

Covid-19 in Italy: Modelling, communications, and collaborations

When Covid-19 arrived in Italy in early 2020, a group of statisticians came together to provide tools to make sense of the unfolding epidemic and to counter misleading media narratives. Here, members of **StatGroup-19** reflect on their work to date

The scientific community's ongoing battle against Covid-19 has led to several major advances. A first vaccine was made available in roughly one year, for example – a remarkable achievement that would have seemed unlikely, if not impossible, at the beginning of the pandemic. At the same time, new statistical and epidemiological tools have been developed to understand the dynamics of the virus's spread, predict its evolution, and evaluate the impact of economic, political, clinical, and public health interventions.

Such progress, however, has not come without setbacks. From the statistical and epidemiological perspective, the need for accurate and timely information on the virus brought to light the inadequacy of national and international structures – government departments, public health agencies, and non-governmental organisations – to collect reliable data at speed and in the appropriate form. In Italy, for example, each area of the country could use a different data collection system, which too often relied almost entirely on humans who were not adequately trained in the task. Delays in data reporting and measurement error were common, so much so that in some cases restrictions were imposed on certain regions because of poor data quality (i.e.,

the inability to monitor the occurrence of outbreaks).

As in Italy, in many countries publicly available data were scarce and often proved unreliable, especially during the first year of the pandemic. At the same time, access to the (limited) sources of more accurate information (e.g., data about clinical history and the contact networks of individual patients) was restricted to very few research teams, thus limiting the potential of these data to inform our understanding of the dynamics of the pandemic and to guide public decisions.

These and other data-related issues are not unique to the Covid-19 situation, but they have mostly been of minor concern to the general public before now. With the arrival of the pandemic, awareness of the need for high-quality data and the role of such data in guiding decision-making has increased, and their importance for policy-making and day-to-day management has finally been properly acknowledged.

Monitoring the pandemic

Our research group has been involved in modelling Covid-19 data since the beginning of April 2020. Motivated by media coverage of some poorly conducted and far-fetched forecasts by well-intentioned but otherwise

ill-equipped analysts, we wanted to add a statistically sound voice to the choir.

We decided to rely only on publicly available data sources, in the spirit of transparent and reproducible science. Our goal was to provide the general public with reliable tools for the interpretation and short-term prediction of the most relevant epidemic indicators. We built an interactive web application (statgroup19.shinyapps.io/Covid19App) where one could obtain daily updated descriptive analyses and get access to forecasts based on our models. In particular, we wanted to contribute to developing the risk literacy of the population – we sought to help the user in distinguishing relevant information from harmful and dangerous misinterpretations. That is why the app also includes essential summaries of the epidemic indicators, interactive graphs, and maps. (Source code is available at github.com/minmar94/StatGroup19.)

One of our first modelling attempts was a parametric regression model to fit the numbers of Covid cases and deaths (incidence data) based on the Richards curve.¹ Richards curves form a family of flexible logistic curves that have been widely used in different fields to model various growth phenomena. They can naturally adapt to the typical phases of epidemic

Modelling with the Richards curve

Our reparameterised version of the Richards curve depends on five parameters, and can capture the different growth trends characterising an epidemic wave: the endemic rate, the maximum size, the growth rate, the peak position, and the asymmetry between the ascending and descending phase of the outbreak. Our simple idea was to express the typically constant baseline parameter through a linear temporal trend (envisioning an endemic state for the disease), and then specify a nonlinear generalised regression model.

The Richards curve can be used to model the expected value of a distributions like the Poisson and negative binomial, which are the most suitable for the parametric modelling of count (e.g., incidence) data. Many researchers have been working on the same logistic growth idea, but relying on a Gaussian specification for the log counts.^{6,7} The latter assumption, however, is in clear contrast to the nature of the data and can invalidate inferential conclusions.

► waves – that is, an endemic, baseline level of infections that seems to build slowly at first, then rapidly, before reaching its peak, after which case numbers start to decrease, falling back over time to an endemic level (see box, “Modelling with the Richards curve”). Our predictions proved to be reliable both for monitoring the situation in real time (a technique known as “nowcasting”) and for describing trends retrospectively (see Figure 1(a) for an example). In addition, the ability of our model to anticipate the day with the maximum number of incident cases (e.g., the day of the peak) – a topic of much interest in the Italian media, especially during the first outbreak – led to a scientific paper.²

This work had some limitations. The need for immediate forecasts and a fast-fitting procedure made us rely on working assumptions, such as stochastic

independence between daily records at different times and in different regions. What this means is that, loosely speaking, the model assumed that (conditionally on the trend) cases in one region on any given day were independent of case numbers at previous times or in neighbouring regions. This is obviously not the case in reality. We dealt with this misspecification using a robust estimation procedure, but we were aware that improved performance could be attained by embedding spatio-temporal (space and time) dependence in the model. A more recent published paper expands our work in this direction.³ Interestingly enough, the results show that substantial spatial and temporal dependence occurred in both the first two Italian epidemic waves, even with a lockdown in place during the first wave. In particular, during the first outbreak,

spatial dependence was mainly driven by geographical proximity (as shown in Figure 1(b)), whereas in the second wave it was driven by transportation (e.g., high-speed trains, flights, and ferries). Furthermore, the model yielded accurate predictions at both the regional and national levels.

The incidences of cases and deaths are only two of the possible indicators of the momentum of an epidemic, describing the pace at which the disease is spreading. Prevalence indicators instead describe the proportion of the population currently infected by the virus or at different stages of disease progression. In particular, the current occupancy of intensive care units (ICUs) is a key metric for the proper allocation of health resources. The aggressiveness of SARS-CoV-2, especially in older populations, put a strain on public health systems worldwide, particularly at the beginning of the pandemic. Emergency rooms and ICUs were overwhelmed by the need for a large number of beds and the lack of effective procedures to fight the most severe symptoms of Covid-19. Therefore, in Italy and elsewhere, ICU occupancy became a crucial indicator to monitor in order to anticipate the possible overloading of hospitals.

In response to this, we devised a model for the short-term prediction of ICU occupancy.⁴ This is based on an ensemble of two different models that look at the same data from different perspectives, combining the best of both worlds. Average occupancy is expressed through polynomial trends in time, and the model fit exploits only records from the most recent two weeks in order to avoid unwanted effects stemming from the global fit on counts too far back in time. Our forecasts proved to be extremely accurate from 1 to 5 days ahead, with observed values within 99% prediction intervals at the nominal rate (see Figure 2) and a median prediction error of two hospital beds. Extrapolation over the proposed horizons has not been tested, and is not advised: the mathematical expression of the trends may yield unreasonable behaviours as observed data get further out in time.

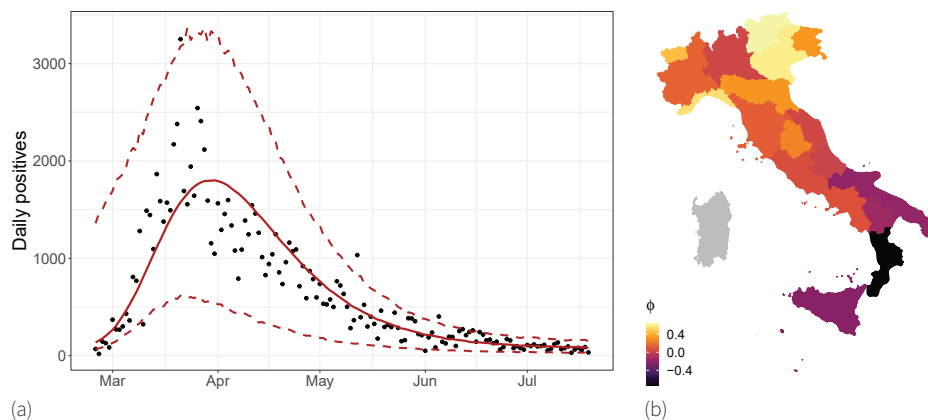


Figure 1: (a) Observed (black dots) and fitted values (red solid line) with 95% confidence intervals (red dashed lines) for the daily number of positive cases of Covid-19 in the Lombardy region of Italy, 5 February 2020 to 19 July 2020. (b) Average estimated spatial random effect for Italy’s first wave.

Predictions proved reliable both for “nowcasting” and for describing trends

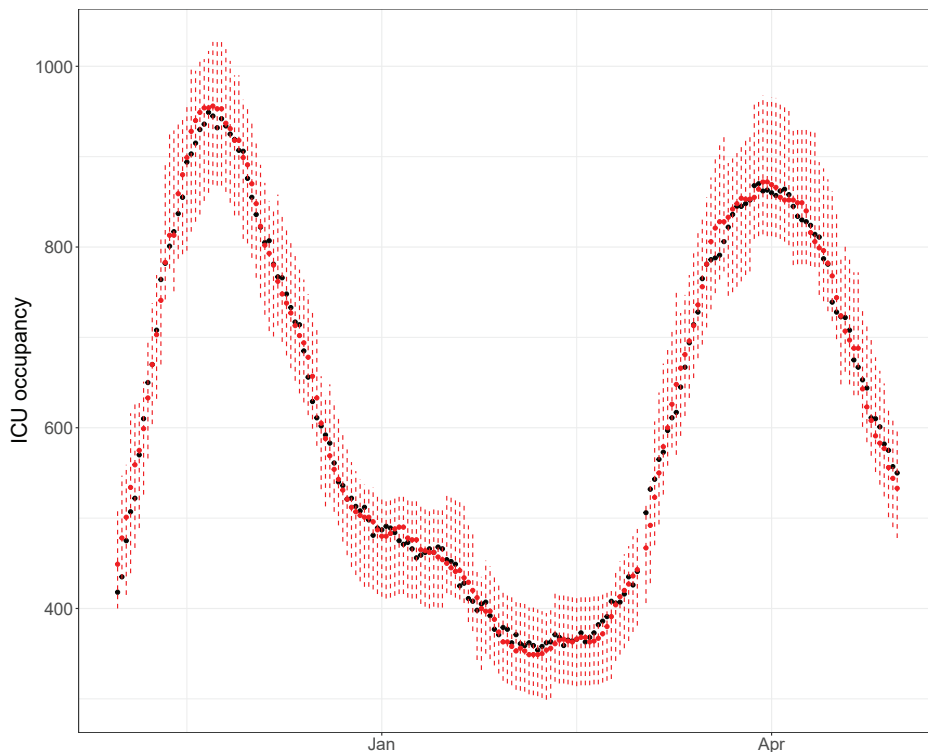


Figure 2: Observed (black dots) and predicted values (red dots) with 95% confidence intervals (red dashed lines) for ICU occupancy in Lombardy during the second and third wave of Italy's Covid-19 epidemic, 1 November 2020 to 30 April 2021.

Hopes and lessons

Alongside our research, our group devoted much energy to the correct communication of Covid-19 epidemic data and statistical analyses. The frenzy and complexity of the situation, especially in the early months of the pandemic, and the overwhelming amount of data, models, and forecasts, inevitably led to some questionable reporting – with stories designed more to grab attention than to provide reliable and complete information.

In an effort to remedy this, we established contacts with several data journalists. We became (and are) routinely involved in illustrating the evolution of the pandemic in Italy for SkyTG24's *I numeri della pandemia*, a TV news programme that airs every day in the late afternoon, shortly after the latest Covid-19 data become available. We also created social media pages and a blog, and we continue to do our best to connect with people and to help them in the correct reading of the pandemic figures. We have also contributed

to the scientific discussion on appropriate communication of statistical results.⁵

The Covid-19 pandemic has brought into sharp focus the need for a better relationship between media and science, with journalists and scientists constantly collaborating and overcoming the urge to simply make headlines. More than that, this global crisis shows that we need improved mechanisms for data collection at national and supra-national levels, and that the collected data must be accurate, consistent, coherent, and timely. They also must be shared freely with the scientific community.

Within the statistical community, there should be a push for production of, and access to, better data; the use of rigorous statistical methods for analysing those data; and improvements in the communication of results to decision-makers, media professionals, and the public.

Ultimately, we hope that the experiences of the past two years will lead political and individual decisions to be driven by attention

towards the health, safety, and well-being of every human being, leaving aside the national and cultural differences that too often limit the development of society and scientific progress. ■

Note

StatGroup-19 is an independent research group formed by Pierfrancesco Alaimo Di Loro (Sapienza – University of Rome), Fabio Divino (University of Molise), Alessio Farcomeni (University of Rome “Tor Vergata”), Giovanna Jona Lasinio (Sapienza – University of Rome), Gianfranco Lovison (University of Palermo and Swiss Tropical and Public Health Institute, Basel), Antonello Maruotti (Libera Università Maria Ss. Assunta and University of Bergen), and Marco Mingione (Sapienza – University of Rome).

Disclosure statement

The authors declare no competing interests.

References

1. Richards, F. (1959) A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**(2), 290–301.
2. Alaimo Di Loro, P., Divino, F., Farcomeni, A. *et al.* (2021) Nowcasting COVID-19 incidence indicators during the Italian first outbreak. *Statistics in Medicine*, **40**(16), 3843–3864.
3. Mingione, M., Alaimo Di Loro, P., Farcomeni, A. *et al.* (2021) Spatio-temporal modelling of COVID-19 incident cases using Richards' curve: An application to the Italian regions. *Spatial Statistics*. <https://doi.org/10.1016/j.spasta.2021.100544>
4. Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G. and Lovison, G. (2021) An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical Journal*, **63**(3), 503–513.
5. Divino, F., Ciccozzi, M., Farcomeni, A., Jona-Lasinio, G., Lovison, G. and Maruotti, A. (2021) Unreliable predictions about COVID-19 infections and hospitalizations make people worry: The case of Italy. *Journal of Medical Virology*, **94**(1), 26–28.
6. Grasselli, G., Pesenti, A. and Cecconi, M. (2020) Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. *Journal of the American Medical Association*, **323**, 1545–1546.
7. Sebastiani, G., Massa, M. and Riboli, E. (2020) COVID-19 epidemic in Italy: Evolution, projections and impact of government measures. *European Journal of Epidemiology*, **35**, 341–345.