outcome from COVID-19 infection. Alb, albumin; ALT, alanine transaminase; Bili, bilirubin; BP, blood pressure; Cr, creatinine; CRP, C reactive protein; eGFR, estimated glomerular filtration rate; Hb, haemoglobin; Plt, platelets; RR, respiratory rate; SpO₂, oxygen saturation; SX, symptom; Ur, urea; WCC, white cell count.

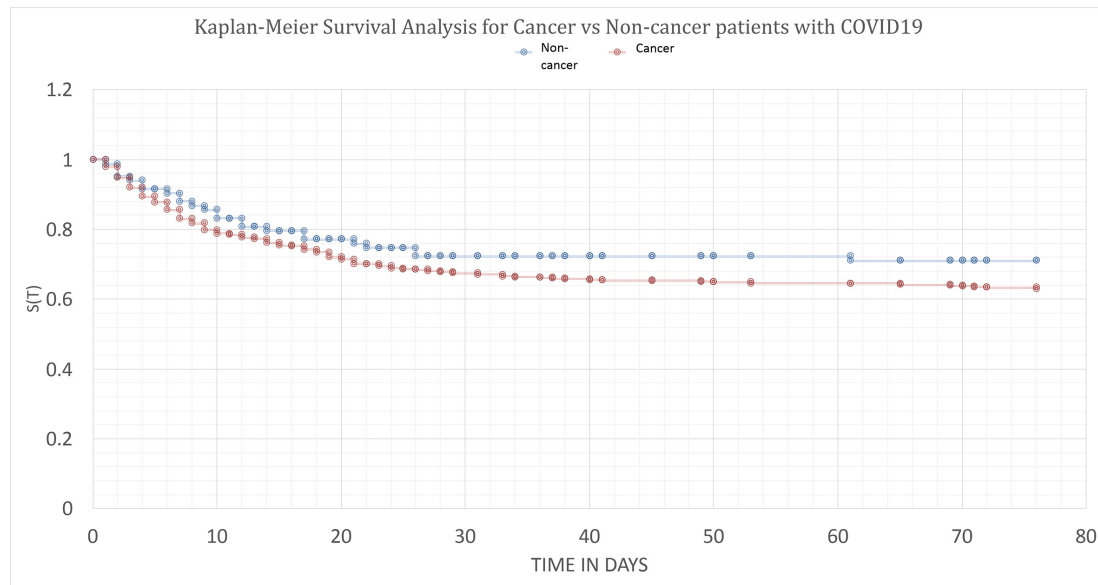


Figure 4 Kaplan-Meier survival analysis and log-rank test to determine overall survival in patients with cancer and without cancer who contracted COVID-19.

intensive care support to these patients are essential to mitigate risk.^{10 30–32} This may be supplemented where possible with COVID-19 free ‘cold’ sites to reduce risk of transmission and prevent anti-cancer treatment-induced COVID-19.³² Thus, categorisation of patients according to risk, minimising patient exposure and considering alternative regimens to control cancer forms the basis of current recommendations including the European Society for Medical Oncology (ESMO) expert consensus and UK National Health Service (NHS) guidelines.^{4 32} Furthermore, these data do not support delays in cancer treatment to reduce risk of COVID-19 transmission in patients with cancer.

Several biochemical markers have been associated with a severe COVID-19 disease course. Zhou *et al* identified a raised D-dimer above 1 µg/mL to be associated with a higher mortality risk.³³ Furthermore, they identified low albumin, raised LDH, troponin, ferritin and interleukin-6 to be more prevalent in non-survivors. In addition, raised CRP and low GFR were associated with a more severe disease outcome with 18% deaths recorded in a renal transplant cohort in keeping with the severe disease course predicted in this group of immunocompromised patients.³⁴ Our model identified raised urea and CRP in addition to transplant status as predictors of greater mortality risk which may lower threshold for admission or earlier referral for intensive care support. However, our algorithm could not specify the direction or size of this interaction which is a limitation of such models.

MLAs are increasingly being used to support healthcare applications including cancer diagnosis, outcomes and recurrence.^{16 17 35} MLA can be used to learn from established data sets and identify hidden patterns between a large number of variables to support individualised decision-making.³⁶ Nonetheless, the technique requires training data sets, appropriately selected analysis method,

as well as testing data sets to establish internal and external validity.^{16 35} MLAs have been shown to improve the accuracy of predicting cancer susceptibility, recurrence and mortality by 15%–25%.³⁷ Moreover, we have shown through our modelling that the findings of both MLA and traditional statistical analysis are complementary and may be used to generate a risk prediction scoring system in patients with cancer with COVID-19.

However, there are several limitations in the method and data presented. MLA remains an experimental technique and still very dependent on the quality of input data. Issues including noise, bias, outliers, missing or duplicate data can lead to misclassifications in any risk prediction model which may be mitigated by larger data sets.¹⁵ As such, few MLAs have achieved validation or widespread clinical application. In our study, confounders including smoking status or respiratory comorbidity were not assessed which could influence outcomes in patients with cancer. Patients with active cancer who tested positive for COVID-19 in the community but did not require hospital admission could not be evaluated. Having a broad inclusion criterion with all solid organ cancers while beneficial for looking at overall impact on patients with cancer does not capture the granularity of how individual cancers may differ in their impact on patients with COVID-19. For example, haematological malignancy, lung cancer and metastatic disease were associated with adverse outcomes from COVID-19 infection.³⁸ Our data set was underpowered to perform relevant subgroup analyses on these patients. Although all patients with active cancer were analysed, variation in the stage of cancer and treatment protocols were not accounted. Thus, the MLAs are limited by the quality of data input and rely on imputation as part of model development which needs external validation once developed, which we have not performed. Nonetheless, this study provides proof of concept to investigate

this question in a collaborative manner using larger data sets.

CONCLUSIONS

COVID-19 has impacted both individuals and healthcare systems in an enormous way. How we deliver safe and effective care to these patients in the confines of our healthcare systems is predicated on identifying those most at risk from this disease. MLAs provide an additional tool for risk assessment to delineate factors with poor prognosis. This will enable us to reconfigure our healthcare systems to provide safe care to these more vulnerable patients.

Twitter Olorunseun O Ogunwobi @seunogunwobi

Contributors AA—conceptualisation, protocol development, ethics approval, proofreading, manuscript writing and approval of final manuscript. FM—data analysis, manuscript drafting and proofreading. NA—data analysis, machine learning algorithm and figures. MR—conceptualisation and data collection. PN—conventional statistical analysis. OOO—conceptualisation, reading and correcting manuscript drafts, and approval of final manuscript. AA and OOO are responsible for overall content and serve as guarantors.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. However, OOO is supported by the National Cancer Institute (grant # CA221704).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by Russells Hall Ethics Committee (REC Reference 20/EE/0139; IRAS ID28233). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Akinfemi Akingboye <http://orcid.org/0000-0002-0325-8400>

Olorunseun O Ogunwobi <http://orcid.org/0000-0003-3388-2137>

REFERENCES

- Huang C, Wang Y, Li X, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497–506.
- Holshue ML, DeBolt C, Lindquist S, *et al*. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020;382:929–36.
- Emami A, Javanmardi F, Pirbonyeh N, *et al*. Prevalence of underlying diseases in hospitalized patients with COVID-19: a systematic review and meta-analysis. *Arch Acad Emerg Med* 2020;8:e35.
- Curigliano G, Banerjee S, Cervantes A, *et al*. Managing cancer patients during the COVID-19 pandemic: an ESMO multidisciplinary expert consensus. *Ann Oncol* 2020;31:1320–35.
- Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. *BMJ* 2020;368:m1198.
- 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;323:2052–9.
- MacMillan Trust. Macmillan Cancer Support Cancer in numbers. [Internet]. Macmillan Cancer Support, 2020. Available: <https://www.macmillan.org.uk/about-us/media-centre/facts-and-figures/cancer-in-numbers.html>
- UK Coronavirus Cancer Monitoring Project team. The UK coronavirus cancer monitoring project: protecting patients with cancer in the era of COVID-19. *Lancet Oncol* 2020;21:622–4.
- Liang W, Guan W, Chen R, *et al*. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *Lancet Oncol* 2020;21:335–7.
- Abdihamid O, Cai C, Kapesa L, *et al*. The landscape of COVID-19 in cancer patients: prevalence, impacts, and recommendations. *Cancer Manag Res* 2020;12:8923–33.
- Luo W, Phung D, Tran T, *et al*. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323.
- Scott IA. Demystifying machine learning: a primer for physicians. *Intern Med J* 2021;51:1388–400.
- Sarijaloo F, Park J, Zhong X, *et al*. Predicting 90 day acute heart failure readmission and death using machine learning-supported decision analysis. *Clin Cardiol* 2021;44:230–7.
- Ibrahim A, Gamble P, Jaroensri R, *et al*. Artificial intelligence in digital breast pathology: techniques and applications. *Breast* 2020;49:267–73.
- Kourou K, Exarchos TP, Exarchos KP, *et al*. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015;13:8–17.
- Nghiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019;20:e262–73.
- Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol* 2020;7:e541–50.
- Ravaud M, Harish V, Sadeghi H, *et al*. Development and validation of a machine learning model using administrative health data to predict onset of type 2 diabetes. *JAMA Netw Open* 2021;4:e2111315.
- Zarifkar P, Kamath A, Robinson C, *et al*. Clinical characteristics and outcomes in patients with COVID-19 and cancer: a systematic review and meta-analysis. *Clin Oncol* 2021;33:e180–91.
- Fernandes GA, Feriani D, França E Silva ILA, *et al*. Differences in mortality of cancer patients with COVID-19 in a Brazilian cancer center. *Semin Oncol* 2021;48:171–80.
- Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-Level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environ Res* 2020;188:109890.
- Russell B, Moss CL, Shah V, *et al*. Risk of COVID-19 death in cancer patients: an analysis from Guy's cancer centre and King's College hospital in London. *Br J Cancer* 2021;125:939–47.
- Lee KA, Ma W, Sikavi DR, *et al*. Cancer and risk of COVID-19 through a general community survey. *Oncologist* 2021;26:e182–5.
- Pinato DJ, Scotti L, Gennari A, *et al*. Determinants of enhanced vulnerability to coronavirus disease 2019 in UK patients with cancer: a European study. *Eur J Cancer* 2021;150:190–202.
- Mehta V, Goel S, Kabarriti R, *et al*. Case fatality rate of cancer patients with COVID-19 in a new York hospital system. *Cancer Discov* 2020;10:935–41.
- Liu C, Zhao Y, Okwan-Duodu D, *et al*. COVID-19 in cancer patients: risk, clinical features, and management. *Cancer Biol Med* 2020;17:519–27.
- Vuagnat P, Frelaut M, Ramtohl T, *et al*. COVID-19 in breast cancer patients: a cohort at the Institut Curie hospitals in the Paris area. *Breast Cancer Res* 2020;22:55.
- The Lancet Oncology. COVID-19: global consequences for oncology. *Lancet Oncol* 2020;21:p. 467.
- Sud A, Torr B, Jones ME, *et al*. Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: a modelling study. *Lancet Oncol* 2020;21:1035–44.
- Al-Quteimat OM, Amer AM. The impact of the COVID-19 pandemic on cancer patients. *Am J Clin Oncol* 2020;43:452–5.
- Guckenberger M, Belka C, Bezjak A, *et al*. Practice recommendations for lung cancer radiotherapy during the COVID-19 pandemic: an ESTRO-ASTRO consensus statement. *Radiotherapy and Oncology* 2020;146:223–9.

- 32 van de Haar J, Hoes LR, Coles CE, *et al.* Caring for patients with cancer in the COVID-19 era. *Nat Med* 2020;26:665–71.
- 33 Zhou F, Yu T, Du R, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
- 34 Sran K, Olsburgh J, Kasimatis T, *et al.* COVID-19 in kidney transplant patients from a large UK transplant center: exploring risk factors for disease severity. *Transplant Proc* 2021;53:1160–1168. Dec.
- 35 Huang S, Yang J, Fong S, *et al.* Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett* 2020;471:61–71.
- 36 Zimmerman A, Kalra D. Usefulness of machine learning in COVID-19 for the detection and prognosis of cardiovascular complications. *Rev Cardiovasc Med* 2020;21:345–52.
- 37 Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007;2:59–77.
- 38 Dai M, Liu D, Liu M, *et al.* Patients with cancer appear more vulnerable to SARS-CoV-2: a multicenter study during the COVID-19 outbreak. *Cancer Discov* 2020;10:783–91.