Investigating factors that may influence COVID-19 survival

36112985

ABSTRACT

Objective *To investigate, amongst a set of probable prognostic factors, which factors may influence coronavirus 19 disease survival probabilities.*

Design *Prospective cohort study; ongoing.*

Setting *The International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) World Health Organization (WHO) Clinical Characterisation Protocol UK (CCP-UK) study, which enrols in-patients at at-least 260 England, Scotland, and Wales hospitals; the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC 4C) conducts the study. The investigation's patients are patients enrolled in a participating England hospital between 10 February 2020 and 5 July 2020.*

Participants *Each study enrollee has a confirmed SARS-CoV-2 infection, or a high infection likelihood. The investigation's patients are members of age groups* 30 - 39*, 40 - 49, and higher.*

Outcome Patient survival.

Results Inconclusive.

Conclusions Insufficient spectrum of factors.

Introduction

During the past 2 years, infections due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have led to a high number of coronavirus 19 disease deaths and in-hospital care. With time, case fatality rates, and in-hospital survival rates, have improved [. . .] as our understanding of coronavirus 19 disease pathogenesis has developed. Consequently, disease prevention, mitigation, and treatment measures have improved.

Effective disease treatment measures depend on our understanding of the disease's prognosis factors (Riley 2013), the focus of prognostic factors research. Prognostic factor research *aims to identify factors associated with subsequent clinical outcome in people with a particular disease or health condition* (Riley 2013).

This brief investigates the prognostic factor potential of a set of measures associated with a prospective cohort study. The study's hypothesis is that *one or more of the measures has a significant impact on each patient's survival probability.* The ISARIC 4C study outcomes are: discharged alive, remains in hospital, death, transferred, and palliative discharge. Herein, the overarching outcomes of interest are: dead, alive. The predictors, factors, under investigation

Methods

Study design and setting

This investigation's data is an indirect ISARIC (International Severe Acute Respiratory Infection Consortium) 4C (Coronavirus Clinical Characterisation Consortium) study data set. The study is an ongoing prospective

cohort study based in more than 200 United Kingdom acute care hospitals. This investigation's data is an England only data set. The reporting ...

Participants

Inclusion criteria were people of all ages who were admitted to one of 208 acute care hospitals in England, Scotland, and Wales with proven or high likelihood of infection with a pathogen of public health interest, defined as SARS-CoV-2 for this event by Public Health England. Reverse transcriptase polymerase chain reaction was the only mode of testing available during the period of study. The decision to test was at the discretion of the clinician attending the patient, and not defined by protocol. The enrolment criterion "high likelihood of infection" reflects that a preparedness protocol cannot assume that a diagnostic test will be available for an emergent pathogen. Site training emphasises that only patients who tested positive for covid-19 were eligible for enrolment.

National guidance was provided by Public Health England and other UK public health agencies that advised who to test based on clinical case definitions for possible covid-19. We also included patients who had been admitted for a separate condition but had tested positive for covid-19 during their hospital stay. Patients were only enrolled during their index admission. Patients had clinical information from their routine health records uploaded into the case report form. Consent was not required for collection of depersonalised routine healthcare data for research in England and Wales. A waiver for consent was given by the Public Benefit and Privacy Panel in Scotland.

Data

Although the consortium collects a large variety of demographic, clinical, and outcome data, this investigation's data only included a limited subset of the study's variables. The demographic data variables are sex and age group, the comorbidity variables are asthma (physician diagnosed), mild liver disease, moderate/severe liver disease, renal disease, chronic pulmonary disease (excluding asthma), chronic neurological disorder, malignant neoplasm.

The ISARIC 4C study collects its baseline data via paper case report forms developed by ISARIC, and the WHO (World Health Organization), for outbreak investigations. Additionally, the data brief states that the baseline data records are *uploaded [on] admission, and usually before hospital episodes were complete, to a REDCap database (Research Electronic Data Capture, Vanderbilt University, US, hosted by University of Oxford, UK).*

Exploratory Data Analysis

- Relationships
- Correlations

Missing Data

Most of the data's variables have missing values. Prior to deciding how to address missing values, it is important to understand the missing values patterns (Steyerberg 2010). Rubin Little and Rubin (2019) outlines three fundamental missing data mechanisms

- missing completely at random (MCAR) / administrative errors, accidents

- missing at random (MAR) / missing data associated with known patient characteristics [independent variables], or the outcome
- missing not at random (MNAR) / missing data associated with missing values of the factor/predictor in question or with unobserved predictors

If the missing values of a data set in question are *missing completely at random* then complete case analysis will suffice because the complete case excerpt is akin to a random sample from a complete population. If MCAR does not hold, e.g., data is *missing at random*, then the complete case excerpt is not representative of the underlying population, therefore population inference is not possible via complete case analysis. (Steyerberg 2010)

Herein, missing data mechanisms analysis is via logistic regression. Almost all the independent variables have missing values patterns that violate MCAR; addressed via multiple imputation in-line with the advice and protocols of (Sterne et al. 2009) & (Steyerberg 2010).

Factor Analysis

The Cox Hazard, Restricted Mean Survival

Results

Characteristics

Correlation graph

Missing Data

The Null Kaplan-Meier Curves

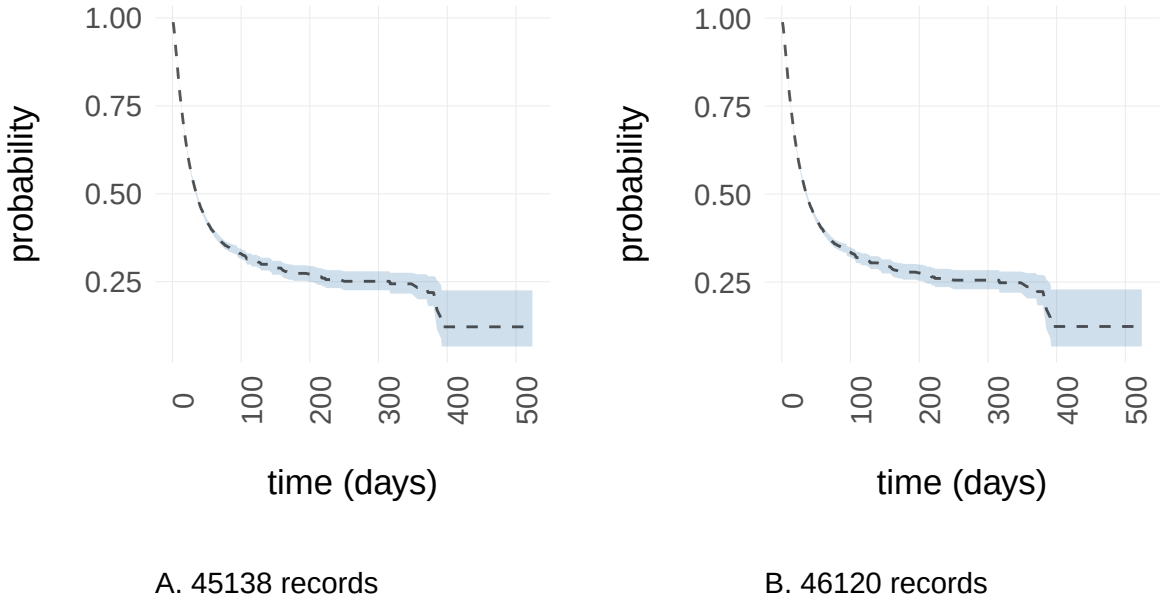

Figure 1: The null survival curves, whereby (A) is based on complete cases only, whilst (B) is based on all plausible records after multiple imputation.

In Hospital Mortality

	no	**yes**	**NA**
deceased	31724	14049	347
	0.688	0.305	0.008

Comorbidities

	no	**yes**	**NA**
asthma	37376	5872	2872
	0.81	0.13	0.06
mild liver disease	42021	666	3433
	0.91	0.01	0.07
renal disease	35158	8056	2906
	0.76	0.17	0.06
pulmonary	35265	8078	2777
	0.76	0.18	0.06
neurological disorder	37227	5806	3087
	0.81	0.13	0.07
liver mod severe	41901	902	3317
	0.91	0.02	0.07
malignant neoplasm	38164	4734	3222
	0.83	0.10	0.07

Demographics

Age Group	**Male**	**Female**	**NA**
80-89	7067	6197	31
	0.153	0.134	0.001
70-79	6388	4445	19
	0.139	0.096	0.000
60-69	4321	2601	9
	0.094	0.056	0.000
50-59	3463	2151	9
	0.075	0.047	0.000
90+	2085	2816	13
	0.045	0.061	0.000
40-49	1846	1110	5
	0.040	0.024	0.000
30-39	784	756	4
	0.017	0.016	0.000

Uni-variate & Multivariate Analysis: Training

The results of Table @ref(tab:univariateTraining) summarise the separate effects of each potential prognostic factor on survival probability; investigated via Cox Proportional Hazard uni-variate analysis → coefficient: regression coefficients, hazard ratio & confidence interval: effect size, p_{value}: statistical significance in relation to survival.

In terms of age groups, the reference age group is *80 – 89*, and their p_{value} scores suggests that **(a)** each age group, except *90+*, has a lower risk of death compared to age group *80 – 89*, and **(b)** the risk of death increases with age. The sex p_{value} suggests that females have a lower risk of death compared to males. The p_{value} scores for patients with an asthma, renal disease, pulmonary disease, or malignant neoplasm, diagnosis suggests that these patients have a higher risk of death compared to their respective disease/disorder free counterparts.

Alas, the proportionality tests invalidate most of the preceding observations; age group and malignant neoplasm have a score test value below 0.05, the renal disease score test value is approximately 0.05, and neurological disorder score is barely greater than 0.05. The multivariate analysis confirms the trend.

Uni-variate & Multivariate Analysis: All

Due to the failure of the Cox Proportional Hazard Modelling & Analysis strategy, ...

Table 1: Univariate Analysis

			hazard ratio CI			
	coefficient	hazard ratio	lower	upper	p_{value}	U
Age Group 30 - 39	-1.586	0.205	0.162	0.259	0.000	0.000
Age Group 40 - 49	-1.377	0.252	0.219	0.291	0.000	0.000
Age Group 50 - 59	-0.856	0.425	0.393	0.460	0.000	0.000
Age Group 60 - 69	-0.442	0.643	0.605	0.682	0.000	0.000
Age Group 70 - 79	-0.133	0.876	0.836	0.918	0.000	0.000
Age Group 90+	0.128	1.137	1.074	1.203	0.000	0.000
Asthma Yes	-0.097	0.907	0.857	0.961	0.001	0.265
Liver Disease, (Mild) Yes	-0.106	0.899	0.763	1.059	0.205	0.792
Liver Disease, (Moderate, severe) Yes	0.078	1.081	0.951	1.229	0.232	0.439
Malignant Neoplasm Yes	0.156	1.168	1.104	1.236	0.000	0.012
Neurological Disorder Yes	0.097	1.101	1.045	1.161	0.000	0.056
Pulmonary Disease Yes	0.223	1.250	1.195	1.308	0.000	0.368
Renal Disease Yes	0.280	1.324	1.266	1.384	0.000	0.052
Sex Female	-0.202	0.817	0.786	0.848	0.000	0.813

[1] CI: Confidence Interval
[2] U: Proportionality score test

Adjusted Analysis via Restricted Mean Survival Time

The proportional hazard assumption does not hold ...

Discussion

The Cox Model Assumptions

The observations are independent: This assumption requires a set of observations wherein the patients are not related or associated. If there are clusters of relations or associates, the clusters should be, must be, indicated as-such. By virtue of the > study's design, it is quite possible that the observations are not independent - families, friends, and colleagues living in the same area will probably be admitted to the same health centre. Alas, the data set does not have cluster indicators, hence it is impossible to definitively test and state the independence, or otherwise, of the observations.

Independent Censoring: There is neither evidence to suggest nor rule-out outcomes due to systematic drop-outs, induced by study design flaws. What are the prognosis of patients released to palliative care or transferred? Is the decision to transfer or release to palliative care influenced or dictated by *lower* survival probability? If the answer to the latter question is yes - it will be in breach of independent censoring.

Proportional Hazards: The curves.

Table 2: Multivariate Analysis

			hazard ratio CI			
	coefficient	hazard ratio	lower	upper	p_{value}	U
Age Group 30 - 39	-1.563	0.209	0.166	0.265	0.000	0.000
Age Group 40 - 49	-1.381	0.251	0.218	0.290	0.000	0.000
Age Group 50 - 59	-0.858	0.424	0.391	0.459	0.000	0.000
Age Group 60 - 69	-0.458	0.633	0.596	0.672	0.000	0.000
Age Group 70 - 79	-0.151	0.860	0.821	0.901	0.000	0.000
Age Group 90+	0.156	1.169	1.104	1.237	0.000	0.000
Asthma Yes	0.003	1.003	0.947	1.063	0.921	0.383
Liver Disease, (Mild) Yes	-0.067	0.935	0.793	1.101	0.421	0.660
Liver Disease, (Moderate, severe) Yes	0.227	1.255	1.104	1.427	0.001	0.270
Malignant Neoplasm Yes	0.045	1.046	0.989	1.107	0.118	0.010
Neurological Disorder Yes	0.046	1.047	0.993	1.103	0.088	0.120
Pulmonary Disease Yes	0.114	1.121	1.070	1.173	0.000	0.368
Renal Disease Yes	0.149	1.160	1.109	1.214	0.000	0.075
Sex Female	-0.289	0.749	0.721	0.779	0.000	0.226

[1] CI: Confidence Interval
[2] U: Proportionality score test

Missing Data

Within a missing values setting, there are 2 options w.r.t. predictor effect analysis (Steyerberg 2010)

- complete case analysis for the uni-variate analysis, and complete case predictors & imputed confounding variables for adjusted analysis.
- uni-variate and adjusted analysis after imputation of all missing values.

The investigation opted for the latter, especially because ...

Bias

Possible bias points

- By virtue of the study's design pre-study–admission differences between hospitals is possible due to the ambiguity of the study's patient recruitment terms, and this might lead to post-study differences. In a nutshell, the study is susceptible to selection bias. [Trochim]
- Missing data (Sterne et al. 2009)

Conclusion

The analysis of missing values hints at flaws in data collection.

Table 3: The disease frequencies

Comorbidity	No	Yes	Unknown	Total
Asthma	37376	5872	2872	46120
Mild Liver Disease	42021	666	3433	46120
Moderate, Severe Liver Disease	41901	902	3317	46120
Malignant Neoplasm	38164	4734	3222	46120
Neurological Disorder	37227	5806	3087	46120
Pulmonary Disease	35265	8078	2777	46120
Renal Disease	35158	8056	2906	46120

Appendices

Exploratory Analysis

Graphs

Missing Data

Clusters, patterns, associations.

References

Little, Roderick, and Donald B. Rubin. 2019. *Statistical Analysis with Missing Data, Third Edition*. Wiley. https://doi.org/10.1002/9781119482260.

Riley, Jill A. AND Steyerberg, Richard D. AND Hayden. 2013. "Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research." *PLOS Medicine* 10 (2): 1–9. https://doi.org/10.1371/journal.pmed.1001380.

Rubin, Donald B. 1976. "Inference and missing data." *Biometrika* 63 (3): 581–92. https://doi.org/10.1093/biomet/63.3.581.

Sterne, Jonathan A C, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. 2009. "Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls." *BMJ* 338. https://doi.org/10.1136/bmj.b2393.

Steyerberg, Ewout W. 2010. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer. https://doi.org/10.1007/978-1-4419-2648-7.