

Analysis of the SARS-Cov-2 epidemic in Lombardy (Italy) in its early phase. Are we going in the right direction?

Angelo Riccio*¹

¹Dept. of Science and Technology, University of Naples Parthenope,
80143–Naples (Italy)

Abstract

BACKGROUND

We described the epidemiological features of the covid-19 outbreak, and evaluated the impact of interventions measures on the epidemic in the Lombardy region, Italy.

METHODS

Laboratory-confirmed covid-19 cases reported through the beginning of April were extracted from the Italian Civil Protection database. Based on key events and interventions, we divided the epidemic into three periods: before February 21, from February 22 to early March, after early March. We compared epidemiological characteristics across periods and

*e-mail: angelo.riccio@uniparthenope.it

developed a modified susceptible-exposed-infectious-recovered model to study the epidemic and evaluate the impact of interventions. We explicitly took into account for unascertained cases (positive cases with no symptoms or mild symptoms that have not been accounted for in official statistics).

RESULTS

Currently, the number of positive active cases has increased to around 30,000 in the Lombardy region. Due to restriction measures, the effective reproduction number dropped from 3.33 (95% CI: 2.03–3.69) during the first period, to 2.36 (95% CI: 2.21–2.70) during the second period. In the third period, the effective reproduction number is estimated to have dropped to 1.49 (95% CI: 1.35–1.62). The model estimates a great proportion of unascertained cases, about 90% of infected people has not been accounted for in official statistics.

CONCLUSIONS

Considerable countermeasures have slowed down the covid-19 outbreak in the Lombardy region. However, notwithstanding the long-lasting lockdown period, the epidemic is still not under control. The effective reproduction number, according to the model used in this work, is still greater than 1.0. Estimation of unascertained cases has important implications on continuing surveillance and interventions.

1 Introduction

- 2 The novel coronavirus (SARS-CoV-2) emerged in Wuhan, China, in late 2019, and quickly spread rapidly to all Chinese provinces and other Asian countries

4 (NHC, 2020). On March 11, 2020, the WHO (World Health Organization) de-
clared the covid-19 outbreak pandemic, after the disease caused by the new
6 coronavirus infected more than 100,000 people and spread to more than 100
countries (WHO, 2020).

8 In the late night of February 20, 2020, the first case of novel coronavirus
disease was confirmed in the Lombardy Region, northern Italy, around the city
10 of Codogno. In the following week, the Codogno area, as well as several neigh-
boring towns in southern Lombardy, experienced a very rapid increase in the
12 number of detected cases, which rose to confirmed 172 positive samples by
February 24 and 5,791 by March 10. From the early beginning of the outbreak,
14 the Italian authorities have adopted a series of restrictive containment mea-
sures, including the creation of a ‘red zone’ around the city of Codogno and
16 other small towns, progressively extended to several regions of northern Italy,
and then, from March 10, also throughout Italy.

18 Efforts to contain the virus are ongoing; however, given the many uncer-
tainties related to the transmissibility and virulence of this pathogen, the ef-
fectiveness of these efforts is unknown. Several recent studies have reported
20 a nonnegligible proportion of asymptomatic cases and transmissibility of the
asymptomatic or presymptomatic cases (Mizumoto et al., 2020; Li et al., 2020;
22 Hu et al., 2020). Moreover, epidemiological data (dates of symptoms onset,
24 clinical features, respiratory tract specimen results, hospitalization and con-
tact tracing) collected by the Local Health Authority (ATS: Agenzia di Tutela
26 della Salute) have allowed to establish that the epidemic in Italy began long
before the clinical evidence of the first case ascertained in Codogno, when the
28 virus had already spread to most of the municipalities of southern Lombardy

(Cereda et al., 2020).

30 Here, to study the epidemic trend of this disease, we use a model inference
to estimate the proportion of undocumented infections in the Lombardy re-
32 gion during the early phase of the epidemic, taking into account for interven-
tion measures, ascertainment rate, transmission rate, and duration from ill-
34 ness onset to hospitalization. Model predictions have been compared against
actual reported cases to evaluate the overall impact of interventions.

36 Using the data available in Italy, our objectives are to estimate:

- the number of people infected with covid-19 in the Lombardy region;
- 38 - the parameters of a SEIR-type model representing the early phase of the
outbreak;
- 40 - the temporal modulation of the effective reproduction due to interven-
tion measures.

42 **2 Methods**

2.1 Source of data

44 In Italy positive cases are first confirmed and validated in the laboratory and
then distributed to the public through the Italian Civil Protection Department
46 website (<https://github.com/pcm-dpc/COVID-19>). Positive cases
are ascertained if the patient had a positive test of SARS-CoV-2 virus by the
48 reverse transcription-polymerase chain reaction (RT-PCR) assay from the se-
quencing of nasal and pharyngeal swab specimens.

50 **2.2 Classification of time periods**

To better reflect the dynamics of the covid-19 epidemic and corresponding in-
52 terventions, we classified the outbreak into three periods based on important
dates that could affect the virus transmission:

- 54 1. the time before February 20;
2. the interval between February 20 and early March,
- 56 3. and the time after early March.

The first date corresponds the date of the first confirmed case in Italy, and this
58 period was considered because no intervention was imposed before then. Dur-
ing the second period a series of progressively more stringent restriction mea-
60 sures have been adopted to contain the infection, precisely:

- 62 • February 24: closing of schools in Lombardy, Emilia Romagna and Vene-
to regions;
- March 1: partial restriction of productive activities throughout Lombar-
64 dy;
- March 5: schools closure throughout Italy;
- 66 • March 8: lockdown of mobility in Lombardy.

After March 10, a severe lockdown of almost all activities has been extended to
68 the whole of Italy.

2.3 Statistical analysis

70 Due to the severe shortage of medical resources early on in this epidemic, many
suspected cases were unable to receive timely treatment and were self-quar-
72 antined at home. It is also believed that many deaths related to the covid-19
epidemic have not been correctly accounted for in official statistics, so that the
74 real extent of the epidemic has been probably underestimated in Lombardy.

To infer SARS-CoV-2 transmission dynamics during the early stage of the
76 outbreak, we simulated observations starting from February 24, 2020 (i.e., when
data from the Italian Civil Protection were made available), extending the clas-
78 sic SEIR model to account for quarantine measures and unascertained cases.
We divided the population into six compartments including susceptible indi-
80 viduals, exposed cases, ascertained cases, unascertained cases, hospitalized
cases, and recovered individuals. Here, unascertained cases included asymp-
82 tomatic cases and those with mild symptoms that could recover without seek-
ing medical care and thus were not reported to the authorities. The model dy-
84 namic is illustrated in Figure 1.

The key parameters in the model are:

- 86 1. ρ (effect of intervention measures),
2. r (ascertainment rate), and
- 88 3. β (transmission rate).

For unreported infectious individuals, the transmission rate may be different
90 from that of infectious; this is taken into account by the parameter δ .

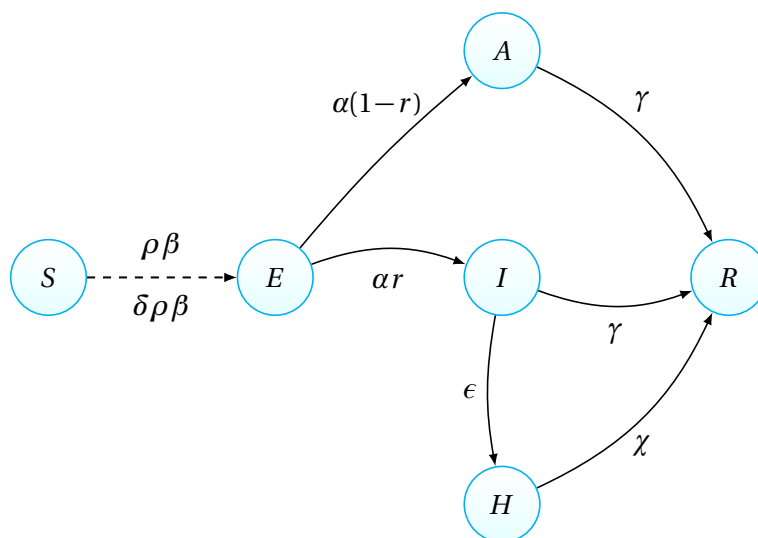


Figure 1: Illustration of the extended SEIR model. The symbols for compartments are: S (susceptible), E (exposed), I (reported infectious), A (unreported infectious), H (hospitalized), and R (recovered). The rate associated with each path is indicated by the corresponding basic parameter.

Considering the impacts of major interventions, we allowed the reproduction number to vary over time, rather than simply fix this value, to capture possible variation in transmission as a result of control measures and behaviour change. We assumed that the ascertainment rate and transmission rate were different for the three time periods. The effective reproduction number, R_t , defined as the expected number of secondary cases infected by a primary case, was computed for each period. Initial states and parameter settings of the SEIR model are described in detail in the Appendix A.

3 Results

100 The epidemic curve is illustrated in Figure 2. During the first phase, the in-
fection curve grew rapidly and, despite containment measures, no significant
102 slowdown was observed. For this reason, on 8 March it was decided the lock-
down of all activities. Only after several days, the infection curve began to show
104 a slowdown. Currently, on 10 April, the number of people infected in the Lom-
bardy region is about 30,000.

106 Our SEIR model fit the observed data reasonably well (Figure 3). To adapt to
the observed data, the model estimates that it is possible to date the beginning
108 of the epidemic to the first decade of January, precisely to 10 January (95% CI:
5–14 January). In addition, the effectiveness of lockdown, namely the effects of
110 the progressive restriction measures, started to show their effects on 17 March
(95% CI: 14–19 March).

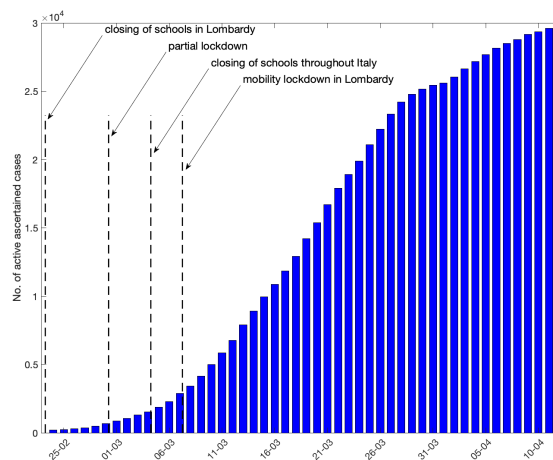


Figure 2: The daily number of ascertained cases. The vertical black dashed lines are drawn at the time when the different control measures were taken.

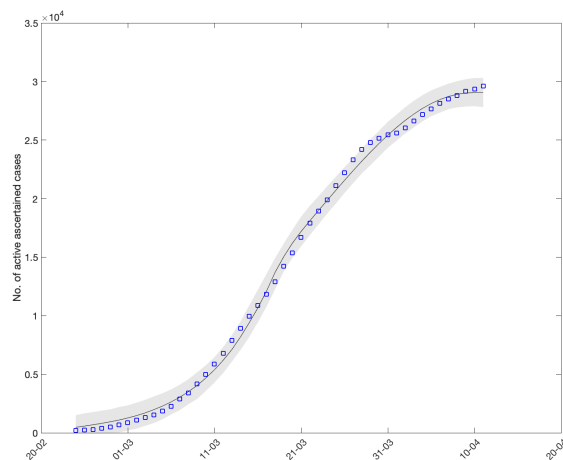


Figure 3: Fit of the SEIR model to the number of active ascertained cases in Lombardy. The grey shaded area represents the 95% CIs.

112 Strikingly, we estimated that the overall ascertainment rate was 0.13, and
similar across the last two periods. Precisely, during the second period the me-
114 dian value of the ascertainment rate, r_1 , is 0.13 (95% CI: 0.10–0.17) and during
the second period the ascertainment rate, r_2 , is 0.13 (95% CI: 0.11–0.17), that
116 is, about 90% of infected people has not been included in official statistics, ac-
cording to our SEIR model. During the first days of April 2020, we estimate
118 about 250,000 unascertained cases in Lombardy, i.e. about 2.5% of the whole
population in this region, see Figure 4. Only 30,000 cases have been accounted
120 for in official statistics.

The estimated effective reproduction number, R_t , is shown in Figure 5. Dur-
122 ing the first period, i.e. before February 24, when no control measure where
taken, the reproduction number is estimated to be 3.33 (95% CI: 2.03–3.69),
124 and slowed down to 2.36 (95% CI: 2.21–2.70) during the second period. Un-

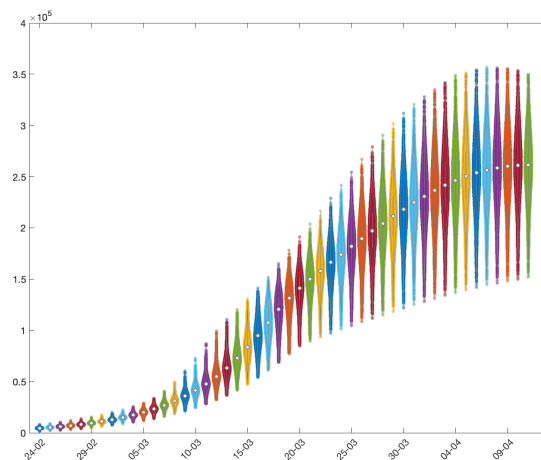


Figure 4: Number of unascertained cases in Lombardy.

fortunately, it should be noted that the curve of active cases is slowing down
126 slowly. This feature translates into an R_t value which is still relatively high. In
the third period it is estimated that R_t is greater than 1, precisely 1.49 (IC 95%:
128 1.35-1.62), despite the length of the lockdown that, on April 10, lasts for more
than a month.

130 As more data become available on the timing of control measures and sub-
sequent dynamics, we agree that it will be crucial to evaluate the effectiveness
132 of measures to provide a robust evidence base for future policy-making.

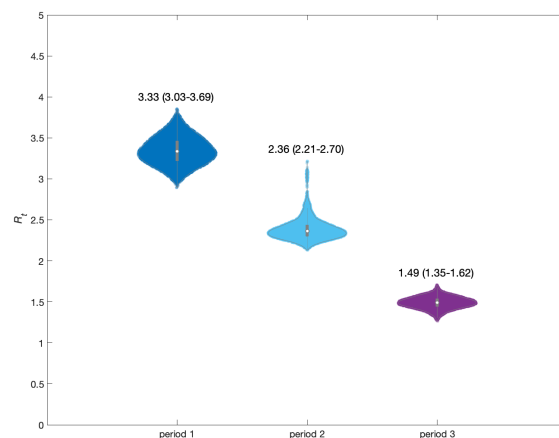


Figure 5: Estimated R_t for the three periods. For each time period, the median and the 95% CI are reported above the violin plot.

References

- 134 D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini,
G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno,
136 M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti,
M. Gramegna, F. Baldanti, A. Melegaro, and S. Merler. The early phase of the
138 COVID-19 outbreak in Lombardy, Italy, 2020.
- H. Haario, M. Laine, A. Mira, and E. Saksman. DRAM: efficient adaptive MCMC.
140 *Statistics and computing*, 16(4):339–354, 2006.
- J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D.
142 Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al. Feasibility of controlling
COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global
144 Health*, 2020.

Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng,
146 et al. Clinical characteristics of 24 asymptomatic infections with COVID-19
screened among close contacts in Nanjing, China. *Science China Life Sci-*
148 *ences*, pages 1–6, 2020.

R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. Substantial
150 undocumented infection facilitates the rapid dissemination of novel coron-
avirus (SARS-CoV2). *Science*, 2020.

152 K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell. Estimating the asymp-
tomatic ratio of 2019 novel coronavirus onboard the princess cruises ship,
154 2020. *medRxiv*, 2020.

NHC. National Health Commission of the People’s Republic of
156 China, Update on the novel coronavirus pneumonia outbreak, 2020.
http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml, last accessed on 8 Febru-
158 ary 2020.

C. Wang, L. Liu, X. Hao, H. Guo, Q. Wang, J. Huang, N. He, H. Yu, X. Lin,
160 A. Pan, et al. Evolving epidemiology and impact of non-pharmaceutical in-
terventions on the outbreak of coronavirus disease 2019 in Wuhan, China.
162 *medRxiv*, 2020.

WHO. WHO Director-General’s opening remarks at the
164 media briefing on COVID-19 - 11 March 2020, 2020.
[https://www.who.int/dg/speeches/detail/who-director-general-s-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020)
166 [opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020.](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020)

A Model description

168 We extended the classic SEIR model to account for unascertained cases and
quarantine measures (Figure 1). We divided the population into S (suscepti-
170 ble), E (exposed), I (ascertained infectious), A (unascertained infectious), H
(hospitalized) and R (recovered) individuals.

172 Starting with a total population size of 10.06 million (equivalent to that of
the Lombardy region) and 1 exposed individual, the model update the number
174 of each compartment, according to the following set of ordinary differential
equations:

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -\rho\beta \frac{SI}{N} - \delta\rho\beta \frac{SA}{N} \\ \frac{dE}{dt} = \rho\beta \frac{SI}{N} + \delta\rho\beta \frac{SA}{N} - \alpha E \\ \frac{dI}{dt} = r\alpha E - \gamma I - \varepsilon I \\ \frac{dA}{dt} = (1-r)\alpha E - \gamma A \\ \frac{dH}{dt} = \varepsilon I - \chi H \\ \frac{dR}{dt} = \gamma I + \gamma A + \chi H \end{array} \right. \quad (1)$$

176 The initial state of the model is reported in Table 1.

β is the basic reproduction number, i.e. the transmission rate without any
178 intervention measure, defined as the number of individuals that a positive case
can infect per day; ρ weights the effect of intervention measures; δ is the ra-
180 tio of the transmission rate of unascertained over ascertained cases; r the as-
certainment rate; $1/\alpha$ and $1/\gamma$ are the latent and infectious time scale; $1/\varepsilon$ is

Table 1: Initial state for the S and E compartments of the SEIR model. All other initial values are set to 0

| Parameter | meaning | before |
|-----------|-------------------------------|--------|
| S_0 | transmission rate | 10.06M |
| E_0 | Number of exposed individuals | 1 |

182 the time scale from illness onset to hospitalization and $1/\chi$ the hospitalization
time scale. Our model was similar to that in another study, which focused on
184 case counts in Wuhan (Wang et al., 2020).

The SEIR model is started at $t = t_0$, which should approach the date of in-
186 troduction of the virus. Here we assume that t_0 is unknown and is a parameter
estimated by the model. Because of the large number of intervention measures
188 taken, it is not clear at what time they have become effective. In the model, t_2 ,
the starting time of the third phase, was considered unknown, and estimated
190 by the data themselves.

For this model the effective reproduction number, R_t , can be calculated as

192

$$R_t = \frac{1}{I+A} \frac{\rho\beta}{\gamma} \left(\delta A + \frac{I/\varepsilon}{1/\gamma + 1/\varepsilon} \right) \quad (2)$$

R_t depends on time in this model, because I and A depend on time and be-
194 cause the effectiveness of intervention measures, i.e. the ρ s, depends on time.
We took the mean across time within a given period as the estimate of R_t .

196 We informed model parameters with estimates from other countries, where
available (Table S1, Hellewell et al. (2020)), and estimated the values of un-
198 known parameters by fitting the model to data on local reported positive cases

in the Lombardy region (see Table 2).

Table 2: Parameters and model values of the SEIR model for the three periods

| variable | meaning | period | | |
|---------------|--|-----------|-----------|-----------|
| | | 1 | 2 | 3 |
| t_0 | starting time of infection | | t_0 | |
| t | starting time of each period (days from January 1) | 0 | 54 | t_2 |
| β_0 | basic reproduction rate | β_0 | β_0 | β_0 |
| ρ | intervention effect | 1 | ρ_1 | ρ_2 |
| r | ascertainment rate | 0 | r_1 | r_2 |
| δ | ratio of unascertained/ascertained transmission rate | 1 | 1 | 1 |
| α | latency rate (days ⁻¹) | 0.19 | 0.19 | 0.19 |
| γ | infection rate (days ⁻¹) | 0.43 | 0.43 | 0.43 |
| ε | rate from illness onset to hospitalization (days ⁻¹) | 0.14 | 0.14 | 0.14 |
| χ | rate of hospitalization (days ⁻¹) | 0.033 | 0.033 | 0.033 |

200 To estimate the unknown parameters (t_0 , β_0 , ρ_1 , ρ_2 , r_1 , r_2 and t_2), we assumed that the number of ascertained cases estimated by the model with illness onset on day t , denoted as y_t , follows a Poisson distribution with rate parameter λ_t , $y_t \sim \text{Poisson}(\lambda_t)$, where λ_t is the number of notified positive cases
 202 on day t . Thus the likelihood function was
 204

$$\mathcal{L}(t_0, \beta_0, \rho_1, \rho_2, r_1, r_2, t_2) = \prod \frac{e^{-\lambda_t} \lambda_t^{y_t}}{y_t!}$$

In order to compute the *a-posteriori* estimator (i.e., the parameters that maximize the posterior function), we used rather vague priors and imposed the
 206

following feasibility intervals to sample from:

$$\left\{ \begin{array}{l} t_0 \in [0, 25] \quad (\text{January 1--January 25}) \\ t_2 \in [54, 84] \quad (\text{February 24--March 25}) \\ \beta_0 \in [0, 10] \\ \rho_1, \rho_2 \in [0, 1] \\ r_1, r_2 \in [0, 1] \end{array} \right. \quad (3)$$

208 The numerical computation of the posterior distribution is performed with
a Metropolis-Hastings (MCMC) algorithm from 100,000 iterations, using the
210 DRAM algorithm (Delayed Rejection Adaptive Metropolis, Haario et al. (2006)).
For prediction, we obtained CIs (Credible Intervals) by stochastic simulations
212 under the SEIR model with the sampled parameter values from MCMC. The
a-priori and *a-posteriori* probability density functions are shown in Figure 6.

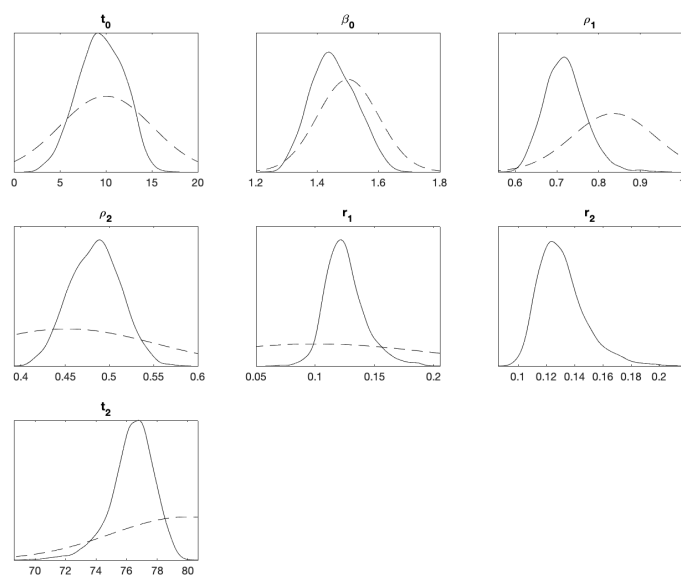


Figure 6: *a-priori* (dashed lines) and *a-posteriori* (solid line) probability density functions for all parameters of the SEIR model.