# BMJ Open

# Trend of pneumonia diagnosis in emergency departments as a COVID-19 surveillance system: a time series study

Rossella Murtas, Adriano Decarli ![ORCID], Antonio Giampiero Russo ![ORCID]

Epidemiology Unit, Agency for the Protection of Health of the Metropolitan Area of Milan, Milano, Lombardia, Italy

**Correspondence to**
Dr Antonio Giampiero Russo;
agrusso@ats-milano.it

## ABSTRACT

**Objective** In Italy, the first diagnosis of COVID-19 was confirmed on 20 February 2020 in the Lombardy region. Given the rapid spread of the infection in the population, it was suggested that in Europe, and specifically in Italy, the virus had already been present in the last months of 2019. In this paper, we aim to evaluate the hypothesis on the early presence of the virus in Italy by analysing data on trends of access to emergency departments (EDs) of subjects with a diagnosis of pneumonia during the 2015–2020 period.

**Design** Time series cohort study.

**Setting** We collected data on visits due to pneumonia between 1 October 2015 and 31 May 2020 in all EDs of the Agency for Health Protection of Milan (ATS of Milan). Trend in the winter of 2019–2020 was compared with those in the previous 4 years in order to identify unexpected signals potentially associated with the occurrence of the pandemic. Aggregated data were analysed using a Poisson regression model adjusted for seasonality and influenza outbreaks.

**Primary outcome measures** Daily pneumonia-related visits in EDs.

**Results** In the studied period, we observed 105 651 pneumonia-related ED visits. Compared with the expected, a lower occurrence was observed in January 2020, while an excess of pneumonia visits started in the province of Lodi on 21 February 2020, and almost 10 days later was observed in the remaining territory of the ATS of Milan. Overall, the peak in excess was found on 17 March 2020 (369 excess visits compared with previous years, 95% CI 353 to 383) and ended in May 2020, the administrative end of the Italian lockdown.

**Conclusions** An early warning system based on routinely collected administrative data could be a feasible and low-cost strategy to monitor the actual situation of the virus spread both at local and national levels.

## Strengths and limitations of this study

► This is a large retrospective study conducted in the territory of the Agency for the Health Protection of Milan, which covers a total population of 3.48 million inhabitants.

► Evaluation of the early presence of COVID-19 was performed using long-term time series and Poisson regression models adjusted for seasonality and influenza epidemics.

► Available data do not allow differentiation of diagnoses of viral pneumonia versus other types of pneumonia, in particular due to potential misclassification of the International Classification of Diseases, Ninth Revision codes.

► We cannot exclude the possible presence of unmeasured variables that may affect the number of emergency department visits.

as introducing incentives to smart working wherever possible in Lombardy. The national lockdown was implemented on 8 March, with closure of all non-essential recreational and commercial activities, schools and universities and prohibition of travel across the country. On 21 March, all non-essential activities were closed nationally. A gradual easing of the restrictive measures began on 4 May, even if until 4 June rules on social distancing, obligation to use masks, travel limitations on the national territory, and closure of schools and other services for childhood remained in place.

To date, COVID-19 surveillance systems have relied only on testing and contact tracing, but these methods imply that the specific virus infection is already diagnosed in the population. In order to promptly implement preventive policies, we need methods that can detect the presence of the virus as early as possible. Recently, researchers have proposed several approaches to early recognition of a COVID-19 outbreak. These include wastewater-based epidemiology studies[1–7] and studies on digital data from Google, Twitter or social media apps in order to recognise

## INTRODUCTION

China confirmed the first cases of COVID-19 in late December 2019. In Italy, the first diagnosis was confirmed on 20 February 2020 in Lombardy. Four days later, the rapid spread of the infection led to defining a 'red zone' (highest risk zone), the area most affected by the pandemic, with related rules limiting displacements and activities that involve gathering of people and travelling, as well

early traces of the virus in the population.[8 9] In March 2020, in order to justify the rapid spread of COVID-19 in Lombardy, it was suggested that in Europe, and specifically in Italy, the virus had already been present in the last months of 2019. The Italian Institute of Health (ISS) tested samples of wastewater from Milan and Turin and showed that genetic virus traces were already present in December.[3 10] In Spain, scientists found virus traces in wastewater collected in mid-January[11] and late February,[5] when evidence of a Spanish outbreak was only starting to circulate.

The major concern in timely detecting patients with COVID-19 is the similarity between influenza-like illness (ILI) and symptoms of COVID-19.[12] Common symptoms are fever, fatigue, cough, dyspnoea, sore throat and myalgia. On the other hand, ILI usually causes mild to severe illness, while COVID-19 may cause more serious conditions especially in the fragile population.[12 13] One of the most common complications is pneumonia, which in the case of influenza may appear as a primary complication or as secondary bacterial pneumonia.[14]

In this study we tested the hypothesis on the early presence of the virus in Italy, before the first diagnosed case, by using long-term time series extracted from the administrative health data of the Agency for Health Protection of Milan (ATS of Milan). The time series of visits to emergency departments (EDs) due to pneumonia in the winter of 2019–2020 were compared with those in the previous 4 years in order to assess unexpected signals potentially related to the occurrence of the pandemic. Most importantly, our aim is also to evaluate the role of this approach as a surveillance system capable of promptly detecting the initial signals of the epidemic outbreak so that important preventive public health measures can be defined early in the pandemic.

## METHODS

This is a retrospective study conducted in the territory of the ATS of Milan, which covers 193 municipalities in the Northern Italian region of Lombardy, with a total population of 3.48 million inhabitants. The study area includes the municipality of Codogno, where the first Italian COVID-19 case occurred. We collected visits due to pneumonia between 1 October 2015 and 31 May 2020 in all EDs of the ATS of Milan (all visits, from residents and from non-residents of the ATS of Milan). In order to evaluate the overall incidence in the population of the ATS of Milan, we also included visits that residents carried out to hospitals not located in the territory of the ATS of Milan.

We defined visits due to pneumonia according to the International Classification of Diseases, Ninth Revision (ICD-9)[15] as viral pneumonia (ICD-9 480), bacterial pneumonia (ICD-9 481–482), pneumonia due to other specified organism (ICD-9 483), pneumonia in infectious diseases classified elsewhere (ICD-9 484), bronchopneumonia organism unspecified (ICD-9 485), pneumonia organism unspecified (ICD-9 486), pneumonia in

influenza (ICD-9 487.0), unspecified alveolar and parietoalveolar pneumonopathy (ICD-9 516.9), acute respiratory failure (ICD-9 518.81), other pulmonary insufficiency not elsewhere classified (ICD-9 518.82) and congenital pneumonia (ICD-9 770.0). Aggregated daily data were collected using current ED databases. To evaluate the total burden of the epidemic in the territory, a patient visiting an ED on two different days was counted twice. No individual-level data were used and patients cannot be identified from aggregated data which do not contain low counts (ie, days with less than five visits). For this reason, and according to the Italian legislation, this study was not submitted for ethics approval.[16]

The number of COVID-19 cases by province of residence, as defined on 17 August 2020, was collected from the ATS of Milan through a web-based platform, specifically developed since the beginning of the outbreak, to trace positive and negative cases as well as related contacts. Prevalence was calculated as the number of cases over the number of residents on 1 January 2020 per 1000 inhabitants.[17]

### Patient and public involvement
Patients were not involved in this research.

### Statistical methods
Aggregated data on daily ED visits were analysed using a Poisson regression model adjusted for seasonality and influenza epidemics.[18] Seasonality was controlled for by including Fourier terms, a series of sine-cosine functions able to approximate periodicity.[19 20] Fourier terms specification does not depend on original data. They are particularly useful because they can be easily integrated in postsample forecasting equations.[19] Weekly data on ILI notifications were taken from the National Health Service Sentinel System (InfluNet).[21] Weekly incidence rates of ILI were expressed as the number of cases per 1000 inhabitants per week. The model was specified as the following:

$$\ln\left(\lambda_i\right) = \beta_0 + \beta_1 \times \sin\theta_i + \beta_2 \times \cos\theta_i + \beta_3 \times \mathrm{ILI}_i$$

where $\lambda_i$ is the expected value of the Poisson variable $Y_i$ (the number of visits on day $i$), $\sin\theta_i$ and $\cos\theta_i$ are the Fourier terms where $\theta_i = 2 \times \pi \times i/365$, and $\mathrm{ILI}_i$ is the weekly incidence rate of ILI for day $i$. The ILI rate was provided on a weekly basis; thus, in the model above, $\mathrm{ILI}_i$ is repeated seven times, for each day of the week.

Data sets were divided into training (data from 1 October 2015 to 30 September 2019) and validation (data from 1 October 2019 to 31 May 2020) sets. We first estimated the parameters in the training set and then predicted the outcome in the validation set. This projection represents a counterfactual situation where COVID-19 had not happened. The number of excess visits was calculated as the difference between observed and predicted values between 1 October 2019 and 31 May 2020. We calculated 95% prediction interval (PI) by sampling from the uncertainty distributions of the

estimated model parameters.[18 22] Results were displayed overall and by province of residence: in the city of Milan, in the province of Milan, in the province of Lodi (the province of the first COVID-19 case) and outside the ATS of Milan. In order to evaluate pneumonia excess attributed to COVID-19 adjusting for natural fluctuations in the number of visits, we performed a sensitivity analysis dividing the data set into training (from 1 October 2015 to 30 September 2018), validation set 1 (from 1 October 2018 to 30 September 2019) and validation set 2 (from 1 October 2019 to 31 May 2020). We performed, on these three data sets, the same analyses described above and compared the pneumonia excess found in validation set 2 with those found in validation set 1 according to t-test statistics and number of statistically significant excesses.

All analyses were performed with R software (V.4.0.2; R Core Team, Vienna, Austria) and Fourier terms were calculated using the Fourier function in the R package forecast[23] after specifying the daily count as a time series with annual periodicity.

## RESULTS

Between 1 October 2015 and 31 May 2020 (amounting to 1704 days), we observed 105 651 pneumonia-related visits, of which 80 086 (76%) were in the training set and 25 565 (24%) were in the validation set. Influenza epidemics were stronger in the epidemic periods of 2015–2019 compared with 2019–2020, with a maximum ILI rate of 14.7 and 12.6 new cases per 1000 inhabitants per week, respectively. In the overall territory of the ATS of Milan we found a COVID-19 prevalence of 14.6 cases per 1000 inhabitants (table 1), similar to the city of Milan and to the province of Milan, but approximately half of that found in the province of Lodi, with 26.3 cases per 1000 inhabitants.
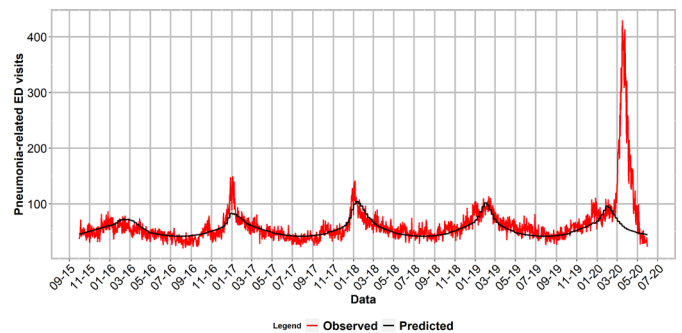


**Figure 1** Daily observed and predicted emergency department (ED) visits due to pneumonia between 1 October 2015 and 31 May 2020. Predictions were obtained by a Poisson regression model adjusted for seasonality and influenza epidemics. Time series in the overall territory of the Agency for Health Protection of Milan (ATS of Milan).

In figure 1 we present the daily observed and predicted ED visits due to pneumonia in the overall period and in the overall territory of the ATS of Milan (time series by province of residence can be found in online supplemental figure 1A–D). Daily visits showed a typical seasonal pattern over the years, well captured by seasonality and influenza epidemics (black line in figure 1). The demographic characteristics of patients by province of residence and period of comparison are described in table 1. Between 1 October 2015 and 31 May 2020, 38 972 (37%) pneumonia-related visits were from residents of the city of Milan, 48 678 (46%) from the province of Milan, 7609 (7%) from the province of Lodi and 10 392 (10%) from residents outside the territory of the ATS of Milan. Majority of visits were attributed to residents of Milan and the province of Milan, with daily time series (online supplemental figure 1A,B) resembling that of the overall territory of ATS (figure 1). Overall, excess pneumonia

**Table 1** Demographic characteristics of patients who visited the emergency departments of the ATS of Milan for pneumonia, by province of residence, between 1 October 2015 and 31 May 2020

| | Overall territory | City of Milan | Province of Milan | Province of Lodi | Outside the ATS of Milan |
|---|---|---|---|---|---|
| | N=105 651 | n=38 972 (37%) | n=48 678 (46%) | n=7609 (7%) | n=10 392 (10%) |
| Missing demographic information | 16 | 6 | 5 | – | 5 |
| Sex: male (%) | 56 911 (54) | 20 170 (52) | 26 645 (55) | 4178 (55) | 5918 (57) |
| Age, mean (SD) | 63 (28) | 64 (29) | 63 (27) | 65 (25) | 58 (27) |
| Age class (years) (%) | | | | | |
| <17 | 13 082 (12) | 5234 (13) | 6030 (12) | 618 (8) | 1200 (12) |
| 17–40 | 6762 (6) | 2225 (6) | 2816 (6) | 529 (7) | 1192 (11) |
| 40–70 | 27 590 (26) | 8913 (23) | 12 831 (26) | 2301 (30) | 3545 (34) |
| ≥70 | 58 201 (56) | 22 594 (58) | 26 996 (56) | 4161 (55) | 4450 (43) |
| COVID-19 prevalence per 1000 inhabitants | 14.6 | 13.2 | 13.1 | 26.3 | – |

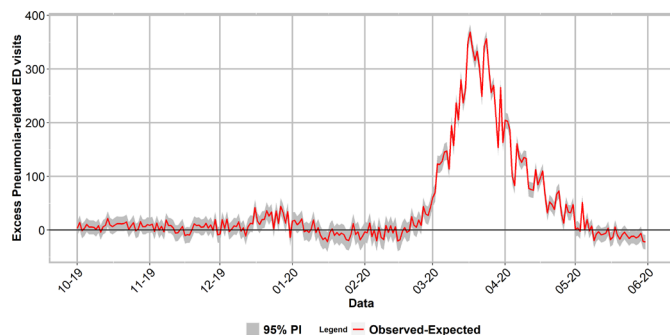ATS of Milan, Agency for Health Protection of Milan.

**Figure 2** Daily excess visits and 95% prediction interval (PI) between 1 October 2019 and 31 May 2020 in the territory of the Agency for Health Protection of Milan (ATS of Milan). ED, emergency department.



**Figure 3** Daily excess visits and 95% prediction interval (PI) between 1 October 2018 and 31 May 2020 in the territory of the Agency for Health Protection of Milan (ATS of Milan): sensitivity analysis. ED, emergency department.

visits (figure 2) were statistically not different from the preceding years up to 15 December, and then significantly higher in the last 2 weeks of 2019. The number of ED visits due to pneumonia, excesses and PIs between 1 December 2019 and 31 May 2020 can be found in online supplemental data ('Results on validation set' columns). In January and in the beginning of February, pneumonia excesses were not statistically significant. The trend increased rapidly since 26 February, reaching a peak on 17 March 2020, with 369 additional ED visits compared with the expected (95% PI 353–383). The estimated excesses ended in May 2020 (the administrative end of the Italian lockdown). Early circulation of the virus was found in the province of Lodi, where the excesses started on 21 February 2020, while in the remaining territory of the ATS of Milan excesses started more than 10 days later, in March. Similar trends to those found for the overall territory of the ATS were found in the time series of Milan and the province of Milan (online supplemental figure 2A–D), with statistically significant pneumonia excesses in the last 2 weeks of December 2019.

The number of ED visits due to pneumonia, excesses and PIs in validation set 1 (on the subset from 1 December 2018 to 31 March 2019) and validation set 2 (on the subset from 1 December 2019 to 30 March 2020) can be found in online supplemental data ('Sensitivity Analysis Results' column). In late December 2018 we found similar excesses to those found in the same period of 2019 (figure 3). Comparing the months of December and January, we found no statistically significant differences between the two validation sets (t-test for mean difference, p=0.77). However, in December 2019 and January 2020 there were 18 days with statistically significant pneumonia excesses (of which 5 consecutive days from 19 to 23 December 2019) compared with 11 days with statistically significant pneumonia excesses in December 2018 and January 2019.

## DISCUSSION

In this work, we evaluated whether the administrative health data of the ATS of Milan supported the hypothesis on the early presence of the virus in Italy, before the
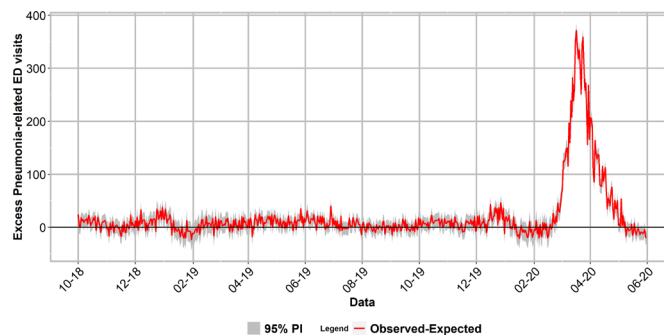
first COVID-19 case in Italy was actually diagnosed in the Lombardy region. Because some symptoms of COVID-19 and influenza are similar,[12] it has been suggested that influenza could mask early COVID-19 cases during the 2019–2020 season.[24]

Here we estimated the expected number of ED visits due to pneumonia by Poisson regression, including as predictors seasonality and influenza epidemics. Results showed a lack of excess in January 2020. The starting date of the excesses corresponded to what was already known on the territory: early circulation of the virus was found in the province of Lodi where the excess started on 21 February 2020, while the excess in the remaining territory of the ATS of Milan appeared with a delay of more than 10 days compared with the province of Lodi. Overall, the peak of excess was found on 17 March 2020. Accounting for a median incubation period of 5.1 days,[25] the hypothetical starting date of the epidemic in the territory of the ATS of Milan was on 16 February for the province of Lodi and on 25 February for the province of Milan.

Furthermore, we found excesses in late December 2019 that could be caused by other stressors than seasonality and influenza epidemics. However, sensitivity analysis showed similar excesses in December 2018, thus ascribing those found in December 2019 as not COVID-19-related. Further work needs to be done in order to explain the double peak found in December 2018 and 2019, while previous years were characterised by one higher peak only. On the other hand, in December 2019 and January 2020 we found a greater number of days with statistically significant excesses compared with the same period of the preceding year. This seems in line with the findings of December traces of COVID-19 in sewer water samples in Milan, as reported by the ISS.[10] However, further evidence has to be provided to strengthen the conclusion of an early circulation of the virus in December 2019, for example by testing those patients for the presence of SARS-CoV-2. In fact, pneumonia excesses were mostly found among residents of the province of Milan rather

than among residents of the province of Lodi, where the outbreak started in February 2020.

The city of Milan, considering the bulk of trade and human movements with China, the population density and the characteristics of urban transportation, would have been an ideal contender to be the first Italian area to be interested by the spread of the virus.

Instead, surprisingly, the city of Milan had a lower prevalence of PCR-positive cases from nasal swab and also a much lower general mortality than many adjacent areas.[26]

This could be explained by an early immunisation of residents in the province of Milan, which was indeed reached by the outbreak later than the province of Lodi. To date, the proportion of asymptomatic carriers is estimated around 15%,[27 28] but it may likely be underestimated and close to one-third of the total infected population.[29] On the other hand, after the first pandemic wave, studies in Europe found a low seroprevalence[30] (5% in Spain, 2.5% in Italy and 7.5% in Lombardy). It is hard to think that an early circulation in December, apparently related to a small excess of pneumonias, could produce immunisation of a significant proportion of the population.

A major limitation of this study is the possibility that unmeasured variables, other than seasonality and influenza epidemics, can affect the outcome, for example differences in the population structure that may modify the baseline time series. Another possible limitation is that this system is not able to differentiate the diagnosis of viral pneumonia from other diagnoses of pneumonia, in particular due to a potential misclassification of the ICD-9 codes. However, this information system based on routinely collected ED data can be a potential low-cost surveillance system for an early alert on the spread of the epidemic in the population, not only for COVID-19 but also for other respiratory pathogens. This would be critical for the early implementation of measures of prevention and containment, such as social distancing, by public health authorities and hospitals.

**ORCID iDs**
Adriano Decarli http://orcid.org/0000-0003-1451-8292
Antonio Giampiero Russo http://orcid.org/0000-0002-5681-5861

## REFERENCES

1 Patel M, Chaubey AK, Pittman CU, *et al*. Coronavirus (SARS-CoV-2) in the environment: occurrence, persistence, analysis in aquatic systems and possible management. *Sci Total Environ* 2020:142698.
2 Jafferali MH, Khatami K, Atasoy M, *et al*. Benchmarking virus concentration methods for quantification of SARS-CoV-2 in RAW wastewater. *Sci Total Environ* 2021;755:142939.
3 La Rosa G, Mancini P, Bonanno Ferraro G, *et al*. SARS-CoV-2 has been circulating in northern Italy since December 2019: evidence from environmental monitoring. *Sci Total Environ* 2021;750:141711.
4 Mao K, Zhang H, Yang Z. An integrated biosensor system with mobile health and wastewater-based epidemiology (iBMW) for COVID-19 pandemic. *Biosens Bioelectron* 2020;169:112617.
5 Randazzo W, Cuevas-Ferrando E, Sanjuán R, *et al*. Metropolitan wastewater analysis for COVID-19 epidemiological surveillance. *Int J Hyg Environ Health* 2020;230:113621.
6 Street R, Malema S, Mahlangeni N, *et al*. Wastewater surveillance for Covid-19: an African perspective. *Sci Total Environ* 2020;743:140719.
7 Mohapatra S, Menon NG, Mohapatra G, *et al*. The novel SARS-CoV-2 pandemic: possible environmental transmission, detection, persistence and fate during wastewater and water treatment. *Sci Total Environ* 2020:142746.
8 Kogan NE, Clemente L, Liautaud P. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real-time, 2020. Available: http://arxiv.org/abs/2007.00756 [Accessed 3 Nov 2020].
9 Wang L, He W, Yu X, *et al*. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect* 2020;80:639–45.
10 La Rosa G, Iaconelli M, Mancini P, *et al*. First detection of SARS-CoV-2 in untreated wastewaters in Italy. *Sci Total Environ* 2020;736:139652.
11 Chavarria-Miró G, Anfruns-Estrada E, Guix S. Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. *medRxiv* 2020.
12 CDC. Similarities and differences between flu and COVID-19. centers for disease control and prevention, 2020. Available: https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm [Accessed 10 Aug 2020].
13 Manzanares-Meza LD, Medina-Contreras O. SARS-CoV-2 and influenza: a comparative overview and treatment implications. *Bol Med Hosp Infant Mex* 2020;77:262–73.
14 Krammer F, Smith GJD, Fouchier RAM, *et al*. Influenza. *Nat Rev Dis Primers* 2018;4:1–21.
15 ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification, 2019. Available: https://www.cdc.gov/nchs/icd/icd9cm.htm [Accessed 10 Mar 2020].
16 Italia. Provvedimento del Garante n. 2 del 16 giugno. *Codice di deontologia e di buona condotta per I trattamenti di dati personali per scopi statistici e scientifici*, 2004.
17 Demo-Geodemo. - Mappe, Popolazione, Statistiche Demografiche dell'ISTAT. Available: http://demo.istat.it/ [Accessed 21 Jan 2019].
18 Weinberger DM, Chen J, Cohen T, *et al*. Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. *JAMA Intern Med* 2020;180:1336.
19 Marcilio I, Hajat S, Gouveia N. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Acad Emerg Med* 2013;20:769–77.
20 Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice*. OTexts, 2018.

21  Ministero della Salute. Influenza. Available: http://www.salute.gov.it/portale/influenza/homeInfluenza.jsp [Accessed 10 Mar 2020].

22  Lauer SA, Sakrejda K, Ray EL, *et al*. Prospective forecasts of annual dengue hemorrhagic fever incidence in Thailand, 2010-2014. *Proc Natl Acad Sci U S A* 2018;115:E2175–82.

23  Forecast: Forecasting functions for time series and linear models, 2020. Available: https://CRAN.R-project.org/package=forecast [Accessed 14 Nov 2020].

24  Coma Redon E, Mora N, Prats-Uribe A, *et al*. Excess cases of influenza and the coronavirus epidemic in Catalonia: a time-series analysis of primary-care electronic medical records covering over 6 million people. *BMJ Open* 2020;10:e039369.

25  Lauer SA, Grantz KH, Bi Q, *et al*. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020;172:577–82.

26  ATS di Milano. Valutazione degli eccessi di mortalità nel corso dell'epidemia COVID-19 nella popolazione della ATS di Milano, 2020. Available: https://www.ats-milano.it/portale/Epidemiologia/Valutazione-dellepidemia-COVID-19

27  Zhao H, Lu X, Deng Y, *et al*. COVID-19: asymptomatic carrier transmission is an underestimated problem. *Epidemiol Infect* 2020;148:e116.

28  He J, Guo Y, Mao R, *et al*. Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *J Med Virol* 2021;93:820–30.

29  Pollán M, Pérez-Gómez B, Pastor-Barriuso R, *et al*. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 2020;396:535–44.

30  Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2, 2020. Available: https://www.istat.it/it/archivio/246156 [Accessed 24 Nov 2020].