Identify deaths due to or involving COVID-19 in absence of the death certificate

Identificare i decessi riconducibili al COVID-19 in assenza del certificato di morte

Rossella Murtas, Antonio Giampiero Russo
UOC Epidemiology Unit, Agency for Health Protection of the Metropolitan Area of Milan (Italy)
Corresponding author: Antonio Giampiero Russo; agrusso@ats-milano.it

ABSTRACT

INTRODUCTION: during 2020, Italy was one of the first nation hit by SARS-CoV-2, but it was not the hardest-hit country in terms of deaths. In absence of the death certificate, the burden of COVID-19 on mortality is usually calculated from overall deaths or from deaths of patients tested positive for COVID-19. However, these measures do not express the real burden of the disease on the population.
OBJECTIVES: identify deaths due to or involving COVID-19 in absence of the death certificates.
DESIGN: deaths for all causes, cause-specific deaths, COVID-19 hospitalization and COVID-19 confirmed cases between 01.01.2020 and 31.12.2021 observed in subjects residing in the territory of the ATS of Milan. Potential deaths due to or involving COVID-19 as those occurring in an optimal time period between the date of death and the date of positive swab and/or COVID-19 hospitalization, were identified. Optimal time period was defined maximizing sensitivity and specificity, comparing potential COVID-19 deaths with 2020 cause-specific mortality as gold standard, stratifying results by time of deaths, age, and number of comorbidities. Then, this method was further validated using a time-series approach to estimate the excess mortality during the COVID-19 outbreak in comparison with the pre-outbreak period 2015-2019. Accuracy of predictions was evaluated with the Root Mean Square Error (RMSE) between observed and predicted values.
SETTING AND PARTICIPANTS: 78,202 deaths for all causes, of which 8,815 due to or involving COVID-19 as classified by the Milan Register of Death Causes for 2020.
MAIN OUTCOME MEASURES: all-cause mortality, cause-specific mortality.
RESULTS: from the beginning of the epidemic, 30% (23,495) died in the first semester of 2020, 26% (19,988) in the second semester of 2020, 23% (18,189) in the first semester of 2021, and 21% (16,530) in the second semester of 2021. COVID-19 hospitalizations were 13.826 (17%), while confirmed COVID-19 cases were 17,548 (22%). The optimal time intervals capable to identify a potential death due to or involving COVID-19 were 0-61 between the date of death and the date of positive swab and 0-11 between the date of death and the date of COVID-19 hospitalization, with an overall sensitivity of 90%, a specificity of 95%, and a RMSE of 3.6. Comparing the method proposed with the time-series approach, a RMSE in 2021 of 15.8 was found. Results showed different optimal time intervals for 2021 vs 2020 and by years of age and comorbidities.
CONCLUSIONS: this study found that deaths due to or involving COVID-19 could be sensitively identified from the date of positive swab and/or COVID-19 hospitalization. This method can be used for public health interventions which provided so far measures in terms of total deaths instead of real numbers of COVID-19 death, in particular those involving the effective reproduction number usually calculated from overall mortality.

Keywords: COVID-19, excess mortality, sensitivity, specificity

WHAT IS ALREADY KNOWN

- **In 2020, Italy reported 77,165 deaths of patients tested positive for COVID-19 over 746,146 deaths for all-cause, with an overall excess of 15.6% compared to 2015-2019.**
- **However, these numbers are underestimated compared to reality, especially in the first wave of the epidemic characterized by limited testing capacity where the number of confirmed cases were seriously underestimated.**
- **A report from the Italian National Health Institute and the Italian National Institute of Statistics reported that median time between death and a positive swab is of 12 days and that 89% of deaths of patients tested positive for COVID-19 happened within 30 days from a positive swab.**

WHAT THIS STUDY ADDS

- **This study proposes a method capable to identify whether a death is attributable to COVID-19 before the death certificate is available. Usual methods do not permit to evaluate deaths individually, but rather the excess attributable to COVID-19 calculating the difference between overall mortality and expected deaths as arising from the same process of the pre-outbreak period.**
- **This study found that defining a potential COVID-19 death as that occurring at most 61 days from a positive swab and/or 11 days from a COVID-19 hospitalization permits to correctly classify 90% COVID-19 deaths. Including COVID-19 hospitalization in the algorithm, the sensitivity was increased by 4%.**
- **In the first semester of 2020, there was a peak attributable to chronic ischaemic heart disease, malignant neoplasm of bronchus and lung, pneumonia, complications and ill-defined descriptions of heart disease, other chronic obstructive pulmonary disease, unspecified dementia and Alzheimer disease that were actually related to COVID-19 as from the time series approach.**

RIASSUNTO

INTRODUZIONE: nel corso del 2020, l’Italia è stata una delle prime nazioni colpite dal virus SARS-CoV-2, ma non è stato il Paese più colpito in termini di decessi. In assenza del certificato di morte, il carico della pandemia sulla mortalità è solitamente calcolato a partire dalla mortalità generale o dai decessi avvenuti in pazienti diagnosticati con COVID-19. Tut-

tavia, queste misure non esprimono il carico reale della malattia sulla popolazione.
OBIETTIVI: identificare in anticipo i decessi riconducibili a COVID-19 in assenza del certificato di morte.
DISEGNO: sono stati raccolti i decessi per tutte le cause e specifici per causa, i ricoveri per COVID-19 e i casi positivi alla malattia tra il 01.01.2020 e il 31.12.2021 osservati in soggetti residenti nel territorio dell'ATS di Milano. Sono stati identificati i decessi potenziali riconducibili al COVID-19 come quelli che si sono verificati in una finestra temporale tra la data del decesso e la data del tampone positivo e/o del ricovero per COVID-19. La finestra temporale ottimale è stata definita massimizzando la sensibilità e la specificità, confrontando i decessi potenziali con la mortalità per causa del 2020 come gold standard, i risultati sono stati stratificati per periodo di decesso, età e numero di comorbidità. Questo metodo è stato ulteriormente convalidato utilizzando un approccio di serie temporali per stimare la mortalità in eccesso durante l'epidemia di COVID-19 rispetto al periodo pre-epidemico 2015-2019. L'accuratezza delle previsioni è stata valutata con il Root Mean Square Error (RMSE) tra i valori osservati e quelli previsti.
SETTING E PARTECIPANTI: 78.202 decessi per tutte le cause, di cui 8.815 riconducibili al COVID-19 classificati dal Registro delle cause di morte di Milano per l'anno 2020.
PRINCIPALI MISURE DI OUTCOME: mortalità per tutte le cause, mortalità specifica per causa.
RISULTATI: dall'inizio dell'epidemia, il 30% (23.495) delle persone incluse nello studio è morto nel primo semestre del 2020, il 26% (19.988) nel secondo semestre del 2020, il 23% (18.189) nel primo semestre del 2021 e il 21% (16.530) nel secondo semestre del 2021. I ricoveri per COVID-19 sono stati 13.826 (17%), mentre i casi confermati di COVID-19 sono stati 17.548 (22%). Le finestre temporali ottimali in grado di identificare i decessi potenziali riconducibili al COVID-19 erano 0-61 tra la data del decesso e la data del tampone positivo e 0-11 tra la data del decesso e la data di ricovero per COVID-19, con una sensibilità complessiva del 90%, una specificità del 95% e un RMSE di 3,6. Confrontando il metodo proposto con l'approccio delle serie temporali, è stato trovato un RMSE nel 2021 di 15,8. I risultati hanno mostrato finestre temporali differenti nel 2021 rispetto al 2020 per età e numero di comorbidità.
CONCLUSIONI: in questo lavoro, è stato riscontrato che i decessi riconducibili al COVID-19 potrebbero essere sensibilmente identificati a partire dalla data di positività del tampone e/o di ricovero per COVID-19. Questo metodo può essere utilizzato per interventi in sanità pubblica che finora hanno fornito misure in termini di decessi totali, anziché di decessi realmente riconducibili al COVID-19, in particolare quelli che coinvolgono l'indice di contagio Rt solitamente calcolato dalla mortalità complessiva.

Parole chiave: COVID-19, eccesso di mortalità, sensibilità, specificità.

INTRODUCTION

During 2020, Italy was one of the first nation hit by the SARS-CoV-2 virus.[1] However, it was not the hardest-hit country in terms of deaths: Italy reported 74,159 deaths of patients tested positive for COVID-19[2] with a maximum z-score in the first wave of 15.81, followed by Netherlands, Belgium, France, UK, and finally Spain (the most hit European country in terms of deaths).[3] However, these numbers are underestimated compared to reality, as they are calculated from deaths of patients tested positive for COVID-19,[4] especially in the first wave of the epidemic characterized by limited testing capacity where the number of positive subjects (233,019 according to official statistics)[2] were seriously underestimated.[5] Modi and colleagues[6] found that the number of COVID-19 deaths in Italy in 2020 until 9th September varied between 59,000 and 62,000, much bigger than official numbers of 36,000 deaths.[2]

For this reason, overall excess mortality has been chosen by the World Health Organization as a proxy of COVID-19-related mortality, excess cases were calculated comparing the total number of deaths (746,146 in 2020 compared to 645,620 in 2019 in Italy according to the Italian National Health Institute – ISS) during COVID-19 pandemic with previous years.[7-10] All-cause mortality excesses had a natural spatial and temporal heterogeneity with a peak for Italian Northern regions in 2020 (overall excess 15,6% vs 24,6% in Northern regions and 7.7% in Southern regions) and a more marked trend in the Southern regions in 2021 (overall excess 9.8% vs 8.2% in Northern regions and 13% in Southern regions).[8] In the provinces of Milan and Lodi – area covered by the Agency for Health Protection (ATS) of Milan, Lombardy region – from January to April 2020, researchers observed a total number of death of 17,959 with an excess of 49% compared to previous years.[9]

The 'natural' solution would be to timely classify the underlying cause of death from the local/national registries, but this process has an inevitable delayed time that depends on various reason: registration delay, staff availability, timing between processing the information, and forwarding to the agency responsible for compiling data. The median time between a death occurring and being registered (registration delay) in England and Wales increased from four days in 2019 to five days in 2020; this could be explained by an increase in the number of registered deaths (530,841 in 2019 and 607,922 in 2020).[11]

To date, the law imposes the notification of the death causes within 2 years from death,[12] but, in order to quickly provide epidemiological and health-related information, a shorter time period would be desired. One of the six aims of the Italian Recovery and Resilience

Plan[13] is the digitalization of the Italian National Health System and; in particular, the digitalization of the death causes directly by certifying doctors is at the attention of the privacy guarantor.[14] However, we are very far from the aimed digitalization.

United Kingdom provides, from the beginning of the epidemic, the daily numbers of deaths of people whose death certificate mentioned COVID-19 as one of the causes and deaths within 28 days of positive test which amount to 186,094 and 164,454, respectively (the second is 88% of the first).[15] Both these numbers might be underestimated if:

1. deaths that should have been attributed to COVID-19 were not tested and identified;[16-18]

2. limited testing capacity. As previously asserted, likely situations happened in the first wave of COVID-19 (where the Italian Institute of Statistics – Istat – and the Italian Ministry of Health estimated, by a seroprevalence survey, that the ratio between notified and real cases was at least of 1 to 6), but also in periods of scarce testing inclination of the population.

Lacking timely classification of the underlying cause of deaths, there was the need to choose methods which allow to timely identify deaths due to or involving COVID-19 in absence of the death certificates. This paper proposes a method to select the best time interval between the date of deaths and the date of positive swab and/or COVID-19 hospitalization, validating this method by time-series analysis. This would be fundamental for public health interventions which provided so far public health measures in terms of total deaths instead of real numbers of deaths due to or involving COVID-19.

METHODS

SETTINGS AND PARTICIPANTS

Data collected included deaths for all causes between 01.01.2020 and 31.12.2021 observed in subjects residing in the territory of the ATS of Milan, which covers 193 municipalities in the Lombardy Region (Northern Italy), with a total population of 3.48 million inhabitants. Deaths were retrieved form the database of the office of vital statistics of the residents in the city of Milan and from the New Regional Register (*Nuova Anagrafe Regionale*, NAR) which registers all residents and healthcare users of the 193 municipalities of the ATS of Milan.[19]

The underlying cause of death was collected from the local Register of Causes of Death (ReNCaM) of the ATS of Milan; deaths due to or involving COVID-19 were classified according to the ICD-10 codes U071 and U072.[4] To be more conservative as possible, both the principal and the secondary causes of death were considered. The term "due to COVID-19" was used when referring only to deaths where COVID-19 were recorded as the underlying cause of death, "involving COVID-19" was used when referring to dead people that had COVID-19 mentioned anywhere on the death certificate, whether as an underlying cause or not.[20] ReNCaM of the ATS of Milan completely classified, in ICD-10 codes, all deaths happened in 2020 and started to classify deaths happened in 2021.

Hospitalizations due to COVID-19 were identified form the administrative data flows, updated at December 2021, as from ICD-9 codes 043, 0431, 0432, 0433, 4804, 5189, 5197, 04311, 04312, 04321, 04322, 04331, 04332, 48041, 48042, 51891, 51892, 51971, 51972, V0185, V0700, V0708, V1204, V7184.[21] Additional information on hospitalizations of patients with COVID-19 were derived from a specific daily data flow instituted by the Lombardy Region from the beginning of the epidemic.

Confirmed cases and date of first positive swab were collected from a web-based platform, specifically developed since the beginning of the outbreak, to trace positive and negative cases as well as related contacts. A confirmed-case is defined as a person with a real-time polymerase chain reaction (RT-PCR) positive result, irrespective of clinical signs and symptoms.[22] In this study, only the first positivization was considered, without considering any reinfections.

There were few notification errors in the administrative databases resulting in positive swabs or hospitalizations posterior to death: patients characterized by this problem have been considered as negative subjects and/or subjects with no hospitalizations.

IDENTIFICATION OF POTENTIAL DEATHS DUE TO OR INVOLVING COVID-19

This study proposes a method to identify potential deaths due to or involving COVID-19, in absence of the death certificate. Validation was carried out comparing potential COVID-19 deaths with those due to or involving COVID-19 as from ReNCaM (hereafter called 'COVID-19 related deaths'). In order to identify potential COVID-19 deaths:

1. the time interval between the date of positive swab and date of death and between the date of hospitalization due to COVID-19 and of death were calculated;

2. a potential COVID-19 death was defined as one falling in a specific time period, i.e., k-days before death, separately for positive swab and hospitalization date;

3. k was made varying between all possible values;

4. the optimal time period which maximize the Youden index[23] by comparing potential COVID-19 deaths with COVID-19 related deaths was calculated.

It was chosen to directly maximize the Youden index as a composite measure of both sensitivity and specificity. Given that there were numerous optimal time periods which maximized this index, the choice was to select the smallest time period capable to maximize the Youden index.

Optimal time windows were calculated also stratifying the cohort by:

- time of deaths: 2020-2021, 2020 alone, 2021 alone, first semester of 2020, and second semester of 2020;
- age: 0-49, 50-64, 65-79, 80+ years of age;
- number of comorbidities: 0, 1, 2+.

Accuracy of predictions was evaluated with the Root Mean Square Error (RMSE) between observed and predicted values which express estimation errors in the same unit of the outcome. The RMSE was chosen in place of the usual Mean Absolute Percentage Error for Poisson data, because small number of deaths will significantly impact MAPE leading to skewed distributions.[24]

SENSITIVITY ANALYSIS

To evaluate the method proposed, a time-series approach was used to estimate the excess mortality during the COVID-19 outbreak in comparison with the pre-outbreak period 2015-2019. The COVID-19 epidemic has driven a global reduction in influenza incidence, both in 2020 and 2021, mostly because viral circulation were limited by behavioural changes (social distancing, mask wearing, and hygiene measures) and travel and movement restrictions.[25] For this reason, deaths due to or involving Influenza and Pneumonia (ICD-10 codes J10-J17) were removed from overall mortality between 2015-2019.[20]
A quasi-Poisson regression model was used (dispersion parameter in the data was 1.33):[26,27]

$$\ln [E(Z_t)] = \alpha + dow + h_1 \text{(day of the year, 2)} + h_2 (T_t, 2)$$

where:
Z_t is the number of deaths (all-cause mortality not due to or not involving Influenza and Pneumonia) on day t
dow is a dummy indicator for the day of the week (to control for weekly variation in mortality).
To deal with time-varying confounders, it was included a natural cubic spline with 2 df for day of the year (h_1) and a natural cubic spline with 2 df for temperature (h_2). Daily mean temperature was collected from 1 monitoring stations located in the centre of the city of Milan. To avoid misalignment in the time series, 29th of February of leap years was removed. Df have been chosen with k-fold cross validation[28] minimizing the RMSE in test sets. A linear term for date to control for long-term trends has been evaluated, but it was not included in the model, because it resulted in worst diagnostics than the selected model. Diagnostics of the final model included plots of model residuals, observed vs fitted values, autocorrelation (ACF) and partial autocorrelation (PACF) function of the residuals to determine adequate adjustment for seasonal trends, and k-fold cross validation to evaluate overfitting.
ACF and PACF of the selected model showed moderate correlation among residuals up to 6 days of lag (see Figure S1, online Supplementary Materials), which highlighted the need for an autoregressive component in the model[26,27,29] (autocorrelation between residuals significantly improved after including an autoregressive component) (Figure S2). However, this would not be mathematically feasible with the aim of this study to estimate 2020-2021 excess deaths as of 2015-2019 and, for this reason, no autoregressive terms in the model were included. Firstly, the parameters were estimated in the 2015-2019 set, then the outcome in 2020-2021 was predicted, and the number of excess deaths calculated as the difference between observed and predicted values.
Finally, the effective reproduction number (R_t at 14 days) was calculated from daily overall mortality and from potential COVID-19 deaths.[30] These public health measures, usually calculated on confirmed cases, are considered fundamental to calculate the effective reproduction number as independent from testing strategies.[31]
Dataset extraction, calculation of sensitivity, specificity, and Youden Index were performed with SAS software by programming ad hoc macro functions. RMSE, time series analysis, and the quasi-Poisson regression model (using the function glm) were performed with R software (V.4.0.2; R Core Team, Vienna, Austria).

RESULTS

Between 01.01.2020 and 31.12.2021, 78,202 deaths were observed for all causes in residents in the ATS of Milan, of which 48% (37,848) was males and 86% (67,487) was older than 70 years of age (Table 1). From the beginning of the epidemic, 30% (23,495) died in the first semester of 2020 (containing the first wave of the epidemic), 26% (19,988) in the second semester of 2020 (containing second wave of the epidemic), 23% (18,189) in the first semester of 2021 (containing the third wave of the epidemic), and 21% (16,530) in the second semester of 2021 (containing the fourth wave of the epidemic). Among 43,483 deaths in 2020, 100% were classified in ReNCaM, of which 8,815 (20%) due to COVID-19. Among 34,719 deaths in 2021, about 23% (7,905) were classified in ReNCaM, of which 94% happened in the first semester of 2021. Among subjects who died during follow-up, COVID-19 hospitalizations in 2020 and 2021 were, respectively, 7,831 (18%) and 5,995 (17%), while confirmed COVID-19 cases were 9,945 (23%) and 7,603 (22%).
There were 106 patients with positive swabs (among 17,548) and 44 patients with hospitalization (among 13,826) posterior to death which have been considered as negative subjects and/or subjects with no hospitalizations.
Minimum and maximum time interval between the date of positive swab and date of death were [0-531], ROC

curve for all possible values is described in Figure S3 (for simplicity, only the time intervals between 0 and 100 were plotted). Increasing the time interval will increase the sensitivity of the test, the ability of a test to correctly identify those who actually died due to or involving COVID-19, but will decrease specificity, the ability of the test to correctly identify those who do not died due to or involving COVID-19. Sensitivity varied between 0% and 88%, while specificity varied between 95% and 99%. Given that sensitivity did not improve considerably above 60 days of lag, the optimal time interval was considered as the one below 60 days of lag which maximize sensitivity, specificity, and Youden index. The optimal time interval was 0-61 with a sensitivity of 86% and a specificity of 96%, i.e., defining a potential COVID-19 death as that occurring at most 61 days from a positive swab, 86% COVID-19-related deaths was correctly classified.
Minimum and maximum time interval between the date of hospitalization due to COVID-19 and of death were [0-510], ROC curve for all possible values is described in Figure S4. Again, increasing the time interval will increase the sensitivity of the test, but will decrease specificity. Hospitalization due to COVID-19 alone is a weaker marker of death than positive swabs with sensitivity varying between 0% and 74% and specificity varying between 96% and 99%. The optimal time interval was 0-11 with a sensitivity of 73% and a specificity of 97%, i.e., defining a potential COVID-19 death as that occurring at most 11 days from a COVID-19 hospitalization, 73% COVID-19-related deaths was correctly classified. Combining the time periods selected above, a potential COVID-19 death was defined as death occurring at most 61 days from a positive swab or 11 days from a COVID-19 hospitalization with an overall sensitivity of 90% and a specificity of 95%.
Weekly moving average of daily COVID-19-related deaths and potential COVID-19 deaths are represented in Figure 1 (weekly moving average of daily COVID-19-not-related deaths and potentially not due to or involving COVID-19 are represented in Figure S5). RMSE were 3.6 for both deaths due to or involving COVID-19 and for deaths not due to COVID-19.
Stratifying the cohort by time of death (Table 2), it was found a smaller optimal time interval, according to positivization, in 2020 compared to 2021 (66 vs 88 days from a positive swab), but a higher optimal time interval, according to hospitalization (16 vs 3 days from a COVID-19 hospitalization). Optimal time interval, according to positivization, increased with age, except among subjects with 80+ years of age which resulted comparable to the overall cohort. On the other hand, optimal time interval, according to hospitalization, was considerably smaller (around 3 or 4 days) for subjects with comorbidities.

CHARACTERISTICS	DEATHS FOR ALL CAUSES (N. 78,202)	
	N.	%
Sex		
Male	37,848	48
Age (years)		
0-49	1,666	2
50-64	5,401	7
65-79	18,599	24
80+	52,536	67
Number of comorbidities		
0	8,929	11
1	64,330	83
≥2	4,943	6
Epidemic Period		
01.01.2020-30.06.2020 (includes first wave)	23,495	30
01.07.2020-31.12.2020 (includes second wave)	19,988	26
01.01.2021-30.06.2021 (includes third wave)	18,189	23
01.07.2021-31.12.2021 (includes fourth wave)	16,530	21
Deaths for all causes		
2020	43,483	56
2021	34,719	44
Deaths classified in ReNCaM		
2020	43,483	100
2021	7,905	15
COVID-19 deaths classified in ReNCaM		
2020	8,815	20
2021	1,306	17
Hospitalizations due to COVID-19		
2020	7,831	18
2021	5,995	17
COVID-19 cases		
2020	9,945	23
2021	7,603	22
Estimated deaths as of 2015-2019		
2020*	33,167	50
2021	33,176	50
Excess deaths as of 2015-2019		
2020*	10,209	31
2021	1,543	5
Time period range between death and positive swab		
min-max	[0-531]	-
Time period range between death and hospitalization		
min-max	[0-510]	-

* To avoid misalignment in the time series, 29th of February of leap years (amounting to 100 deaths in 2020) was removed. / *Per evitare disallineamenti nelle serie temporali, il 29 febbraio degli anni bisestili (per un totale di 100 decessi nel 2020) non è stato considerato.*

Table 1. Descriptive statistics of death for all causes between 01.01.2020 and 31.12.2021 in residents in the ATS of Milan.
Tabella 1. Statistiche descrittive dei decessi per tutte le cause avvenuti fra il 01.01.2020 e il 31.12.2021 nei residenti dell'ATS di Milano.

TIME OF DEATH	OPTIMAL TIME INTERVAL	SENSI-TIVITY	SPECIFI-CITY	YOU-DEN INDEX
2020-2021	0-61;0-11	0.90	0.95	0.86
2020	0-66;0-16	0.90	0.95	0.85
1th semester of 2020	0-62;0-20	0.87	0.95	0.82
2nd semester of 2020	0-59;0-13	0.93	0.96	0.89
2021	0-88;0-3	0.96	0.93	0.89
Age				
0-49	0-55;0-10	0.86	0.98	0.84
50-64	0-78;0-12	0.93	0.96	0.89
65-79	0-84;0-3	0.94	0.95	0.88
80+	0-55;0-15	0.89	0.95	0.84
Comorbidities				
0	0-76;0-16	0.87	0.95	0.83
1	0-61;0-3	0.91	0.95	0.86
2+	0-75;0-4	0.89	0.96	0.85

Table 2. Optimal time intervals, according to positivization and hospitalization, by time of death, age, and comorbidity.
Tabella 2. Intervalli temporali ottimali, secondo la positivizzazione o il ricovero, per data di decesso, età e comorbitità.

In the first semester of 2020, an abnormal peak of COVID-19-not-related deaths was found (Figure S5), where the most prevalent ICD-10 codes were: chronic ischaemic heart disease, malignant neoplasm of bronchus and lung, pneumonia (organism unspecified), complications and ill-defined descriptions of heart disease, other chronic obstructive pulmonary disease, unspecified dementia and Alzheimer disease.
To evaluate the proposed method, a time-series approach was used to estimate the excess mortality during the COVID-19 outbreak in comparison with the pre-outbreak period 2015-2019. During 2015-2019, 166,866 deaths for all cause were observed in residents in the ATS of Milan, 33,373 on average per year, of which 491 dues to or involving Influenza or Pneumonia. Estimated deaths as of 2015-2019 would be, in a counterfactual situation where COVID-19 had not happened, 33,167 (-31% than observed) in 2020 and 33,176 (-5% than observed) in 2021 (Table 1). Figure 2 provides the time series of daily COVID-19-related deaths, estimated deaths as of 2015-2019, and excess deaths as of 2015-2019 for the year 2020. Figure 2 shows that the abnormal peak found in the first semester of 2020, not directly attributable to COVID-19 as from ICD-10 cause, seems actually attributable to COVID-19 according to the time-series approach chosen in this study. The same happened at the beginning of the second wave of 2020.
Figure 3 provides the weekly moving average of potential COVID-19 deaths, potential not-COVID-19 deaths, estimated deaths as of 2015-2019 and excess deaths as of 2015-2019 for the year 2021. The trends, potential COVID-19 deaths vs excess deaths and potential not-COVID-19 deaths vs estimated deaths, are similar whichever of the two methods proposed is chosen. RMSE were 15.8 for both potential COVID-19 deaths vs excess deaths and potential deaths not due to or involving COVID-19 vs estimated deaths.
Figure 4 represents the effective reproduction number (R_t) by time of death calculated from overall mortality, from potential COVID-19 deaths, and the number of potential COVID-19 deaths. The R_t trends are quite different from each other and the ones calculated from potential COVID-19 deaths showed considerable confidence intervals in specific periods of the epidemic characterized by low number of deaths. However, it increases with the numbers of potential COVID-19 deaths and suggests additional periods of an uncontrolled level of the epidemic that are not highlighted by the one calculated from overall mortality; for example, in December 2021, where the number of confirmed cases increased exponentially in contrast with deaths for all causes.

DISCUSSION

This study aims at propose a method to identify deaths due to or involving COVID-19 in absence of the death certificates by selecting the best time interval between the date of deaths and the date of positive swab and/or COVID-19 hospitalization. ISS defines deaths due to COVID-19 as those occurring in confirmed positive patients, in patients with specific COVID-19 symptoms, in absence of other causes different from the disease and of clinical recovery period.[4] A report form ISS and Istat reported that median time between death and a positive swab is 12 days and that 89% of deaths of patients tested positive for COVID-19 happened within 30 days from a positive swab.[8] Here, 86% of COVID-19 deaths was found to have happened within 61 days from a positive swab and that 73% of COVID-19 deaths have happened within 11 days from a COVID-19 hospitalization. The major distinction of these results could be attributed to the different definitions of COVID-19: Istat considers only deaths due to COVID-19, while the aim of the present study was to evaluate deaths due to or involving COVID-19. The management of all deaths attributable to COVID-19, so not only those due to COVID-19, could therefore have lengthened the window between the date of swab and the date of death. Combining the time periods, an overall sensitivity of 90% and a specificity of 95% were obtained, with an increase of 4% in sensitivity including the information on hospitalization versus a decrease of only 1% in specificity.
The proposed method aimed to individually identify whether a death was attributable to COVID-19 before the death certificate is available. Usual methods do not

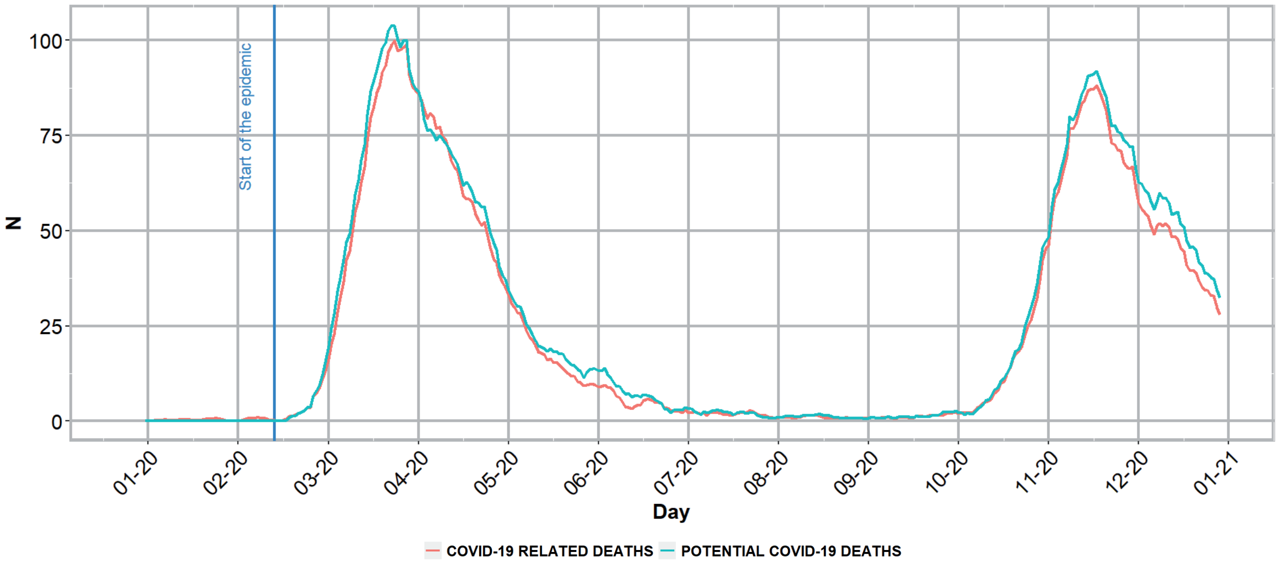

Figure 1. Weekly moving average of daily COVID-19-related deaths and potentially due to COVID-19 (occurring at most 61 days from a positive swab or 11 days from a COVID-19 hospitalization).
Figura 1. Media mobile settimanale dei decessi giornalieri dovuti a COVID-19 e potenzialmente dovuti a COVID-19 (occorsi al massimo a 61 giorni dal tampone positivo o 11 giorni dalla data di ricovero per COVID-19).

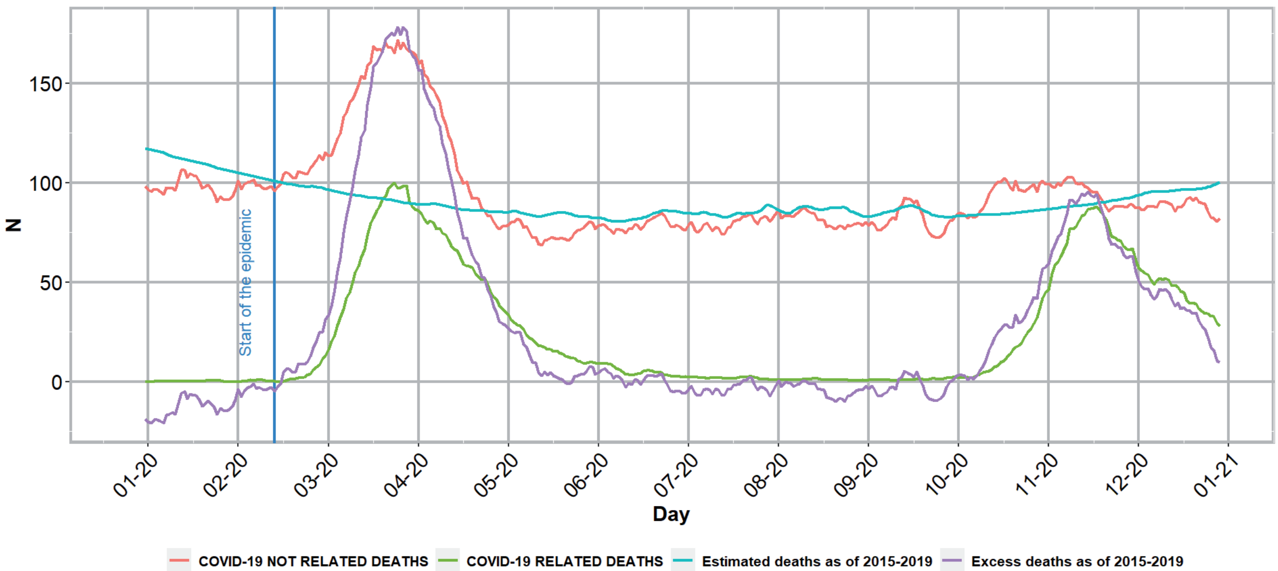

Figure 2. Weekly moving average of daily COVID-19-related deaths, estimated deaths as of 2015-2019, and excess deaths as of 2015-2019 for the year 2020.
Figura 2. Media mobile settimanale dei decessi giornalieri dovuti a COVID-19, decessi stimati rispetto al periodo 2015-2019 ed eccessi di mortalità rispetto al 2015-2019 per l'anno 2020.

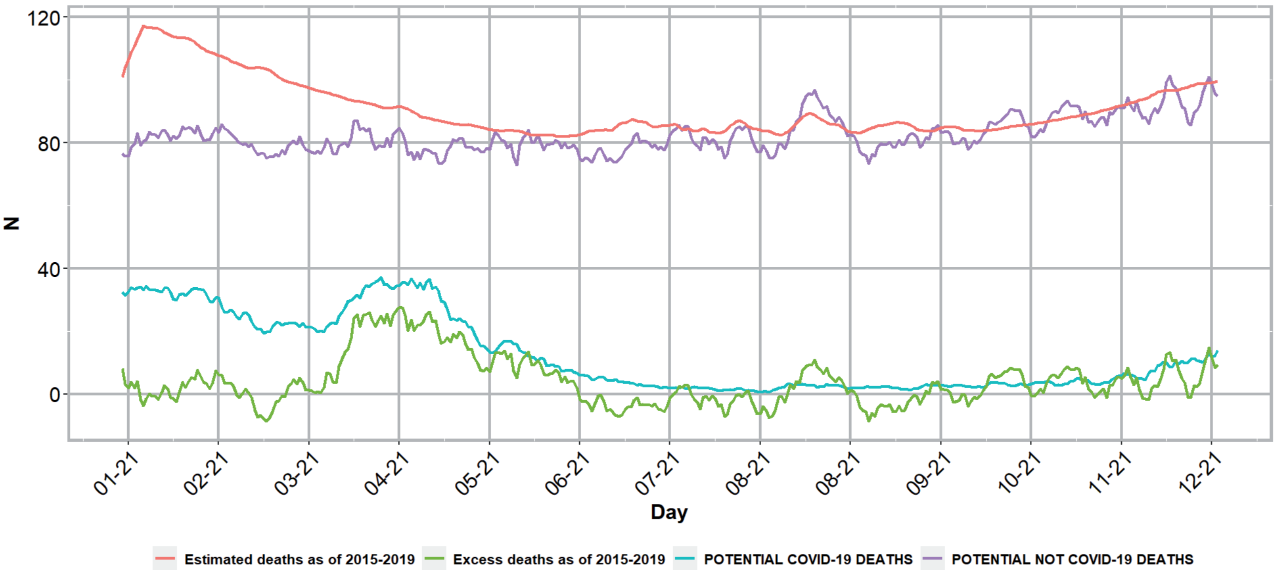

Figure 3. Weekly moving average of potential COVID-19 deaths, potential not-COVID-19 deaths, estimated deaths as of 2015-2019, and excess deaths as of 2015-2019 for the year 2021.
Figura 3. Media mobile settimanale dei decessi potenzialmente dovuti a COVID-19, decessi potenzialmente non dovuti a COVID-19, decessi stimati rispetto al periodo 2015-2019 ed eccessi di mortalità rispetto al 2015-2019 per l'anno 2021.

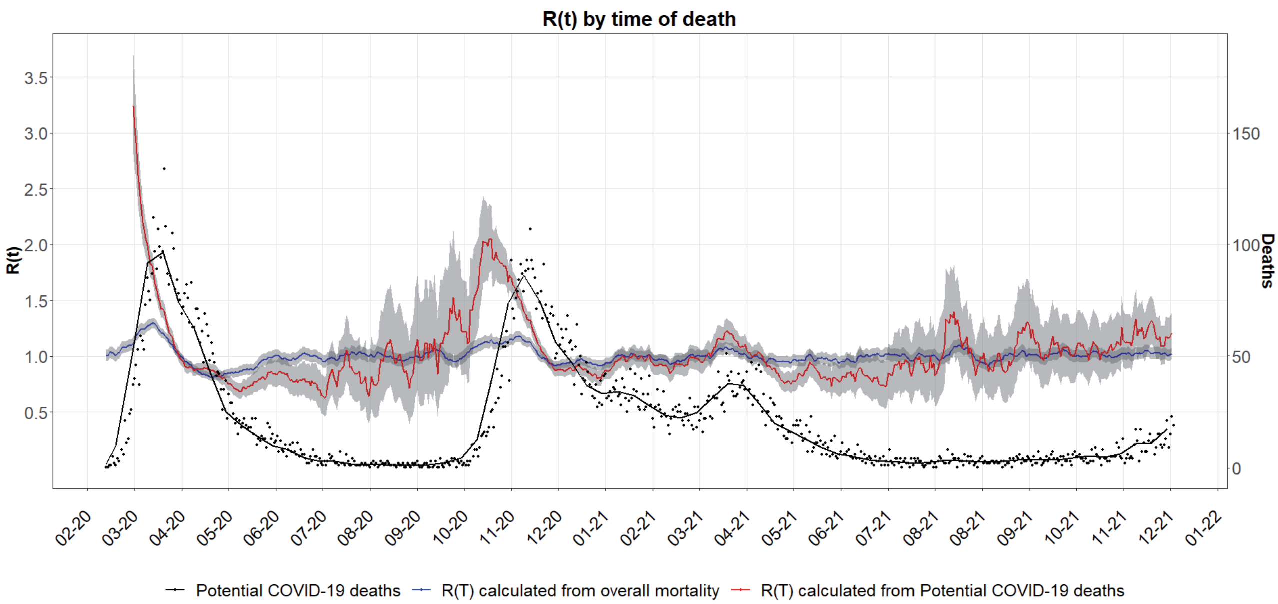

Figure 4. Effective reproduction number (R_t), by time of death calculated from overall mortality, from potential COVID-19 deaths, and number of potential COVID-19 deaths.
Figura 4. Numero di riproduzione effettiva (R_t), per data del decesso calcolata dalla mortalità totale, dai decessi potenzialmente dovuti a COVID-19 e numero di decessi potenzialmente dovuti a COVID-19.

permit to evaluate deaths individually, but only to evaluate the excess attributable to COVID-19 calculating the difference between overall mortality and expected deaths as arising from the same process in the pre-outbreak period.[5-7,9,10] The proposed method showed a good fit in comparing daily COVID-19 related deaths and potential COVID-19 deaths with an error of around 4 deaths (small compared to the range of observations [0-129]).
In addition, the proposed method was validated by comparing potential COVID-19 deaths with excess deaths as from 2015-2019, showing a good fit with an error of around 16 deaths in 2021.
Due to the dynamics of COVID-19, where asymptomatic cases are not properly tested and limited testing capacity might underestimate the real burden of the disease, it is sometimes difficult to calculate the effective reproduction number. It has been suggested that overall mortality would be a better indicator than the number of positive patients as independent from testing strategies.[32]
However, overall mortality does not evaluate the pressure of the epidemic on death rates, as they might increase or decrease due to other mechanisms different from COVID-19; for example, for the summer excess in the death rates. This study shows as effective reproduction number calculated from overall mortality is quite different from that calculated from potential COVID-19 deaths, suggesting additional periods of an uncontrolled level of the epidemic that might be masked using only overall mortality.
The present study reveals a peak of COVID-19-not-related deaths in the first wave and at the beginning of the second wave of the epidemic (which, in contrast, should be described as COVID-19-related deaths according to the time-series analysis). These were mostly ascribed as due to chronic ischaemic heart disease, malignant neoplasm of bronchus and lung, pneumonia (organism unspecified), complications and ill-defined descriptions of heart disease, other chronic obstructive pulmonary disease, unspecified dementia, and Alzheimer disease. This is indicative of potential delays in diagnosis and/or treatment, but also of a substantial degree of misclassification in 2020. Asta and colleagues[18] found excesses in cause-specific mortality either for acute and chronic pathologies, which might be ascribed to misclassification of death causes, as highlighted also by other authors.[33,34] This is, in fact, a potential limitation of this study, which is based on the specific testing strategy of the period. This underestimation problem has been completely solved during the epidemic and the method proposed was equivalent to the time-series approach. Moreover, considering also the hospitalization in addition to COVID-19 positivity, this problem was overcome.
Another limitation of this study is that it was considered the same time period during the entire course of the epidemic. According to a report from ISS,[35] the median time between death and a positive swab changed over the course of the epidemic, from 7 days during the first wave to 14 days in summer 2020 and 11 days during the second wave, reflecting in a reduction in access times to diagnosis and treatment.
In conclusion, here it is proposed a simple, feasible, low-cost strategy to identify deaths due to or involving COVID-19 in absence of the death certificates which can be used as public health measures as an alternative to overall mortality.

Conflict of interest: none declared.

REFERENCES

1. Cerqua A, Di Stefano R. When did coronavirus arrive in Europe? Stat Methods Appl 2022;31(1):181-95.
2. Emergenza Coronavirus. Available from: https://emergenze.protezionecivile.gov.it/it/sanitarie/coronavirus (ultimo accesso: 23.03.2022).
3. EUROMOMO EuroMOMO Bulletin, Week 10, 2022. EUROMOMO. Available from: https://euromomo.eu/dev-404-page/ (ultimo accesso: 23 marzo 2022).
4. Istituto superiore di sanità. Rapporto ISS COVID-19 n. 10/2021. COVID-19: rapporto ad interim su definizione, certificazione e classificazione delle cause di morte. Aggiornamento del Rapporto ISS COVID-19 n. 49/2020. Versione del 26 aprile 2021. Roma, ISS., 2021 Available from: https://bit.ly/3PJRP8w (ultimo accesso: 09.02.2022).
5. Odone A, Delmonte D, Gaetti G, Signorelli C. Doubled mortality rate during the COVID-19 pandemic in Italy: quantifying what is not captured by surveillance. Public Health 2021;190:108-15.
6. Modi C, Böhm V, Ferraro S, Stein G, Seljak U. Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. Nat Commun 2021;12(1):2729.
7. Karlinsky A, Kobak D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. In: Davenport MP, Lipsitch M, Lipsitch M, Simonsen L, Mahmud A (eds). eLife 2021;10:e69336.
8. Istat, ISS. Impatto dell'epidemia covid-19 sulla mortalità totale della popolazione residente. Anni 2020-2021 e gennaio 2022. Roma: Istat-ISS; 2022. Disponibile : https://www.istat.it/it/archivio/266865
9. Sandrini M, Andreano A, Murtas R, et al. Assessment of the Overall Mortality during the COVID-19 Outbreak in the Provinces of Milan and Lodi (Lombardy Region, Northern Italy). Epidemiol Prev 2020;44(5-6) Suppl. 2:244-51.
10. Levi M, Cipriani F, Romeo G, Balzi D. Analysis of the excess mortality and factors associated with deaths from COVID-19 versus other causes in Central Tuscany (Italy) in 2020. Epidemiol Prev 2021;45(6):496-503.
11. Office for National Statistics. Impact of registration delays on mortality statistics in England and Wales. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/impactofregistrationdelaysonmortalitystatisticsinenglandandwales/2020 (last accessed: 24.03.2022).
12. Regolamento (CE) n. 1338/2008 del Parlamento europeo e del Consiglio del 16 dicembre 2008 relativo alle statistiche comunitarie in materia di sanità pubblica e di salute e sicurezza sul luogo di lavoro. Available from: https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32008R1338&from=CS#:~:text=Oggetto-,1.,sicurezza%20sul%20luogo%20di%20lavoro.
13. Italian Ministry of Economy. Il Piano Nazionale di Ripresa e Resilienza (PNRR). Available from: https://www.mef.gov.it/focus/Il-Piano-Nazionale-di-Ripresa-e-Resilienza-PNRR/ (last accessed: 28.03.2022).
14. Chellini E. Un cambiamento epocale nel flusso della mortalità ci aspetta. Ma siamo preparati? Epidemiol Prev 2021;45(3):130-31.
15. Deaths in the UK | Coronavirus in the UK. Available from: https://coronavirus.data.gov.uk/details/deaths (last accessed: 28.03.2022).
16. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020;368(6490):489-93.
17. Achilleos S, Quattrocchi A, Gabel J, et al. Excess all-cause mortality and COVID-19-related mortality: a temporal analysis in 22 countries, from January until August 2020. Int J Epidemiol 2022;51(1):35-53.

18. Asta F, Michelozzi P, De Sario M,et al. Impatto dell'epidemia di COVID-19 sulla mortalità totale e per causa a Roma nel 2020. Epidemiol Prev 2022;46(1-2):59-67.
19. Sistema Informativo Socio Sanitario Regione Lombardia. Anagrafe Regionale degli assistiti e delle strutture. Anagrafe Regionale degli Assistiti e delle strutture. Available from: https://www.siss.regione.lombardia.it/wps/portal/site/siss/il-sistema-informativo-socio-sanitario/principali-servizi-offerti/anagrafe-regionale-degli-assistiti-e-delle-strutture
20. Office for National Statistics. Deaths registered weekly in England and Wales, provisional. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending27november2020#deaths-registered-by-week (last accessed: 17.03.2022).
21. Integrazione dei sistemi di classificazione adottati per la codifica delle informazioni cliniche contenute nella scheda di dimissione ospedaliera e per la remunerazione delle prestazioni ospedaliere in conseguenza della nuova malattia da SARS-CoV-2 (COVID-19). Modifiche al decreto del 18 dicembre 2008. Italian Law Decree 28.10.2020. Available from: https://www.gazzettaufficiale.it/eli/id/2021/02/01/21A00441/sg
22. 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.
23. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb) 2016;26(3):297-307.
24. Vandeput N. Data Science for Supply Chain Forecasting 2nd Edition. De Gruyter 2021Available from: https://www.google.com/search?q=Data+Science+for+Supply+Chain+Forecasting+2nd+Edition (last accessed: 21.03.2022).
25. Koutsakos M, Wheatley AK, Laurie K, Kent SJ, Rockman S. Influenza lineage extinction during the COVID-19 pandemic? Nat Rev Microbiol 2021;19(12):741-42.
26. Onozuka D, Tanoue Y, Nomura S, et al. Reduced mortality during the COVID-19 outbreak in Japan, 2020: a two-stage interrupted time-series design. Int J Epidemiol 2022;51(1):75-84.
27. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. Int J Epidemiol 2013;42(4):1187-95.
28. Jung Y, Hu J. A K-fold Averaging Cross-validation Procedure. J Nonparametr Stat 2015;27(2):167-79.
29. Brumback BA, Ryan LM, Schwartz JD, Neas LM, Stark PC, Burge HA. Transitional Regression Models, with Application to Environmental Time Series. J Am Stat Ass 2000;95(449):16-27.
30. Inglesby TV. Public Health Measures and the Reproduction Number of SARS-CoV-2. JAMA 2020;323(21):2186-87.
31. Na J, Tibebu H, De Silva V, Kondoz A, Caine M. Probabilistic approximation of effective reproduction number of COVID-19 using daily death statistics. Chaos Solitons Fractals 2020;140:110181.
32. Grande E, Fedeli U, Pappagallo M, et al. Variation in Cause-Specific Mortality Rates in Italy during the First Wave of the COVID-19 Pandemic: A Study Based on Nationwide Data. Int J Environ Res Public Health 2022;19(2):805.
33. Italian National Insitute of Statistics. Impatto dell'epidemia covid-19 sulla mortalità: cause di morte nei deceduti positivi a sars-cov-2. Roma: Istat; 2020. Available from: https://www.istat.it/it/files/2020/07/Report_ISS_Istat_Cause-di-morte-Covid.pdf
34. Italian National Health Institute. Il case fatality rate dell'infezione SARS-CoV-2 a livello regionale e attraverso le differenti fasi dell'epidemia in Italia. Rapporto ISS COVID-19 n. 1/2021. Roma: ISS, 2021. Available from: https://www.iss.it/documents/20126/0/Rapporto+ISS+COVID-19+n.+1_2021.pdf/eef324b0-983d-c257-96fd-e8d430e1ca82?t=1612179039051 (last accessed: 22.02.2022).