Original Article

Phases 1 and 2 of Covid-19 Epidemic in the Three Geographical Areas of Italy: An Estimation of Italian Government Measures Based on a Bayesian Changepoint Detection Method

Maria Laura Manca[1,2]*, Francesco Russo[3], Vladimir Georgiev[2], Stefano Taddei[1]

[1]Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy.
[2]Department of Mathematics, University of Pisa, Pisa, Italy.
[3]Medical Hydrology, University of Pisa, Pisa, Italy.

ARTICLE INFO   ABSTRACT

Background: Based on data from the Ministry of Health, which highlighted the earlier onset of Covid-19 epidemic in Italy, compared with the Europe, we would like to present a statistical elaboration on the impact of measures taken by the Government, during the phase 1 and the start of phase 2.
Methods: After the implementation of a Bayesian changepoint detection method, we looked for a best fit model, based on the first part of time series data, in order to observe the progress of the data in the presence and absence of the restriction measures introduced.
Results: Both the implementation of changepoint detection method and the analysis of the curves showed that the decree that marked the start of lockdown has had the effect of slowing down the epidemic by allowing the start of a plateau between 21 and 25 March. Moreover, the decree that decided the beginning of phase 2 on 4 May did not have a negative impact.
Conclusion: This statistical analysis supports the hypothesis that stringent measures decreased hospitalization, thanks to a slowing down in the evolution of the epidemic compared with what was expected.

INTRODUCTION

The novel coronavirus-19 (Covid-19) epidemic has started in China, rapidly spreading across the world and, until today, more than 18.000.000 people got infected, and approximately 700.000 died (Center for Systems Science and Engineering at Johns Hopkins University, https://coronavirus.jhu.edu). Almost all nations were forced to declare lockdown as a measure to manage the dramatic impact of the Covid-19 epidemic, in terms of deaths and load on health systems. Based on official data (http://www.salute.gov.it), which highlighted the earlier onset of the epidemic in Italy, compared with the Europe, we would like to present a statistical analysis on the impact of measures taken by the Italian Government, during the phase 1 and the start of phase 2. In Italy, from 21st of February (date of diagnosis of the first Covid-19 patient), the number of deaths and Intensity Care Unit (ICU) patients was quickly jumping for days. The Italian Government reacted by introducing stringent measures from the date of 9th of March, when a historic decree ordered the lockdown of the entire nation, restricting the people movement, unless for health emergency or unavoidable work needs. Only after data demonstrated that the epidemic may have started to slow, the

---

*. Corresponding Author E-mail: laura.manca@unipi.it

Government slowly begins to mitigate restrictions, since the 4th of May, the date of the first decree of epidemic' phase 2.

Numerous statistical methods are available in order to design, implement, and evaluate public health actions, in epidemiological surveillance systems [1, 2]. In the present work, we focused on methodologies aimed to detect changepoints during the course of infectious diseases, aimed to evaluate health measures. So, the main goals were to identify the effects of the policies in Italy to contain the epidemic at different times (phase 1 and start of phase 2), and, secondly, to compare the possible different impact in the three geographical areas of Italy: Continental, Peninsular and Insular areas.

Materials and Methods

Daily data elaborated in this work, gathered from the Italian Minister of Health (http://www.salute.gov.it) in strict agreement with local Ethical statement, correspond to 80 days, precisely the period that stretches between 1st March and 19th May 2020. We started from the 1st of March because, after first case of Covid-19 was diagnosed in Lombardy region, on that date the epidemic began to spread (1.577 subjects tested positive for the virus, patients hospitalized were 639 of which 140 in ICU, and the deceased were 34), so much so that the Government had adopted the first containment measures in limited geographical zonas. Concerning the date of May 19, we considered that on May 4 the Government began to loosen, albeit prudently, the restrictions, and that both the World Health Organization (https://www.who.int/home), and some Authors [3] accounted for the estimated range of the Covid-19 incubation period between 2 and 14 days. Incidentally, by 19th of May, in Italy 65.129 subjects tested positive for the virus, patients hospitalized were 9.991 of which 716 in ICU, and the deceased were 32.169.

In the present work, we have chosen the hospitalization data because they are more stable and precise than, for example, the number of the confirmed cases that could be biased by a non-stationary testing rate. On time series of hospitalization data, we have implemented a changepoint detection methodology, aimed to investigate the problem of finding abrupt changes in data, following the Governmental interventions [4]. Changepoint problems are used in many application fields, including the epidemiological sector. We have adopted a Bayesian changepoint detection methodology,

originally proposed by Barry and Hartigan (BH) [5], based on Markov Chain Monte Carlo (MCMC) methods. A similar procedure has already been used, for example, by Lee and Ong [6], or Gregori and coworkers [7]. In the present work, BH approach was utilized to detect when the epidemic changed speed as compared with what was to be expected. In the original paper, BH method assumes that all the observations are independent with Gaussian distribution $N(\mu_i, \sigma^2)$, and that the probability of a changepoint at time t_i is p, independently at each t_i. This assumption can be weakened, "because all that is required is that, given the partition and the parameters, observations in different blocks are mutually independent" [5]. Therefore, the prior distribution of μ_{ij} (the mean of the block beginning at time t_i and ending at time t_j) is chosen as $N(\mu_0, \sigma_0^2/(t_j - t_i))$. In our case, we have considered the prior distribution of the mean μ_{ij} of hospitalizations beginning at time t_i and ending at time t_j.

The posterior distribution of the changepoints was obtained, combining the prior distributions and the likelihood derived from the hospitalization data, via MCMC. The algorithm uses a partition $\rho = (U_1, U_2, ..., U_n)$, where *n* is the number of observations (in our case n = 80) and $U_i = 1$ indicates a changepoint at time t_i, and initializes U_i to 0 for all $t_i < t_n$, with $U_n = 1$. In each step of the Markov chain, at each time t_i, a value of U_i is drawn from the conditional distribution of U_i, given the data and the current partition.

Although an exact implementation of BH procedure is possible [5], the calculations are $O(n^3)$, so we have implemented a good MCMC approximation that is $O(n^2)$ [8], using the R - bcp package for Mac Os X.

Then, we looked for a best fit model, based on the first part of time series data, in order to observe the progress of the data in the presence (raw data) and absence (extrapolated data) of the restriction measures introduced. In particular, for each of the 3 geographical areas, starting from the hospitalization data series $y_1, y_2, ..., y_n$ (n = 80), we have extracted the data series $y_1, y_2, ..., y_m$ (m = 25) on which to adapt the model to be used for the forecasts $y_{m+1}, y_{m+2}, ..., y_N$ (N = 40), and obtained a series of estimates for the days from 1 to m, which we first compared with the observed values $y_1, y_2, ..., y_m$ to evaluate the goodness of fit. We have calculated R^2 to evaluate the goodness of fit. At

121

this point we have constructed the set of expected values: F_{m+1}, F_{m+2}, ..., F_N, and compared them with the values observed for the same period y_{m+1}, y_{m+2}, ..., y_N, with the aim of evaluating whether the measures taken by the Government changed the course of the epidemic. We stopped at a 14-day forecast because each model was built from a series of 25 observations.

Results

This analysis detected specific changepoints in the hospitalization curves, depending on the geographical area. In Insular Italy, that encompasses Sardinia and Sicily islands, we have noticed 3 changepoints: 20th of March, 25th of March, 11 and 15 days after the first national decree on the lockdown, respectively, and 11th of May, 7 days after the beginning of the reduction of the restrictions (figure 1, panel 1). In Peninsular Italy, that encloses 10 regions in Central and Southern Italy, 2 changepoints were comparable (16th and 25th of March), while the last was detected the 29th of April (figure 1, panel 2). Regarding Continental Italy, including the 8 regions and 2 autonomous provinces of Northern Italy, where the first two hotspots occurred, we have identified a single changepoint the 21st of March (figure 1, panel 3), close to the first identified in the other two areas of Italy, 12 days after the national decree and 28 days after the local restrictions in the most affected towns.
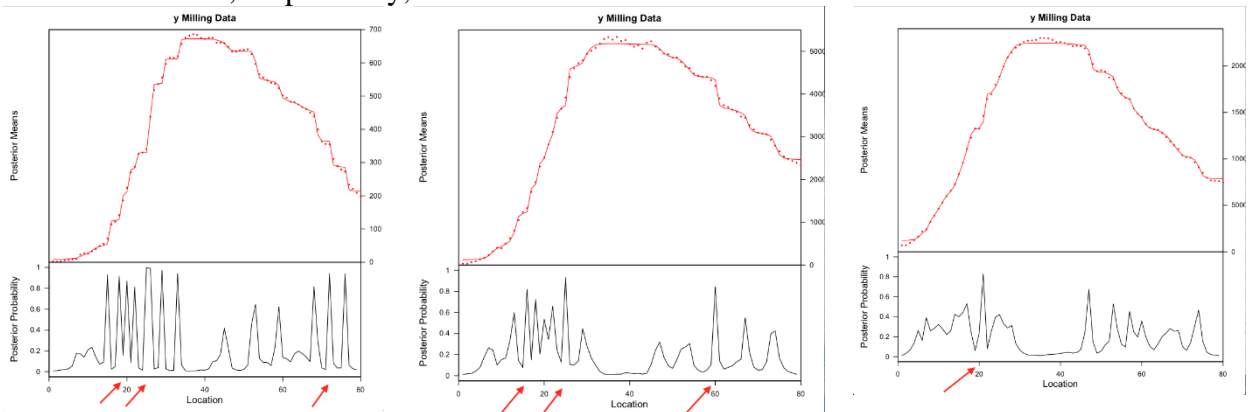

Figure 1. Results of changepoint analysis in the 3 geographical areas, Panel 1 represents the results in Insular Italy, panel 2 in Peninsular Italy, and panel 3 in Continental Italy. Red raw indicates the changepoints.

Considering that on March 25th, all three geographic areas showed a changepoint, we modeled the daily data from March 1st to 25th, extrapolating the values for the following 14 days from the model (figure 2, panel 1, 2, 3). The best fit of the epidemic curves were constituted by a 2-degree polynomial in the first three weeks of virus diffusion, in all areas (mean and standard deviation: $R^2 = 0.992 \pm 0.007$). When we represented the real data, from March 26 to May 19, the significant impact of containment measures compared to prediction was evident in all areas of the country (figure 2, panel 1, 2, 3).

Both from the implementation of changepoint detection method and the analysis of the curves, in all areas, appeared that the decree that marked the start of lockdown has had the effect of slowing down the epidemic, by allowing the start of a plateau between 21 and 25 March. Moreover, the decree that marked the start of phase 2 of Covid-19, on 4 May, did not have a negative impact, anywhere in Italy and especially in Insular area.
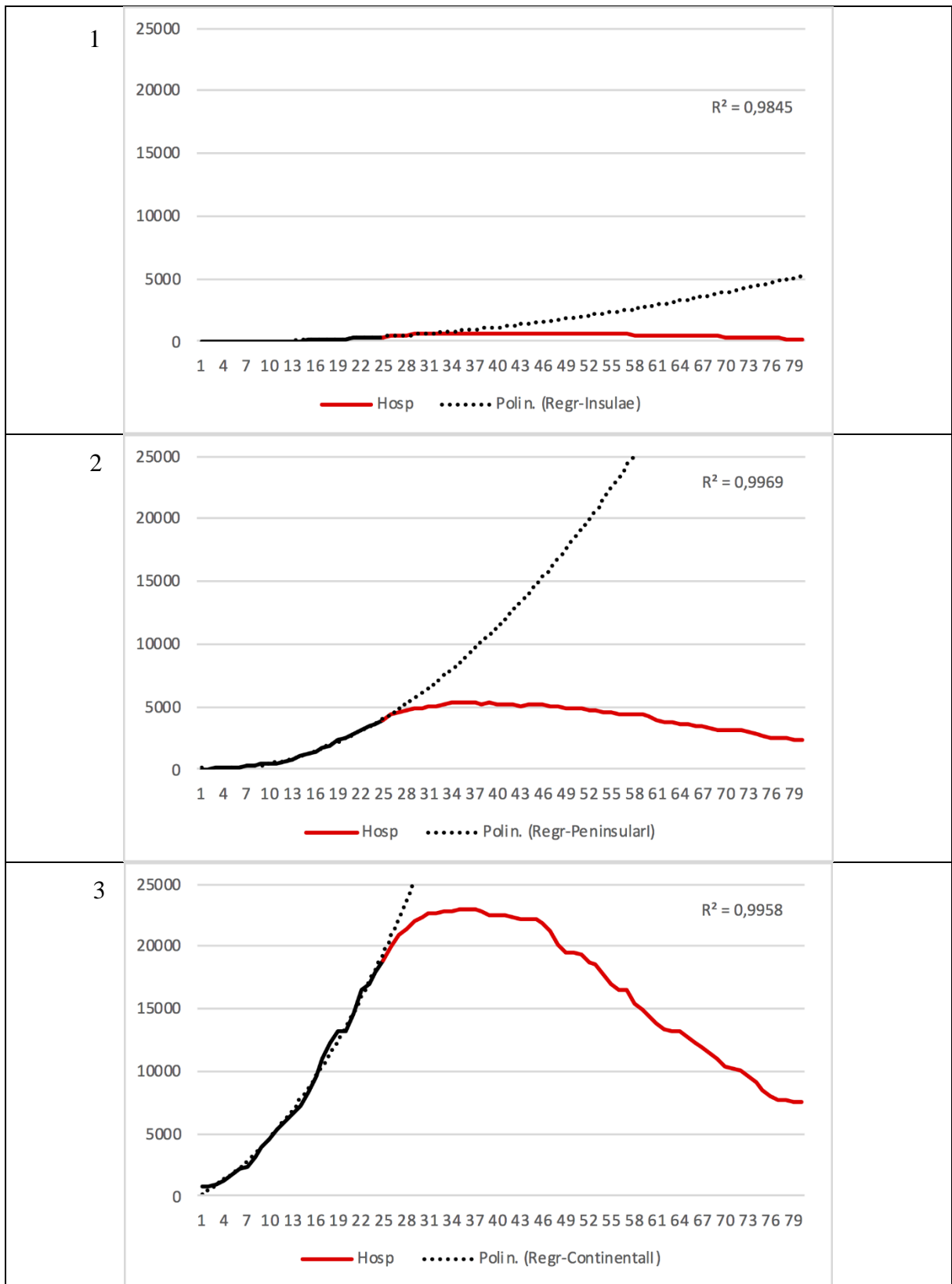
Figure 2. Time series of observed and predicted data in the 3 geographical areas, Panel 1 represents the results in Insular Italy, panel 2 in Peninsular Italy, and panel 3 in Continental Italy. On horizontal axis there is the time in days, and on vertical axis the number of hospitalized patients. Red line represents observed data; dotted black line represents estimated data without containment measures.

Discussion and Conclusions

The present manuscript, based on a combined method of Bayesian changepoint detection and 2-degree polynomial models implemented on the hospitalization case time series, extended to Italy the research of Gregori [7] on the Veneto region. Results showed that containment measures were effective in all 3 geographical areas in which Italy is divisible, in agreement with Chitalapudi [9], who highlighted the relevance of lockdown and self-isolation in control the disease transmissibility among Italian population, using a different approach implemented with the seasonal R-ARIMA forecasting package.

Results also demonstrated that the Covid-19 epidemic was putting overwhelming pressure mainly on Continental Italy, while Peninsular and Insular Italy seemed to be less affected, according to Distante [10], who documented different progression of the epidemic in Southern Italian regions compared to Lombardy. In Continental Italy, the algorithm we implemented revealed a single changepoint on March 21, suggesting that the slowdown was proceeding slowly. In the other 2 areas, we identified 2 changepoints in March, a few days later, and a further changepoint which underlined a significant new reduction in the circulation of the virus.

Given the initial spread of Covid-19 in Northern Italy, the present results, showing greater efficacy in the least affected areas at the time of closure, support the recommendation of both WHO and some Authors [11] that strict containment measures should be introduced as early as possible. Similarly, it is desirable that the slowdown in measures is prudential. The first decree of phase 2 of the Italian government has kept the borders between the regions closed, and results have not revealed a negative impact of the first openings in any area of Italy.

Looking at figures 1 and 2, it can be seen how the beginning of the long plateau, that preceded the reduction of hospitalizations and infections in Italy, occurred between 21 and 25 March. The SIRD model constructed by Fanelli and Piazza [12] also culminated in Italy around March 21. Similarly, Sebastiani and coworkers [10] discovered that, by March 31, the epidemic growth rate has already peaked in 87 out of 107 Italian provinces. The same Authors estimated that the delay between the start of the implementation of the restriction measures and the measurable reduction in the growth rate of Covid-19 was about 7-10 days, in according to the present delay that ranged between 7 and 12 days.

Finally, the best fit of the epidemic curves was represented by a 2-degree polynomial in the first three weeks of diffusion, in all areas, in line with previous studies [7], indicating that COVID-19 does not exhibit exponential growth, not even in Northern Italy.

In conclusion, this study supports the hypothesis that stringent measures decreased hospitalization (and more indirectly the number of deaths, positive people and ICU accesses), thanks to a slowing down in the evolution of the epidemic compared with what was expected. Furthermore, the combined use of changepoint detection methods and time series modeling elaboration seems to provide a useful tool for analyzing Covid-19 epidemic. Future research may include a comparison between the results obtained with the present procedure and those obtained with SIR models.

What is already known on this subject

Italy has been the first nation in Europe where COVID-19 has been spreading, and the Italian Government is the first in Europe to adopt severe containment measures, and one of the most prudent in alleviating them.

What this study adds

The containment strategies would indicate an effective impact on the management of the outbreak throughout all Italy, especially in the Peninsular and Insular areas where the epidemic started less early. Furthermore, the combined use of changepoint detection methods and time series modeling elaboration seems to provide a useful tool for analyzing Covid-19 epidemic.

Conflict of interest

Acknowledgements

References

1. Shmueli G, Burkom H. Statistical challenges facing early outbreak detection. Technometrics 2010; 52: 39.

2. Unkel S, Paddy Farrington C, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society. Series A (Statistics in Society) 2012; 175: 49-82.

3. Lei S, Jiang F, Su W. Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection. Clin Med. 2020; 100331.

4. Kawahara Y, Sugiyama M. Sequential Change-Point Detection Based on Direct Density-Ratio Estimation. SIAM International Conference on Data Mining. 2009: 389–400.

5. Barry D, Hartigan JA. A Bayesian analysis for change point problems. J Am Stat Assoc. 1993; 88: 309.

6. Lee CN, Ong HC. Bayesian analysis of change point problems for time series data. AIP Conference Proceedings 1750, 060025 2016; https://doi.org/10.1063/1.4954630.

7. Gregori D, Azzolina D, Lanera C, et al. A first estimation of the impact of public health actions against COVID-19 in Veneto (Italy). J Epidemiol Community Health 2020; jech-2020-214209.

8. Erdman C, Emerson JW. A fast Bayesian change point analysis for the segmentation of microarray data. Bioinformatics 2008; 24: 2143-2148.

9. Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. J Microbiol Immunol Infect. 2020; 53: 396-403.

10. Distante C, Piscitelli P, Miani A. Covid-19 Outbreak Progression in Italian Regions: Approaching the Peak by the End of March in Northern Italy and First Week of April in Southern Italy. Int J Environ Res Public Health 2020; 17: 3025.

11. Sebastiani G, Massa M, Riboli E. Covid-19 epidemic in Italy: evolution, projections and impact of government measures. Eur J Epidemiol. 2020; 35: 341-345.

12. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 2020; 134: 109761.