

# Strategy to identify priority groups for COVID-19 vaccination: A population based cohort study



Antonio Giampiero Russo<sup>a,\*</sup>, Adriano Decarli<sup>a</sup>, Maria Grazia Valsecchi<sup>b</sup>

<sup>a</sup>Epidemiology Unit, Agency for Health Protection of the Metropolitan Area of Milan, Milan, Italy

<sup>b</sup>Bicocca Center of Bioinformatics, Biostatistics and Bioimaging (B4 Center), School of Medicine and Surgery, University of Milan-Bicocca, Milan, Italy

## ARTICLE INFO

### Article history:

Received 5 February 2021  
Received in revised form 19 March 2021  
Accepted 22 March 2021  
Available online 26 March 2021

### Keywords:

COVID vaccination  
Public Health  
Model selection  
LASSO regression  
Stratification for priorities

## ABSTRACT

**Background:** Evidence from COVID-19 outbreak shows that individuals with specific chronic diseases are at higher risk of severe prognosis after infection. Public health authorities are developing vaccination programmes with priorities that minimize the risk of mortality and severe events in individuals and communities. We propose an evidence-based strategy that targets the frailest subjects whose timely vaccination is likely to minimize future deaths and preserve the resilience of the health service by preventing infections.

**Methods:** The cohort includes 146,087 cases with COVID-19 diagnosed in 2020 in Milan (3.49 million inhabitants). Individual level data on 42 chronic diseases and vital status updated as of January 21, 2021, were available in administrative data. Analyses were performed in three sub-cohorts of age (16–64, 65–79 and 80+ years) and comorbidities affecting mortality were selected by means of LASSO cross-validated conditional logistic regression. Simplified models based on previous results identified high-risk categories worth targeting with highest priority. Results adjusted by age and gender, were reported in terms of odds ratios and 95%CI.

**Results:** The final models include as predictors of mortality (7,667 deaths, 5.2%) 10, 12, and 5 chronic diseases, respectively. The older age categories shared, as risk factors, chronic renal failure, chronic heart failure, cerebrovascular disease, Parkinson disease and psychiatric diseases. In the younger age category, predictors included neoplasm, organ transplantation and psychiatric conditions. Results were consistent with those obtained on mortality at 60 days from diagnosis (6,968 deaths).

**Conclusion:** This approach defines a two-level stratification for priorities in the vaccination that can easily be applied by health authorities, eventually adapted to local results in terms of number and types of comorbidities, and rapidly updated with current data. After the early phase of vaccination, data on effectiveness and safety will give the opportunity to revise prioritization and discuss the future approach in the remaining population.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

On February 19, 2020, the first case of Coronavirus Disease 2019 (COVID-19) was identified in Italy, in the province of Lodi. The first case occurred in a non-metropolitan area, which refers to the Agency for Health Protection (ATS) of Milan, in the municipality of Codogno (approximately 15,000 inhabitants), and had an initial non-specific set of symptoms that delayed the diagnosis of COVID-19. The infection was seen in 46 close contacts, including the healthcare personnel of the Codogno hospital, and this started

the first phase of the COVID-19 pandemic in Italy. Since then, the ATS of Milan is constantly monitoring the pandemic (<https://www.ats-milano.it/portale/Epidemiologia/Valutazione-dellepidemia-COVID-19>).

In the first weeks of March, when the epidemic was growing exponentially, the ATS of Milan implemented an algorithm that stratified patients according to their clinical and demographic characteristic associated with an increased risk of dying. The model was calibrated on 2,981 COVID-19 cases as of March 13, 2020, of which 435 (14%) patients died by March 15, 2020. Characteristics statistically associated with the risk of death were: age 70 years and more, presence of neurological disorders (dementia, Alzheimer's disease, Parkinson's disease), heart failure, ischemic

\* Corresponding author at: Epidemiology Unit, Agency for Health Protection of Milan, C.so Italia, 52, 20122 Milano, Italy.

E-mail address: [agrusso@ats-milano.it](mailto:agrusso@ats-milano.it) (A.G. Russo).

cardiomyopathy, valvular disease, renal failure (including dialysis), and neoplasm diagnosed in the last two years [1].

This model generated a simple caretaking and surveillance system adopted by a group of GPs who actively operated in contacting and monitoring by telephone those patients classified at high risk of dying. The GPs routinely ascertained their clinical conditions, prescribed appropriate treatments, updated the risk of severe outcome and required hospitalization in the most compromised ones. This model of caretaking showed a high-impact in preventing a severe course of the disease. Infected patients monitored by this group of GPs had a 50% decrease in the risk of dying and a 70% decrease in the risk of hospitalization, compared to similar high-risk patients not monitored according to the surveillance system [2].

Worldwide, multiple variables have been described as possible risk factors for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) susceptibility, severity and prognosis, among which age, sex and comorbidities play an important role. The Center for Disease Control and Prevention (CDC) routinely updates evidence on the underlying medical conditions that have shown to increase the risk of severe illness from SARS-CoV-2 (<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html>). In UK the health analytics platform OpenSAFELY, that covers 40% of all patients in England, was created to examine factors associated with COVID-19-related death. Obesity, recent history of solid neoplasm and haematological malignancy, diabetes, severe asthma, respiratory disease, chronic heart disease, liver disease, stroke, dementia, other neurological diseases, reduced kidney function, autoimmune diseases and other immunosuppressive conditions were associated with a higher risk of COVID-19-related death [3]. Evidence from the global outbreak has demonstrated that individuals with pre-existing cardiovascular diseases, type II diabetes, cancer and COPD are at a greater risk of a severe course of the infection or death after COVID-19 diagnosis [4–7].

The identification of individual predisposing factors is important not only to define different surveillance procedures for frail groups of infected people but it is currently necessary to develop vaccination strategies [3,8]. Indeed, as currently the number of available doses cannot cover the target population, health policy strategies that account for the differential risk of individuals and communities are needed for an equitable distribution of vaccines.

The goals of the COVID-19 vaccination program are 1) ensure safety and effectiveness of COVID-19 vaccines; 2) reduce transmission (susceptibility), morbidity (severity), mortality of COVID-19 infection; 3) help minimize disruption to society and economy, including maintaining healthcare capacity and; 4) ensure equity in vaccine allocation and distribution [9,10].

First phase (Phase 1a) of the vaccination program in Italy is directed to all high-risk health workers, frontline health and social care workers, and residents in care homes for older adults and their careers [11]. Phase 1b includes subjects 80 years old or more (80+) and subjects 65 years old or more at significantly higher risk of a severe course because of comorbid and underlying conditions. Phase 1c: subjects aged 16–64 years with high-risk medical conditions, and other essential workers. Phase 2 will see the mass vaccination campaign for all other subjects aged 16 years or older. However, the health policy question is: “which high-risk groups should be offered first COVID-19 vaccination in Phase 1b and 1c” [11].

People with specific comorbid and underlying conditions are at high risk for COVID-19 mortality. The pathway from infection to mortality includes hospitalization and/or access to intensive care. An early access to vaccination may reduce the need of contact with the health system and the hospital, the risk of a severe course of the disease and of death. A prioritization plan for vaccination needs evidence that can be promptly derived and easily implemented by

health authorities based on data routinely collected during the pandemic at the population level.

We propose a simple algorithm, based on the pandemic information collected on a cohort of more than 3 million people referred to the ATS of Milan, that allows to stratify the population in different risk sets, within age classes, suitable for the prioritization plan in the first crucial phases of COVID vaccination.

## 2. Materials and methods

### 2.1. Data source

All cases identified from February to November 2020 were 165,382. Of them 11,716 were excluded because are less than 16 years and 7,579 are resident on long-term care facilities. Therefore, the cohort includes a total of 146,087 consecutive cases reported in the COVID-19 database of the ATS of Milan, who were diagnosed in the study area with a positive nasopharyngeal swab (based on polymerase chain reaction, PCR) from February 19 to November 30, 2020. The study area, covered by the ATS of Milan, includes 193 municipalities in the Northern Italian region of Lombardy, with a total population of 3.49 million inhabitants, of whom approximately 3 million aged 16 years or older.

Individual level comorbidities data were reported in the COVID-19 database using a validated algorithm that defines chronic diseases as specified in the Regional Act X/616418 and X/765519 of 2017. This algorithm is routinely used in the region of Lombardy but can be replicated in any setting based on administrative data.

Vital status was derived from the early notification system of the ATS of Milan, set-up from the beginning of the epidemic in a specific web based information system. All deaths were captured by linking the following information sources: 1) Civil Register of each Municipality; 2) Register of Causes of Death by death certificates completed by physicians, coroners, or medical examiners; 3) reporting of Funeral homes and 4) GPs and Mayor's offices that used a specific web based information system for reporting the COVID cases. The vital status was assessed as of January 21, 2021.

### 2.2. Statistical methods

In order to comply with the phases defined in the vaccination program, the analyses were separately performed in three sub-cohorts based on age: 16–64, 65–79 and 80+. Starting from 42 specific chronic diseases which were present in the infected population, the risk factors potentially prognostic for mortality were selected by means of a LASSO conditional logistic regression approach based on a 5-fold cross-validation in order to avoid overfitting [12].

Results of the logistic regression models that included the selected risk factors, adjusted by age (continuous variable in years) and gender, were reported in terms of odds ratios (ORs) and corresponding 95% confidence intervals in forest plot type graphs. The area under the ROC curve (up to 0.5 (as well as chance), 0.7 to 0.8 (acceptable), 0.8 to 0.9 (excellent), and 0.9 to 1.0 (outstanding classification)) was used to describe the discriminatory ability of the models.

Sensitivity analyses were conducted with the same approaches considering the first (from February 19 to May 31, 2020) and second wave (from October 1 to November 30, 2020) of the epidemic and mortality within 60 days from the date of the diagnostic positive swab. Simplified logistic regression models, aimed at defining priority strata easy to be implemented in practice at the population level, evaluated the subjects with the conditions selected in the final model, as part of the high risk subgroup. The remaining subjects were classified at low risk and an additional sensitivity analysis also

considered the definition of a medium risk group for subjects who presented other non selected comorbidities. Results of these stratification models were reported in terms of ORs and 95% CI.

The analysis were done by SAS Package, Procedure GLMSELECT (SAS/STAT User Guide, 2019) [13].

### 3. Results

#### 3.1. Characteristics of the population

The cohort includes 146,087 nasopharyngeal swab confirmed cases of SARS-CoV-2, of whom 7,667 (5.2%) died. In particular, classes 16–64, 65–79 and 80+ years of age included: 115,115 subjects, with 693 deaths (0.6%), 19,854 subjects, with 2,782 deaths (14.0%) and 11,118 subjects, with 4,192 deaths (37.7%), respectively. As shown in Table 1, in each age class, mortality was higher in males compared to females. A total of 89,982 subjects (61.6%) did not present a chronic condition among those listed in Table 2. The remaining COVID-19 cases had an increasing number of multiple comorbidities with increasing age and an increased mortality rate with increasing number of comorbidities in each age class. Fig. 1 reports the daily distribution of deaths from February 19, 2020 to January 21, 2021, and shows quite similar absolute daily frequencies during the two pandemic phases, with highest values around 100 deaths per day, but a less sharp decrease after the peak in the second wave. The same graph shows the daily number of incident COVID-19 cases, highlighting how the pandemic had a much higher impact in the second wave in the area of the ATS of Milan.

Table 2 shows the distribution of the 42 individual comorbidities which were present in the population cohort of the infected individuals, by age class. Hypertension, hypercholesterolemia and diabetes, represent three chronic conditions with the highest prevalence in the population of COVID-19 cases.

Ischemic heart disease and psychiatric conditions (defined by the codes ICD-10: F00–F99) also have a high prevalence in this population. However, the distribution of other specific conditions varied by age group. The more frequent conditions (above 3% prevalence) were psychiatric conditions, COPD, and hypothyroidism under 65 years; patient with neoplasm, chronic heart failure, ischemic heart disease, COPD, cardiomyopathy with or without arrhythmia and hypothyroidism were relatively frequent (above 6%) in the age class 65–79; patients with neoplasm, chronic heart failure, ischemic heart disease, cardiomyopathy with or without arrhythmia and COPD were common (above 10% prevalence) in subjects aged 80 years or more.

#### 3.2. Regression models on factors influencing mortality

Results of the final regression models on mortality, in each age group, are summarizes in Table 3 (first row, overall results) and Figs. 2.

**Table 1**  
Distribution by gender and number of comorbidities according to age class (ATS of Milan, 2021).

	16–64			65–79			80+			Total
	N. deaths/N. cases	%	OR (95% CI)	N. deaths/N. cases	%	OR (95% CI)	N. deaths/N. cases	%	OR (95% CI)	
Gender										
Females	202/59,256	0.34	1#	815/8,969	9.09	1#	1,857/5,873	31.62	1#	<b>74,098</b>
Males	491/55,859	0.88	2.49 (2.11–2.94)	1967/10,885	18.07	2.20 (2.01–2.40)	2,335/5,245	44.52	1.91 (1.76–2.07)	<b>71,989</b>
N. of comorbidities										
0	215/84,193	0.26	1#	386/4,442	8.69	1#	434/1,347	32.2	1#	<b>89,982</b>
1	166/21,241	0.78	1.62 (1.31–1.99)	631/5,845	10.80	1.11 (0.97–1.27)	1,006/2,866	35.1	1.13 (0.98–1.30)	<b>29,952</b>
2–3	211/8,405	2.51	3.95 (3.23–4.84)	1,036/6,922	14.97	1.48 (1.30–1.68)	1,720/4,665	36.9	1.20 (1.05–1.37)	<b>19,992</b>
4+	101/1,175	7.92	11.01 (8.52–14.23)	729/2,645	27.56	2.93 (2.55–3.37)	1,032/2,240	46.1	1.76 (1.52–2.04)	<b>6,161</b>
Total	<b>693/115,115</b>			<b>2,782/19,854</b>			<b>4,192/11,118</b>			<b>146,087</b>

In the patient population aged 16–64, the final logistic regression model includes as independent conditions associated with mortality 10 chronic conditions, with a good discriminatory ability of model (area under the ROC curve 0.88). Patients who have a neoplasm and are in front line treatment have a 5-fold increase in the odds of death compared to those without this condition during the COVID-19 infection. The other risk factors, which carry approximately a 2-fold increase in the risk of death, are the experience of a solid organ transplantation, the presence of diabetes, of chronic renal failure and of peripheral artery disease. The other conditions (chronic heart failure, cardiomyopathy with arrhythmia and ischemic heart disease) have between 50 and 70% relative increase in the odds of death. Interestingly, presence of cardiomyopathy without arrhythmia and hypertension carry a limited increase in the odds of death, which is borderline significant, as shown by the confidence intervals: estimated OR of 1.38, 95% CI of 1.01 – 1.88 and OR of 1.22, 95% CI of 1.01 – 1.47, respectively (Table 3 and Fig. 2, panel A).

In the patient population aged 65–79, the final logistic regression model includes as independent factors associated with mortality 12 chronic conditions, with an area under the ROC curve of 0.73. Approximately, a significant 2-fold increase in the odds of death is related to the presence of chronic renal failure, Parkinson’s disease, Alzheimer and dementias and psychiatric conditions. All other selected conditions carry an estimated OR between 1.3 and 1.7, significantly affecting mortality (Table 3 and Fig. 2, panel B).

In the patient population aged 80 or more years, the results of the final logistic regression model show as independent factors associated with mortality 5 chronic conditions, with an area under the ROC curve of 0.67. Apart from the presence of psychiatric conditions, which are related to an estimated OR of 1.96 (95% CI of 1.69 – 2.28), the remaining factors (Parkinson’s disease, chronic renal failure, chronic heart failure, ischemic heart disease) have a less marked but significant impact on mortality, with estimated ORs between 1.2 and 1.4 (Table 3 and Fig. 2, panel C).

#### 3.3. Sensitivity analysis

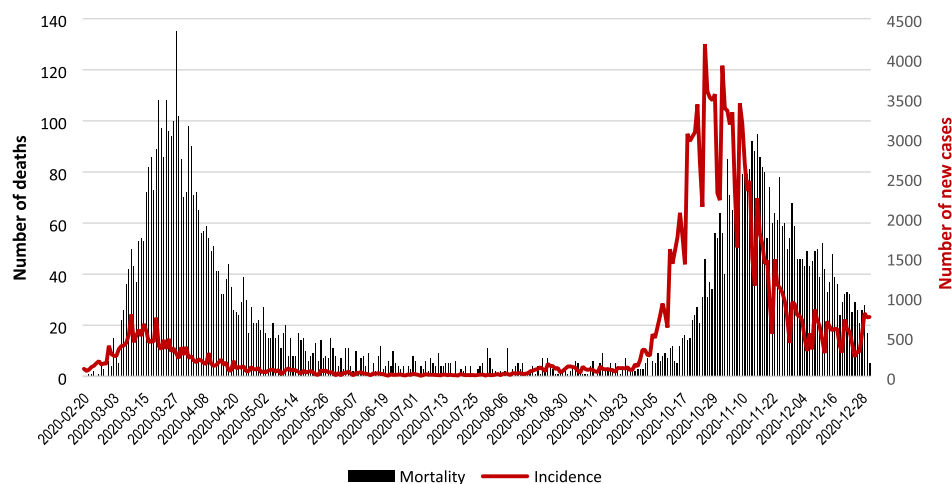
Overall results on major comorbidities impacting on mortality are strongly influenced by the second wave which presents almost 5 times the incident cases of the first one. Yet, as shown in Table 3, the two separate analyses by wave identify a common subset of relevant factors.

Of interest, when we focused on 60 days mortality (from diagnosis of COVID-19 infection), we indeed included the large majority of observed deaths (6,968/7,667) indicating that COVID-19 cases present a relevant early mortality and results of the models were very similar.

**Table 2**

Distribution of individual comorbidities according to age class (counting multiple comorbidities on the same individuals). Percentages are calculated for each single comorbidity on the total number of subjects in the age class. (ATS of Milan, 2021).

	16–65		65–79		80+		Overall	
	Dead	Alive	Dead	Alive	Dead	Alive	146,087 N.	%
	693 N.	114,422 N. (%)	2,782 N. (%)	17,072 N. (%)	4,192 N. (%)	6,926 N. (%)		
C56 Hypertension	290 (41.8)	12,219 (10.7)	1,875 (67.4)	9,567 (56.0)	3,178 (75.8)	5,049 (72.9)	<b>32,178</b>	<b>22.03%</b>
C48 Hypercholesterolemia	105 (15.1)	3,331 (2.9)	800 (28.76)	3,955 (23.2)	1,069 (25.5)	1,856 (26.8)	<b>11,116</b>	<b>7.61%</b>
C52 Diabetes	161 (23.2)	4,050 (3.5)	836 (30.0)	3,064 (17.9)	1,009 (24.1)	1,396 (20.2)	<b>10,516</b>	<b>7.20%</b>
C37 COPD	49 (7.07)	3,765 (3.29)	433 (15.56)	1,536 (9.00)	611 (14.58)	859 (12.40)	<b>7,253</b>	<b>4.96%</b>
C27 Ischemic heart disease	85 (12.3)	1,369 (1.2)	691 (24.8)	2,319 (13.6)	1,137 (27.1)	1,397 (20.2)	<b>6,998</b>	<b>4.79%</b>
C29 Cardiomyopathy with arrhythmia	58 (8.4)	1,284 (1.1)	532 (19.1)	1,855 (10.9)	1,211 (28.9)	1,574 (22.7)	<b>6,514</b>	<b>4.46%</b>
C99 Psychiatric conditions	78 (11.26)	4,592 (4.01)	253 (9.09)	734 (4.30)	415 (9.90)	400 (5.78)	<b>6,472</b>	<b>4.43%</b>
C08 Neoplasm in first line treatment	119 (17.2)	2,284 (2.0)	436 (15.7)	1,527 (8.9)	431 (10.3)	721 (10.4)	<b>5,518</b>	<b>3.78%</b>
C33 Cardiomyopathy without arrhyt.	66 (9.5)	1,358 (1.2)	457 (16.4)	1,628 (9.5)	777 (18.5)	992 (14.3)	<b>5,278</b>	<b>3.61%</b>
C60 Hypothyroidism	25 (3.61)	3,445 (3.01)	154 (5.54)	1,069 (6.26)	194 (4.63)	387 (5.59)	<b>5,274</b>	<b>3.61%</b>
C59 Neoplasm after 5 years	17 (2.4)	1,864 (1.6)	184 (6.6)	1,208 (7.1)	441 (10.5)	694 (10.0)	<b>4,408</b>	<b>3.02%</b>
C20 Chronic Heart Failure	55 (7.94)	528 (0.46)	437 (15.71)	1,041 (6.10)	892 (21.28)	976 (14.09)	<b>3,929</b>	<b>2.69%</b>
C57 Neoplasm in follow-up. 1–5 yrs	24 (3.5)	1,721 (1.5)	166 (6.0)	1,028 (6.0)	358 (8.5)	559 (8.1)	<b>3,856</b>	<b>2.64%</b>
C64 Hashimoto's thyroiditis	2 (0.29)	2,057 (1.80)	34 (1.22)	397 (2.33)	31 (0.74)	87 (1.26)	<b>2,608</b>	<b>1.79%</b>
C02 Chronic Renal failure	47 (6.78)	542 (0.47)	253 (9.09)	525 (3.08)	377 (8.99)	354 (5.11)	<b>2,098</b>	<b>1.44%</b>
C25 Cerebrovascular disease	22 (3.17)	255 (0.22)	223 (8.02)	484 (2.84)	383 (9.14)	458 (6.61)	<b>1,825</b>	<b>1.25%</b>
C41 Chronic hepatitis	22 (3.17)	1,086 (0.95)	49 (1.76)	376 (2.20)	48 (1.15)	129 (1.86)	<b>1,710</b>	<b>1.17%</b>
C17 Peripheral Artery Disease	34 (4.91)	372 (0.33)	214 (7.69)	515 (3.02)	260 (6.20)	274 (3.96)	<b>1,669</b>	<b>1.14%</b>
C28 Valvular heart disease	20 (2.89)	609 (0.53)	130 (4.67)	424 (2.48)	224 (5.34)	255 (3.68)	<b>1,662</b>	<b>1.14%</b>
C38 Epilepsy	28 (4.04)	645 (0.56)	71 (2.55)	172 (1.01)	68 (1.62)	86 (1.24)	<b>1,070</b>	<b>0.73%</b>
C46 Inflammatory Bowel Diseases	7 (1.01)	735 (0.64)	40 (1.44)	168 (0.98)	38 (0.91)	44 (0.64)	<b>1,032</b>	<b>0.71%</b>
C62 Basedow and hyperthyroidism	3 (0.43)	698 (0.61)	23 (0.83)	163 (0.95)	25 (0.60)	33 (0.48)	<b>945</b>	<b>0.65%</b>
C23 Venous diseases	16 (2.31)	412 (0.36)	60 (2.16)	187 (1.10)	95 (2.27)	115 (1.66)	<b>885</b>	<b>0.61%</b>
C47 Alzheimer and Dementias	2 (0.29)	39 (0.03)	103 (3.70)	188 (1.10)	217 (5.18)	270 (3.90)	<b>819</b>	<b>0.56%</b>
C40 Rheumatoid arthritis	2 (0.29)	380 (0.33)	38 (1.37)	219 (1.28)	46 (1.10)	76 (1.10)	<b>761</b>	<b>0.52%</b>
C32 Parkinson's disease	7 (1.01)	76 (0.07)	90 (3.24)	185 (1.08)	146 (3.48)	136 (1.96)	<b>640</b>	<b>0.44%</b>
C26 Cirrhosis	19 (2.74)	283 (0.25)	61 (2.19)	149 (0.87)	45 (1.07)	71 (1.03)	<b>628</b>	<b>0.43%</b>
C06 HIV/AIDS	10 (1.44)	459 (0.40)	11 (0.40)	32 (0.19)	1 (0.02)	3 (0.04)	<b>516</b>	<b>0.35%</b>
C01 Transplanted	24 (3.46)	254 (0.22)	29 (1.040)	80 (0.47)	10 (0.24)	8 (0.12)	<b>405</b>	<b>0.28%</b>
C44 Psoriasis and psoriatic arthritis	2 (0.29)	268 (0.23)	11 (0.40)	44 (0.26)	7 (0.17)	7 (0.10)	<b>339</b>	<b>0.23%</b>
C14 Multiple sclerosis	1 (0.14)	282 (0.25)	6 (0.22)	34 (0.20)	2 (0.05)	6 (0.09)	<b>331</b>	<b>0.23%</b>
C05 Blood and Hematopoietic	2 (0.29)	222 (0.19)	17 (0.61)	21 (0.12)	4 (0.10)	9 (0.13)	<b>275</b>	<b>0.19%</b>
C45 Lupus erythematosus	7 (1.01)	106 (0.09)	7 (0.25)	26 (0.15)	4 (0.10)	5 (0.07)	<b>155</b>	<b>0.11%</b>
C51 Sjogren's disease	0 (0)	85 (0.07)	2 (0.07)	45 (0.26)	9 (0.21)	11 (0.16)	<b>152</b>	<b>0.10%</b>
C21 Systemic sclerosis	0 (0)	76 (0.07)	9 (0.32)	28 (0.16)	3 (0.07)	9 (0.13)	<b>125</b>	<b>0.09%</b>
C15 Chronic pancreatitis	3 (0.43)	38 (0.03)	10 (0.36)	31 (0.18)	12 (0.29)	20 (0.29)	<b>114</b>	<b>0.08%</b>
C49 Hyper and hypoparathyroidism	0 (0)	69 (0.06)	2 (0.07)	32 (0.19)	3 (0.07)	8 (0.12)	<b>114</b>	<b>0.08%</b>
C22 Ankylosing spondylitis	0 (0)	70 (0.06)	2 (0.07)	9 (0.05)	2 (0.05)	0 (0)	<b>83</b>	<b>0.06%</b>
C39 Myasthenia gravis	3 (0.43)	27 (0.02)	11 (0.4)	11 (0.06)	11 (0.26)	7 (0.10)	<b>70</b>	<b>0.05%</b>
C36 Addison's disease	1 (0.14)	29 (0.03)	1 (0.04)	11 (0.06)	3 (0.07)	1 (0.01)	<b>46</b>	<b>0.03%</b>
C04 Acromegaly and gigantism	1 (0.14)	10 (0.01)	3 (0.11)	9 (0.05)	3 (0.07)	1 (0.01)	<b>27</b>	<b>0.02%</b>
C24 Cushing syndrome	0 (0)	15 (0.01)	0 (0)	1 (0.01)	0 (0)	0 (0)	<b>16</b>	<b>0.01%</b>

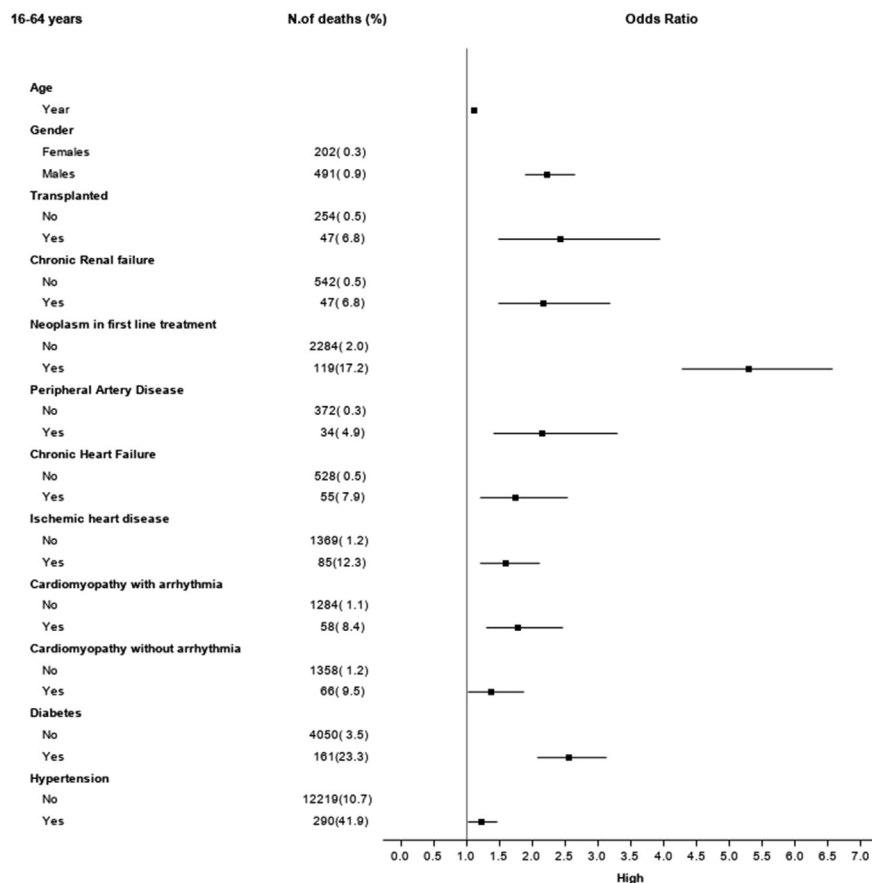


**Fig. 1.** Daily new cases of COVID and cohort mortality trends from February to December 2020. (ATS of Milan, 2021).

**Table 3**

Results of the predictive models in terms of selected comorbidities related to the increased risk of death in the overall cohort and sensitivity analysis restricted to mortality in the first 60 days from diagnosis of SARS-CoV-2 and in the sub-cohorts defined according to the two pandemic waves in Italy. (ATS of Milan, 2021).

Scenarios	16–64 <i>List of comorbidities</i> <i>Dead/alive</i>	65–79 <i>List of comorbidities</i> <i>Dead/alive</i>	80+ <i>List of comorbidities</i> <i>Dead/alive</i>
Overall cohort <i>February 19 -</i> <i>November 30, 2020</i>	<b>693/114,422</b> Transplanted, Chronic Renal failure, Neoplasm in first line treatment, Peripheral Artery Disease, Chronic Heart Failure, Ischemic heart disease, Cardiomyopathy with arrhythmia, Cardiomyopathy without arrhythmia, Diabetes Hypertension	<b>2,782/1,702</b> Chronic Renal failure, Neoplasm in first line treatment, Peripheral Artery Disease, Chronic Heart Failure, Cerebrovascular disease, Ischemic heart disease, Parkinson's disease, COPD, Epilepsy, Alzheimer and Dementias, Diabetes, Psychiatric conditions	<b>4,192/6,926</b> Chronic Renal failure, Chronic Heart Failure, Ischemic heart disease, Parkinson's disease, Psychiatric conditions
Wave I <i>February 19 - May</i> <i>31,2020</i>	<b>451/ 12,133</b> Chronic Renal failure, Neoplasm in first line treatment, Peripheral Artery Disease, Chronic Heart Failure, Ischemic heart disease, Diabetes, Hypertension, Psychiatric conditions	<b>1,624/3,093</b> Chronic Renal failure, Neoplasm in first line treatment, Chronic Heart Failure, Cirrhosis Alzheimer and Dementias, Diabetes, Psychiatric conditions	<b>2,192/1,636</b> Chronic Renal failure, Parkinson's disease, Diabetes, Psychiatric conditions
Wave II <i>October 1 -</i> <i>November 30, 2020</i>	<b>226/96,929</b> Transplanted, Chronic Renal failure, Chronic Heart Failure, Cerebrovascular disease, Ischemic heart disease, Epilepsy, Hypercholesterolemia, Diabetes, Hypertension	<b>1,099/13,428</b> Chronic Renal failure, Peripheral Artery Disease, Chronic Heart Failure, Cerebrovascular disease, Ischemic heart disease, Parkinson's disease, COPD, Diabetes	<b>1,895/5,046</b> Chronic Renal failure, Chronic Heart Failure, Venous diseases, Cerebrovascular disease, Cardiomyopathy with arrhythmia, Parkinson's disease, COPD, Psychiatric conditions
Deaths within 60 days <i>February 19 -</i> <i>November 30, 2020</i>	<b>628/114,487</b> Transplanted, Chronic Renal failure, Neoplasm in first line treatment, Peripheral Artery Disease, Chronic Heart Failure, Cerebrovascular disease, Ischemic heart disease, Diabetes, Hypertension	<b>2,533/17,321</b> Chronic Renal failure, Peripheral Artery Disease, Chronic Heart Failure, Cerebrovascular disease, Ischemic heart disease, Alzheimer and Dementias, Diabetes Psychiatric conditions	<b>3,807/7,311</b> Chronic Renal failure, Peripheral Artery Disease, Chronic Heart Failure, Ischemic heart disease, Parkinson's disease, Diabetes, Psychiatric conditions



**Fig. 2a.** Estimated Odds Ratios\* and corresponding 95% confidence intervals from a multivariable logistic model with comorbidities selected by the LASSO procedure in COVID-19 cases with 16–64 year of age (ATS of Milan, 2021).



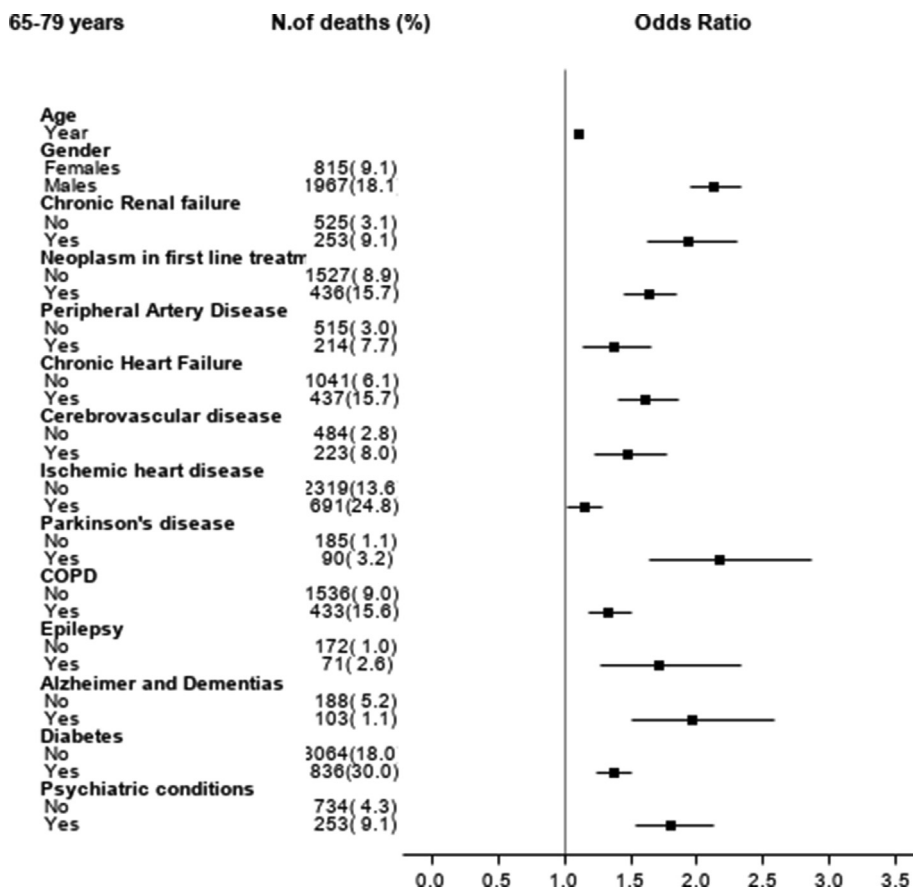


Fig. 2b. Estimated Odds Ratios\* and corresponding 95% confidence intervals from a multivariable logistic model with comorbidities selected by the LASSO procedure in COVID-19 cases with 64–79 year of age (ATS of Milan, 2021).

### 3.4. Stratification models

In order to simplify the identification of priority subgroups for vaccination, two risk categories were defined within each age class: high-risk, with highest priorities, for subjects with at least one of the comorbidities selected in the LASSO regression and a low-risk for the remaining subjects. The stratification models related to this approach are reported in Table 4 for each age class (Model I). The odds of death for patients at high-risk is approximately 3 times and 2 times higher in subjects aged 16–64 and 65–79 respectively, and it increased by 60% in patients 80+ compared to patients without the selected comorbidities. A projection of the number of subjects in the two strata defining priorities in vaccination is shown in Table 5, based on the population of the area of the ATS of Milan as of January 1st 2020. A sensitivity analysis shows that, also within comorbidities selected as those significantly associated with mortality, the number of concomitant conditions in the same subject is per-se a prognostic indicator (Model II, Table 4). These stratification models were also fitted considering 60 days mortality and separately in the two waves of the epidemics, with consistent results in terms of estimated odds ratios (supplementary Table 1).

We also considered the definition, within each age class, of a medium-risk group as a possible second subgroup in the vaccination plan after the high-risk group. This includes subjects with any of the other non selected comorbidities, leaving in the new low-risk group subjects with none of the entire set of 42 chronic diseases. However, the odds of death for patients at medium-risk was not so markedly different compared to that in patients without any chronic condition, except for the younger age class (Model III, Table 4).

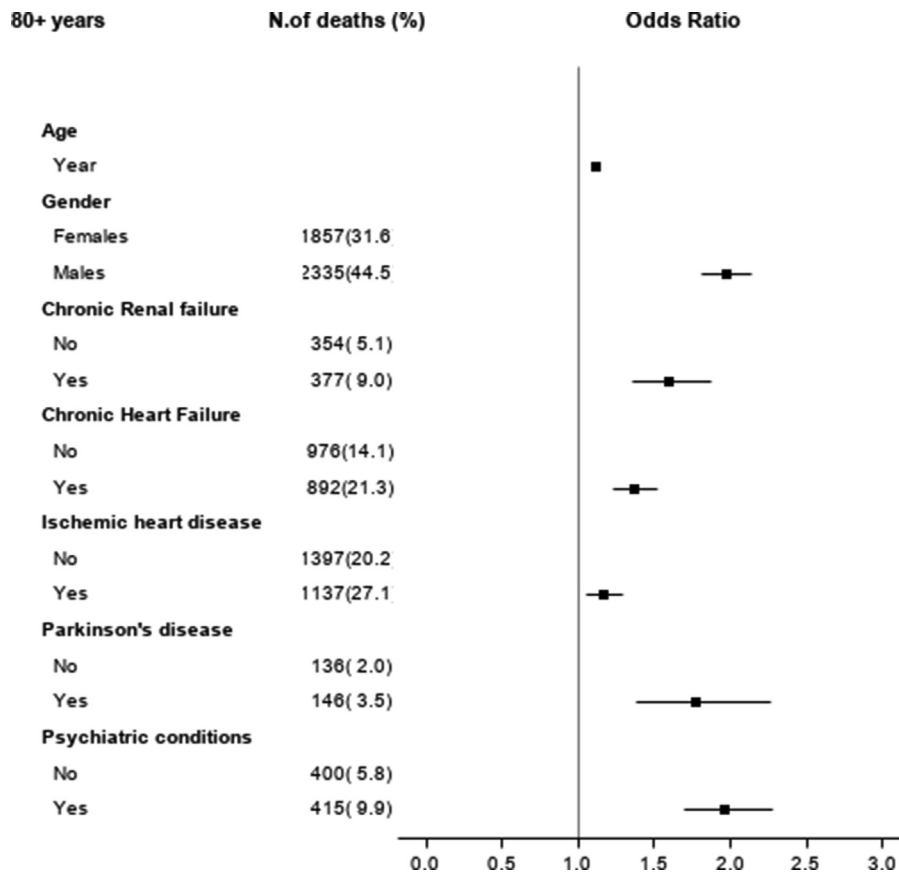
### 4. Discussion

We developed a model to identify which chronic conditions independently affect the individual risk of mortality, on top of age, in subjects with COVID-19 infection. The model used real world data derived by combining multiple electronic health records of a large population in the region of Lombardy (ATS-Milano, about 3.5 million subjects), in the area where the first case of COVID-19 was diagnosed in Italy.

The model was developed with the purpose of optimizing the vaccination programme, by defining which groups of frail individuals should be allocated to vaccination with priority. The idea is to identify those individuals who, if infected, would have a severe course of the disease, increasing the burden on health services, and would experience a fatal event. Similar models might also be useful for developing other preventive strategies, including preventive therapeutic approaches and specific intervention policies at the social level.

Since age per se is undoubtedly a strong factor associated with morbidity and mortality, as confirmed also in our COVID-19 population, the phases of the vaccination campaign will follow priorities by age classes (from older to younger). Our approach separately defined, within each age class of interest, a risk level based classification.

In the two older categories (65–79 years of age and 80+) at COVID-19 diagnosis, existing chronic conditions that were associated with increased risk of death were chronic renal failure, chronic heart failure, cerebrovascular disease, Parkinson disease and psychiatric conditions. In older patients (80+) also the existence of venous diseases was an independent factor for an



**Fig. 2c.** Estimated Odds Ratios\* and corresponding 95% confidence intervals from a multivariable logistic model with comorbidities selected by the LASSO procedure in COVID-19 cases with 80+ year of age (ATS of Milan, 2021).

increased risk of death. In the lower age category (65–79), additional independent factors that increased the risk of death were: having experienced an organ transplantation, presence of blood and hematopoietic disorders, epilepsy, Alzheimer or dementias, having a cancer in first line therapy. All these factors, independently, carried a 1.5 or two-fold increase in the

odds of death, approximately (Fig. 2). In the younger age category (16–64 years of age), neoplasm in first line therapy, epilepsy, psychiatric conditions, organ transplantation, chronic renal failure, cardiomyopathy with arrhythmia, Parkinson, diabetes were independently associated with an increased risk of death.

**Table 4**

Results of simplified logistic regression models based on selected covariates. Model I classifies subjects in two strata: high risk (at least one selected comorbidity) and low risk (none of the selected comorbidities). Model II presents a sensitivity analysis by number of selected comorbidities. Model III classifies subjects in three strata: high risk (at least one of the selected comorbidity), medium risk (other non-selected chronic comorbidities) and low risk (none of the selected comorbidities). (All models are adjusted by age in years and gender). (ATS of Milan, 2021).

	16–64				65–79				80+			
	N. deaths	N. cases	%	OR (95% CI)	N. deaths	N. cases	%	OR (95% CI)	N. deaths	N. cases	%	OR (95% CI)
<b>Model I</b>												
<i>At least one selected comorbidity</i>												
No	267	97,169	39%	1#	893	10,435	32%	1#	2,145	6,650	51%	1#
Yes	426	17,946	61%	3.39 (2.87–3.99)	1889	9,419	68%	2.19 (2.01–2.39)	2,047	4,468	49%	1.63 (1.51–1.77)
<b>Model II</b>												
<i>N. of selected comorbidities</i>												
0	267	97,169	39%	1#	893	10,435	32%	1#	2,145	6,650	51%	1#
1	171	13,100	25%	2.10 (1.72–2.56)	770	5,306	28%	1.57 (1.42–1.74)	1,302	3,003	31%	1.48 (1.35–1.62)
2	122	3,322	18%	4.52 (3.60–5.69)	518	2,348	19%	2.42 (2.14–2.74)	587	1,186	14%	1.88 (1.65–2.14)
3+	133	1,524	19%	10.38 (8.25–13.05)	601	1,765	22%	4.05 (3.58–4.59)	158	279	4%	2.57 (2.01–3.30)
<b>Model III</b>												
Low risk	198	82,282	29%	1#	361	4,309	13%	1#	393	1,255	9%	1#
Medium risk	69	14,887	10%	1.55 (1.18–2.05)	532	6,126	19%	0.93 (0.81–1.07)	1,752	5,395	32%	1.08 (0.95–1.25)
High risk	426	17,946	61%	3.75 (3.13–4.49)	1,889	9,419	68%	2.10 (1.86–2.37)	2,047	4,468	49%	1.75 (1.52–2.00)
<b>Total</b>	<b>693</b>	<b>115,115</b>			<b>2,782</b>	<b>19,854</b>			<b>4,192</b>	<b>11,118</b>		

**Table 5**

Projection of the number of subjects involved in the vaccination programme according to the two strata priority approach: high risk (at least one selected comorbidity) and low risk (none of the selected comorbidities). (ATS of Milan, 2021).

	N.	%	Cumulative frequency	Cumulative %
High risk 80+	85,385	3.02	85,385	3.0
Low risk 80+	202,998	7.18	288,383	10.2
High risk 65–79	182,414	6.45	470,797	16.6
Low risk 65–79	333,748	11.80	804,545	28.4
High risk 16–64	263,659	9.32	1,068,204	37.8
Low risk 16–64	1,760,805	62.24	2,829,009	100.0

The finding that, in subjects 65 years old or more infected with SARS-CoV-2, a pre-existing psychiatric illness is related to more than 1.5–2 fold risk of death is an important effect that should be the object of in depth analyses and targeted intervention. Whether this risk may be due to SARS-CoV-2 that exacerbates the pre-existing illness or to disorganized behavior, poor insight, and marginalized social status that together may cause a delay in diagnosis or treatment, has yet to be determined [14]. Cardiovascular disease has also been highly reported among COVID-19 patients and is associated with an increased mortality and cardiovascular complications such as thromboembolic events, myocarditis, acute coronary syndrome, arrhythmia, cardiogenic shock and heart failure [15].

Several systematic reviews suggested that diabetes is a determinant of severity and mortality of COVID-19 patients [16], but the pathogenesis of increased mortality is still unclear and possibly related to an uncontrolled inflammatory response which determines a high hypercoagulable state [6].

The presence of neoplasm in first line treatment represents a repeated risk factor in any age class and it is unclear whether cancer patients are at high risk of death due to an immunocompromised state linked to a recent chemotherapy or surgery [17].

The proposed system of priority is based on a simplified classification model that, adjusting for age and gender, identifies as the group with highest priority to vaccination the high risk subgroup that presents, within the age class, at least one of the comorbidities significantly affecting the odds of dying after COVID-19. This approach was chosen for its simplicity in targeting the frailest subjects whose timely vaccination would likely minimize future deaths by preventing infection. Indeed, these two strata (high and low risk) can be easily identified in the population of the reference area, due to the availability of updated administrative data. The possible development of a risk score based on the model regression coefficients of each selected comorbidity would complicate the approach and would also be very dependent on the fitted “local” model estimates. An “in between” simple strategy could be that of identifying a gradient of risk by considering a classification with 1, 2 or 3+ comorbidities among those significantly impacting mortality for defining increasing levels of priority. Another strategy that we considered is to define a medium risk group, second in priority after the high risk, based on the presence of chronic conditions other than those identified having a high impact on mortality. Yet this stratum seems to have some relevance only in the younger age category, where mortality is however much less marked. It should be considered only if relatively easy to be implemented and because an additional stratum may facilitate the vaccination programme in this largest age group.

The simple two-level stratification can be seen as a limitation of our work. However, this strategy allows health policy makers to quickly calculate the proportion of subjects that will undergo different phases of vaccination, as presented in the projection of Table 5, and to properly schedule the use of available vaccines.

This approach to the definition of priorities can easily be shared with other Agencies of the Health System in Italy that collect similar administrative data and can easily be exported with the same methodological approach in other countries and eventually adapted to local results in terms of number and types of comorbidities. The use of administrative data, now routinely collected during the pandemic on COVID-19 patients, guarantees that results can be rapidly updated while the pandemic is evolving, if priorities need to be adapted.

The cohort study approach is appropriate for the aim of our work while other designs, such as a case-control-study, could be considered if we aimed at collecting specific additional information and would anyway present difficulties in defining an appropriate control group at the population level, due to the presence of asymptomatic subjects.

One interesting aspect that we did not address in full, but that is worth considering in future works, is the possible actions to be taken, in the vaccination programme, in case of an incident diagnosis that induces a future chronic condition among those identified as at risk. For example, the diagnosis of heart failure during the pandemic should induce the health system to consider a priority for vaccine since chronic heart failure is a condition at risk for 80 +. Also, the first diagnosis of cancer during the pandemic should per se induce a priority.

Generally speaking, the consistency of various significant pathologies associated with mortality after COVID-19, in the different age groups, in the two pandemic waves, must also be carefully considered because priorities of vaccination could be also defined by single comorbidities rather than by age, if vaccination becomes easier and faster to manage.

Last but not least, the focus on comorbidities for defining priorities in the vaccination should not overcome the clinical decision on whether the single patient does not have specific problems that contraindicate the administration of the vaccine.

A limitation in our approach is that it relies on a set of administrative data and has two aspects. On one side, we are not necessarily exhaustive in considering all chronic diseases and we do not have detailed clinical parameters usually available in clinical registries, data on lifestyle and social behaviour. On the other side, in settings where administrative data are less reliable and not timely updated, this approach cannot be feasible. Programme based on age alone would per se, in these settings, achieve vaccine uptake and capture, with older subjects, those where prevalence of risk features and chronic conditions is naturally higher. Another limitation is that we considered factors affecting mortality and not, more generally, the severe course of the infection as, for instance, using data on access to intensive care units. This could be indeed a feasible extension to a less “hard” but relevant endpoint. However, it could also be much more influenced by the local situation in terms of availability of sub-intensive and intensive care units.

Finally, after the early phase of vaccination, data on effectiveness and safety and on the epidemiological situation will give the



opportunity to revise prioritization and discuss the future approach in the remaining population.

## 5. Transparency declaration

Antonio Giampiero Russo declares that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

## 6. Details of ethical approval

Ethics approval and consent to participate were not required, as this is an observational study based on data routinely collected by the Agency for Health Protection (ATS) of Milan, a public body of the Regional Health Service of Lombardy. Among the institutional functions of the ATS, established by the Lombardy regional legislation (R.L. 23/2015), is the management of individual care in the regional social and healthcare system, the evaluation of services provided to patients residing in the area covered by the ATS, and their health outcomes. This study is also ethically compliant with Italian law (Legislative Decree 101/2018) and the “General Authorisation to Process Personal Data for Scientific Research Purposes” (clauses no. 8 and 9/2016, referred to in the Data Protection Authority action of 13/12/2018). Data were anonymized with a unique identifier in the different datasets before being used for the analyses.

## 7. Details of contributors, and the name of the guarantor

Concept and design: AGR, AD and MG. Acquisition, interpretation of data: AGR, AD and MG.

Drafting of the manuscript: AGR, AD and MG. Critical revision of the manuscript for important intellectual content: AGR, AD and MG. Statistical analysis: AGR, AD and MG. Administrative, technical, or material support: All authors. AGR had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. AGR, the guarantor, accepts full responsibility for the work, had access to the data, and controlled the decision to publish. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

## 8. Data sharing statement

The dataset from this study is held securely at the ATS of Milan, Epidemiology Unit. Data sharing agreements prohibit the ATS of Milan from making the dataset publicly available. The full dataset creation plan and underlying analytic code are available from the authors upon request.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We would like to thank the entire personnel involved in the Departments, Services, and Units of the Agency for Health Protection (ATS) of Milan. In particular, the Public Health Unit; Infectious Disease Surveillance Unit; Laboratory of Prevention; Occupational Health and Prevention Unit and Epidemiology Unit.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.vaccine.2021.03.076>.

## References

- [1] Tunesi S, Murtas R, Riussi A, et al. Describing the epidemic trends of COVID-19 in the area covered by Agency for Health Protection of the Metropolitan Area of Milan. *Epidemiol Prev* 2020;44 (5–6)(Suppl 2):95–103. <https://doi.org/10.19191/EP20.5-6.S2.10>.
- [2] Russo AG, Faccini M, Bergamaschi W, Riussi A. A strategy to reduce adverse health outcomes in subjects highly vulnerable to covid-19: results from a population-based study in northern Italy. *BMJ open*. 2021;11(3):. <https://doi.org/10.1136/bmjopen-2020-046044>e046044.
- [3] Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.
- [4] Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. *PLoS One* 2020;15(11). <https://doi.org/10.1371/journal.pone.0241955>. eCollection 2020.
- [5] Mantovani A, Byrne CD, Zheng MH, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: A meta-analysis of observational studies. *Nutr Metab Cardiovasc Dis* 2020;30:1236–48.
- [6] Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with covid-19: Evidence from meta-analysis. *Aging* 2020;12(7).
- [7] Vahidy FS, Pan AP, Ahnstedt H, et al. Sex differences in susceptibility, severity, and outcomes of coronavirus disease 2019: Cross-sectional analysis from a diverse US metropolitan area. *PLoS One* 2021;16(1). <https://doi.org/10.1371/journal.pone.0245556>. eCollection 2021.
- [8] Moore S, Hill EM, Dyson L, Tildesley M, Keeling, MJ. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. doi: <https://doi.org/10.1101/2020.09.22.20194183>.
- [9] WHO-INT -Guidance on developing a national deployment and vaccination plan for COVID-19. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [10] NHS. <https://www.gov.uk/government/publications/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19>.
- [11] Ministero della Salute: Salute Vaccinazione anti-SARS-CoV-2/COVID-19 - Piano strategico. Elementi di preparazione e di implementazione della strategia vaccinale. Update December 15, 2020. [http://www.salute.gov.it/portale/documentazione/p6\\_2\\_2\\_1.jsp?lingua=italiano&id=2986](http://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id=2986).
- [12] Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J Roy Statist Soc Ser B (Methodological)* 1996;58(1):267–88 [accessed January 18, 2021].
- [13] SAS/STAT User's Guide: SAS® 9.4 and SAS® Viya® 3.4 Programming Documentation. PROC Procedure GLMSELECT.
- [14] Li L, Li F, Fortunati F, Krystal JH. Association of a Prior Psychiatric Diagnosis with Mortality among Hospitalized Patients with Coronavirus Disease 2019 (COVID-19) Infection. *JAMANetw Open* 2020;3(9):. <https://doi.org/10.1001/jamanetworkopen.2020.23282>. PMID: 32997123; PMCID: PMC7527869e2023282.
- [15] Guzik TJ, Mohiddin SA, Dimarco A, Patel V, Savvatis K, Marelli-Berg FM, et al. COVID-19 and the cardiovascular system: implications for risk assessment, diagnosis, and treatment options. *Cardiovasc Res* 2020;116(10):1666–87. <https://doi.org/10.1093/cvr/cvaa106>. PMID: 32352535; PMCID: PMC7197627.
- [16] Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diabetes Metab Res Rev* 2020; e3319. doi: 10.1002/dmrr.3319. Epub ahead of print. PMID: 32233013; PMCID: PMC7228407.
- [17] Gosain R, Abdou Y, Singh A, Rana N, Puzanov I, Ernstoff MS. COVID-19 and Cancer: a Comprehensive Review. *Curr Oncol Rep* 2020;22(5):53. <https://doi.org/10.1007/s11912-020-00934-7>. PMID: 32385672; PMCID: PMC7206576.