SHORT COMMUNICATION

Estimating COVID-19-induced excess mortality in Lombardy, Italy

Antonello Maruotti[1,2] · Giovanna Jona-Lasinio[3] · Fabio Divino[4] · Gianfranco Lovison[5,6] · Massimo Ciccozzi[7] · Alessio Farcomeni[8]

Abstract

We compare the expected all-cause mortality with the observed one for different age classes during the pandemic in Lombardy, which was the epicenter of the epidemic in Italy. The first case in Italy was found in Lombardy in early 2020, and the first wave was mainly centered in Lombardy. The other three waves, in Autumn 2020, March 2021 and Summer 2021 are also characterized by a high number of cases in absolute terms. A generalized linear mixed model is introduced to model weekly mortality from 2011 to 2019, taking into account seasonal patterns and year-specific trends. Based on the 2019 year-specific conditional best linear unbiased predictions, a significant excess of mortality is estimated in 2020, leading to approximately 35000 more deaths than expected, mainly arising during the first wave. In 2021, instead, the excess mortality is not significantly different from zero, for the 85+ and 15–64 age classes, and significant reductions with respect to the 2020 estimated excess mortality are estimated for other age classes.

Keywords COVID-19 · Excess mortality · All-cause mortality · Surveillance

Introduction

Official publicly available data mainly report incidence indicators. Incidence indicators measure the number of individuals with a particular condition, related with the epidemic, recorded during a given period. They can be referred

to different time periods; in particular, in the Italian Civil Protection Department dataset (https://github.com/pcm-dpc/COVID-19), daily incidence counts are available for the following indicators

– positives, which are classified into

  – hospitalized (either in regular wards or in ICU)
  – isolated-at-home

– tested
– deceased
– recovered/discharged

The number of positives available from official reporting, however, underestimates the true number of cases since there exists a vast proportion of asymptomatic or mildly symptomatic patients, among all infected individuals, who are not detected [1–4]. Complex methods are then required to provide reasonable forecasts of the epidemic evolution [5–8] and avoid unreliable predictions, which make people worry (on this point, see e.g. [9, 10]). Thus, the attention has recently moved to the analysis of mortality, as the gold standard measure of the impact of COVID-19 worldwide [11, 12]. This is mainly because mortality data suffer

✉ Antonello Maruotti
  a.maruotti@lumsa.it; antonello.maruotti@uib.no

1 Dipartimento GEPLI, Libera Università Maria Ss Assunta, Rome, Italy

2 Department of Mathematics, University of Bergen, Bergen, Norway

3 Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy

4 Dipartimento di Bioscienze e Territorio, Università del Molise, Pesche, Italy

5 Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università di Palermo, Palermo, Italy

6 Department of Epidemiology and Public Health, Swiss TPH, University of Basel, Basel, Switzerland

7 Unità di Statistica Medica ed Epidemiologia, Campus Biomedico, Rome, Italy

8 Dipartimento di Economia e Finanza, Università di Roma "Tor Vergata", Rome, Italy

less of underreporting, measurement errors and other data issues, i.e. are in general more reliable. We would like to contribute to this growing literature by introducing a generalized linear mixed model to estimate COVID-19-induced excess mortality in Lombardy, Italy.

Italy was dramatically hit by the COVID-19 pandemic, with 4.8 million cases and 132000 deaths as of October 31st, 2021. Lombardy, the most populated and industrialized Italian region, was the epicenter of the outbreak, in particular during the first wave in early 2020. It experienced one of highest case-fatality rates with nearly 900,000 cases (about one fifth of those registered in Italy) and nearly 34,000 deaths out of 10 million inhabitants (for a discussion see e.g. [13]). The impact of COVID-19 on mortality rates is rather clear and can be easily depicted by a simple data visualization. Quantifying the excess deaths induced by COVID-19 is more difficult and may reveal both direct and indirect effects of the pandemic on mortality. Statistical methods have been recently introduced ranging from simple averaging methods [14] to structured models [15–17].

Here we introduce a model to estimate excess mortality in 2020 and 2021 in Lombardy stratified by various age classes. The proposed generalized linear mixed model, in the spirit of [18], is able to capture year-specific trends, seasonality and overdispersion in the number of all-cause deaths and is based on historical data collected on a weekly basis from 2011 to 2019.

Data and methods

Open-source daily all-cause Italian mortality data, stratified by Region and age classes, from 2011 to 2021 are available from the Italian National Institute of Statistics (ISTAT, https://www.istat.it/it/archivio/240401). These data were downloaded on October 25th 2021 and temporally aggregated at the week level. The weeks number 53, present in some years, were dropped from the analysis. Data on historical all-cause mortality are displayed in Fig. 1. Seasonality is clearly present, and is shown for all the considered years. Peaks are rather evident at the major COVID-19 waves, with a maximum of 7729 weekly deaths at the end of March 2020, during the first pandemic wave. Similar behaviors, on different scales, are observed for age-specific data, not reported here for the sake of brevity. We report age-specific plots in the Supplementary Material.

To better appreciate age-class peculiarities, different models are fitted to observed deaths by age classes. Five age classes are considered here, namely (0–14), (15–64), (65–74), (75–84), 85+. This is very useful to detect if and how COVID-19 has an impact on mortality at different age classes of the population and to plan interventions to protect at-risk sub-populations.

The crucial aspect in estimating excess mortality is the definition of a reliable benchmark mortality model, i.e. a model providing *good* estimation of the expected mortality under pre-epidemic conditions. We model

Fig. 1 Observed total number of deaths per week (solid black line) and model's estimated values (thick solid blue line). The dashed (red) line shows the expected weekly deaths under the historical model for years 2020 and 2021
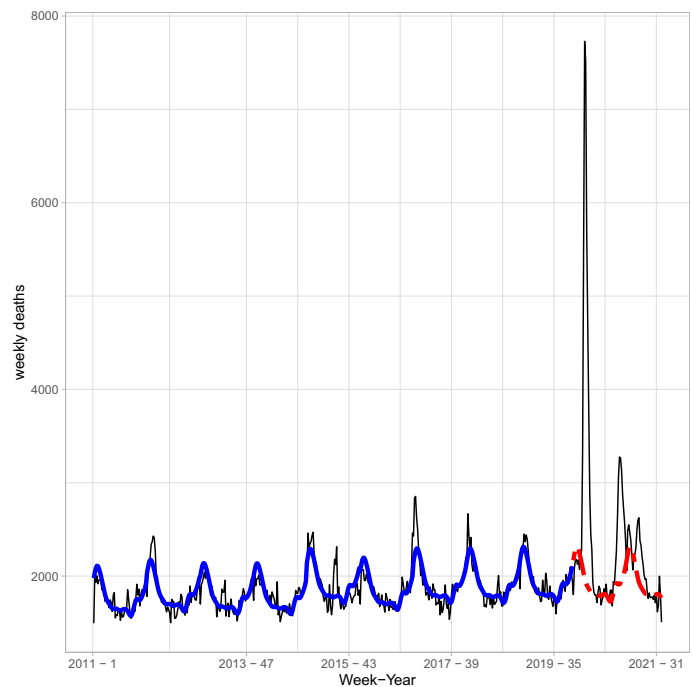
the weekly mortality Y_{tj} at week $t = 1, \ldots, 52$ and years $j = 2011, \ldots, 2019$. The count data process on weekly deaths is modeled with a Negative Binomial distribution, i.e.

$$Y_{tj} \sim NB(\lambda_{tj}, \alpha)$$

with

$$f(y_{tj} \mid \lambda_{tj}, \alpha) = \frac{\Gamma(y_{tj} + \frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})\Gamma(y_{tj} + 1)}$$
$$\left(\frac{\alpha\lambda_{tj}}{1 + \alpha\lambda_{tj}}\right)^{y_{tj}}$$
$$\left(\frac{1}{1 + \alpha\lambda_{tj}}\right)^{1/\alpha}, \quad \lambda_{tj} > 0, \alpha > 0.$$

Count data may be distributed as Negative Binomial if the rate, at which events occur, is heterogeneous, and consequently the counts are characterised by overdispersion compared to the Poisson (as typically happens in our deaths time series). The Poisson distribution is nested within the Negative Binomial, in the sense that if no overdispersion/heterogeneity is present, the Negative Binomial distribution converges to the Poisson distribution.

In count data analysis, the interest is usually focused upon the parameter vector $\lambda = (\lambda_{1,2011}, \ldots, \lambda_{tj}, \ldots, \lambda_{52,2019})$ which is modelled, in a regression context, by defining a generalized linear mixed model with log link for the analysed response, i.e.

$$\log(\lambda_{tj}) = (\beta_0 + b_{0j}) + \sum_{i=1}^{I} \beta_i \sin\left(\frac{2i\pi t}{52}\right)$$
$$+ \sum_{i=1}^{I} \gamma_i \cos\left(\frac{2i\pi t}{52}\right).$$

Cyclical patterns, with-year fluctuations, are captured by Fourier series, whose term I is defined by using model selection criteria; to account for year-specific mortality baselines and to accommodate, at least partially, auto-correlation of the weekly counts, random effects $b_{0j} \sim N(0, \sigma)$ are introduced. Model's fitting to the data is shown in Fig. 1 (see the thick solid blue line). The weekly predictions of mortality data for years 2020 and 2021 are based on the 2019 year-specific conditional best linear unbiased predictions of the generalized linear mixed model. Excess mortality is obtained by subtracting the expected deaths based on the pre-pandemic period from the registered all cause deaths in the pandemic period. Finally, the weekly estimates of excess mortality are summed over the year (or part of it, for 2021).

Results and discussion

Results are shown in Table 1. Prediction intervals are found through parametric bootstrap; if zero is included in the intervals, no difference with the expected number of deaths is estimated, i.e. there is no excess mortality. The first interesting result is that no excess mortality is

Table 1 Estimated excess mortality in Lombardy, with 95% prediction intervals in brackets

Age class	Observed deaths 2020	Excess mortality 2020		
0–14	275	− 33 (− 96; 18)		
15–64	12260	2019 (1575; 2440)		
65–74	18150	5090 (4512; 5639)		
75–84	41275	11568 (10604; 12487)		
85+	63260	15920 (12407; 19226)		
	Observed Deaths Weeks 1 to 34 2020	Excess Mortality Weeks 1 to 34 2020	Observed Deaths Weeks 1 to 34 2021	Excess Mortality Weeks 1 to 34 2021
0–14	182	− 20 (− 61; 13)	146	− 56 (− 96; -24)
15–64	8275	1525 (1232; 1802)	7043	293 (− 1; 572)
65–74	12384	3765 (3384; 4126)	10045	1425 (1047; 1789)
75–84	28253	8594 (7953; 9197)	21315	1655 (1032; 2264)
85+	43177	11864 (9521; 14093)	32004	691 (− 1573; 2966)

estimated in Lombardy neither in 2020 nor in the first 34 weeks of 2021 for the youngest sub-population aged (0–14); indeed, a significant reduction in the number of deaths is estimated for this specific sub-population in 2021. This reinforces the idea that children did not experience excess mortality due to COVID-19 infections. For all the other classes, significant excesses are estimated in 2020, leading to approximately 35,000 more deaths than expected. The largest excess of deaths, in relative terms, is in the classes 65–74 and 75–84, with an excess of approximately 28%. A 25% excess is estimated in the oldest class 85+. For the working age class 15–64, instead, the estimated excess mortality is approximately 18%. There is, thus, a clear age effect on the risk of death induced by COVID-19. As a by-product of the analysis, we are able to detect weeks where the estimated excess mortality is relevant. In Lombardy, a large part of the excess of mortality is estimated during the first wave in March–April 2020, though similar behaviors, with smaller effects, are estimated in Autumn 2020.

According to our analysis, the population 65+ in Lombardy has the highest excess mortality induced by the COVID-19 epidemic. This justifies the priority assigned to the elderly during the vaccination process, as they showed the highest mortality both in absolute and relative terms. As we do not have data for the entire 2021, we focus on the first 34 weeks (up to the end of August), for which we have complete data. Excess mortality is not significantly different from zero for the 85+ and 15–64 age classes, and the estimated excess mortality shows significant reductions with respect to the 2020 estimated values for other age classes. The 65–74 and 75–84 classes are the ones more at risk, with 14.2% and 7.8% of excess mortality estimated; this is somehow not surprising because these sub-populations were the latest to be vaccinated among the elderly (for the relevant data, please, refer to https://github.com/italia/covid19-opendata-vaccini). The positive effect on excess mortality reduction of the vaccine can be better appreciated if we compare excess mortality in the first 34 weeks of 2020 and 2021. A strong reduction in the excess mortality induced by COVID-19 is estimated, even in age classes showing a significant excess. We further discuss results for all age classes except the (0–14) on, which has no excess mortality estimated, focusing on specific weeks where excess mortality arose. In 2020, we estimated an excess mortality on all other classes starting from mid-March to mid-May and, then, from the beginning of November to the end of the year, with peaks in weeks 45–48. Our model does not estimate any excess mortality at the beginning of 2021 for all age classes but the (75–84) one, whose estimated excess mortality on weeks 2 to 4 is statistically significant. The excess mortality in 2021 is mainly concentrated in some specific weeks, from

mid-March to mid-May, with small time-scale differences across (65–74) and (75–84) age classes. In 2021, instead, the vaccine limits the rise of complications possibly leading to death. This result is even more robust as non-pharmaceutical interventions, as lockdowns, are definitely less restrictive during 2021 than during 2020.

Declarations

Conflict of interest We declare that we have no conflict of interest.

Ethical statement This study is compliant with the ethical standards.

Statement of human and animal rights This article does not contain any studies involving human participants performed by any of the authors.

Informed consent Considering the design of the study no informed consent was necessary

References

1. Böhning D, Rocchetti I, Maruotti A, Holling H (2020) Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. Int J Infect Dis 97:197–201
2. Rocchetti I, Böhning D, Holling H, Maruotti A (2020) Estimating the size of undetected cases of the COVID-19 outbreak in Europe: an upper bound estimator. Epidemiologic Methods. https://doi.org/10.1515/em-2020-0024
3. Onder G, Rezza G, Brusaferro S (2020) Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA 323:1775–1776
4. Schneble M, De Nicola G, Kauermann G, Berger U (2021) A statistical model for the dynamics of COVID-19 infections and their case detection ratio in 2020. Biom J. https://doi.org/10.1002/bimj.202100125
5. Alaimo Di Loro P, Divino F, Farcomeni A, Jona-Lasinio G, Lovison G, Maruotti A, Mingione M (2021) Nowcasting COVID-19 incidence indicators during the Italian first outbreak. Stat Med 40:3843–3864
6. Mingione M, Alaimo Di Loro P, Farcomeni A, Divino F, Lovison G, Maruotti A, Jona-Lasinio G (2021) Spatio-temporal modelling of COVID-19 incident cases using Richards' curve: an application to the Italian regions. Spat Stat. https://doi.org/10.1016/j.spasta.2021.100544
7. Agosto A, Campas A, Giudici P, Renda A (2021) Monitoring COVID-19 contagion growth. Stat Med 40:4150–4160
8. Günther F, Bender A, Katz K, Küchenhoff H, Höhle M (2021) Nowcasting the COVID-19 pandemic in Bavaria. Biom J 63:490–502

9. Divino F, Ciccozzi M, Farcomeni A, Jona-Lasinio G, Lovison G, Maruotti A (2021) Unreliable predictions about COVID-19 infections and hospitalizations make people worry: the case of Italy. J Med Virol. https://doi.org/10.1002/jmv.27325

10. Ioannidis JP, Cripps S, Tanner MA (2020) Forecasting for COVID-19 has failed. Int J Forecast. https://doi.org/10.1016/j.ijforecast.2020.08.004

11. Beaney T, Clarke JM, Jain V, Golestaneh AK, Lyons G, Salman D, Majeed A (2020) Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? J R Soc Med 113:329–334

12. Kowall B, Standl F, Oesterling F, Brune B, Brinkmann M, Dudda M, Stang A (2021) Excess mortality due to Covid-19 A comparison of total mortality in 2020 with total mortality in 2016 to 2019 in Germany, Sweden and Spain. PLoS One 16:e0255540

13. Odone A, Delmonte D, Scognamiglio T, Signorelli C (2020) COVID-19 deaths in Lombardy, Italy: data in context. Lancet Public Health 5:e310

14. Modig K, Ahlbom A, Ebeling M (2021) Excess mortality from COVID-19: weekly excess death rates by age and sex for Sweden and its most affected region. Eur J Public Health 31:17–22

15. Dorrucci M, Minelli G, Boros S, Manno V, Prati S, Battaglini M, Bella A (2021) Excess mortality in Italy during the COVID-19 pandemic: assessing the differences between the first and the second wave, year 2020. Front Public Health 9:669209

16. Scortichini M, Schneider dos Santos R, De' Donato F, De Sario M, Michelozzi P, Davoli M, Gasparrini A (2020) Excess mortality during the COVID-19 outbreak in Italy: a two-stage interrupted time-series analysis. Int J Epidemiol 49:1909–1917

17. Blangiardo M, Cameletti M, Pirani M, Corsetti G, Battaglini M, Baio G (2020) Estimating weekly excess mortality at subnational level in Italy during the COVID-19 pandemic. PloS One 15:e0240286

18. Verbeeck J, Faes C, Neyens T, Hens N, Verbeke G, Deboosere P, Molenberghs G (2021) A linear mixed model to estimate COVID-19-induced excess mortality. Biometrics. https://doi.org/10.1111/biom.13578

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.