GCR60

Some epidemiological considerations on COVID mainly in China and Italy

Giorgio Parisi

A disclaimer: I am speaking on my own capacity, not as President of the Accademia dei Lincei

In this talk:

- General epidemiological considerations.
- The evolution of the epidemic looking to the symptoms.
- The evolution of the epidemic looking to the deaths.
- Underreporting of symptoms and deaths.
- Wider impact of public health

The reproductivity number is the average number of cases a given case produces.

It depends on time and is denoted R_t. Its value at the starting point of epidemics is called R_0.

R_0 is a function of the social structure, of local customs, maybe of the weather.

An other parameter is the slope. It is define as:

$$s(t) = \frac{d}{dt} \log(I(t)) \,, \tag{1}$$

where $I(t)$ is the number of infected cases at time t.

At the starting point we should have $s(t) = s_0$ and therefore

$$I(t) = \exp(s_0(t - t_0)) \tag{2}$$

How do you relate R_0 to s_0?

The serial time

A simple model.

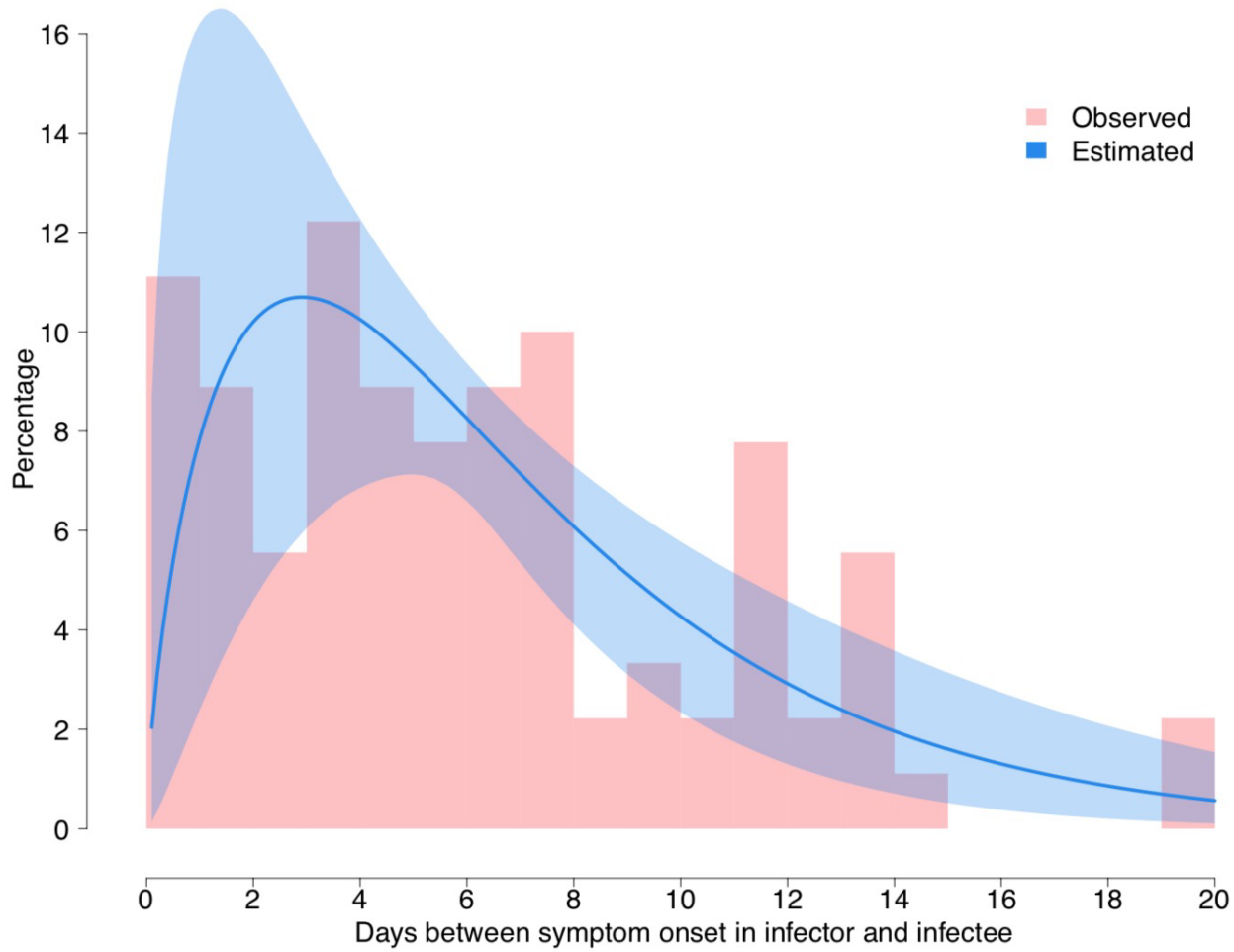
Let us assume that a given case infects all the R_0 (in average) people exactly after a serial time t_S. We get

$$I(t + t_S)/I(t) \equiv \exp(s_0 t_S) = R_0 \tag{3}$$

A shower model.

One assumes that the child infections are Poisson distributed with an average R_0: they are random independent events with a known distribution of the serial time. A shower type Montecarlo is done to compute s_0 from R_0.

The standard is the shower model that gives results that are not too different from the simple model (given also that we do not have a good measure of the serial time). Typical values of the serial time are in Italy 6.6 days (ISS), in Germany (RKI) 4.0 days.



A realistic model: the number or children infections are not Poisson distributed, e.g. it has a fat tail may be power-like (superspreaders), e.g. there are small family clusters of infected people.

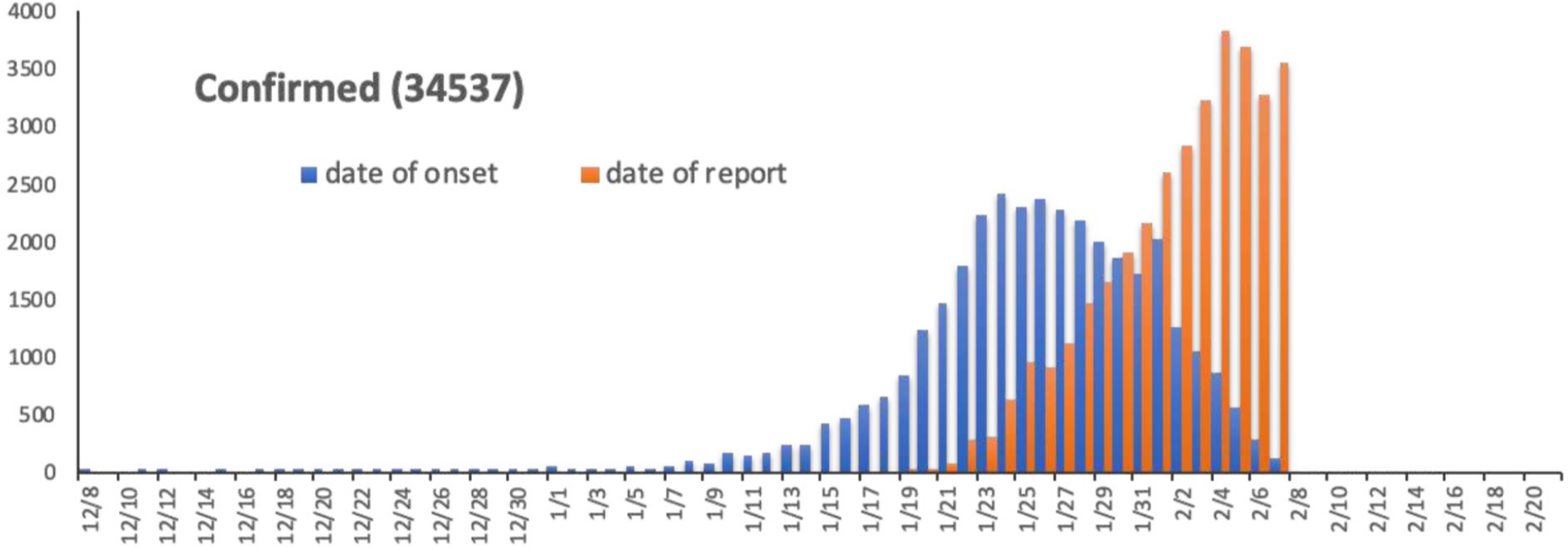
$I(t)$, i.e. the number of infected cases at time t, cannot be measured.

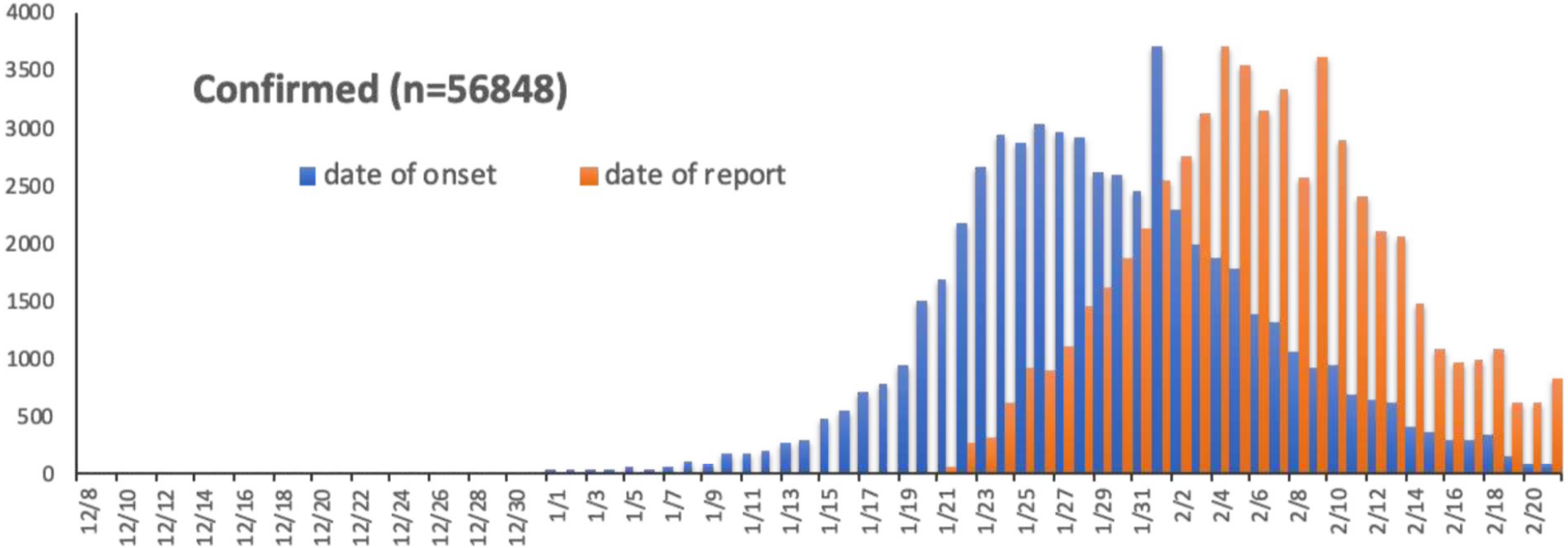
You can measure the number of people that have been tested positive at time t, i.e. $P(t)$, and (using those fractions that become symptomatic (before the test) or after the test (presymptomatic)) you can define the number of people that becomes symptomatic at time t, i.e. $S(t)$.

Let us forget for the time being presymptomatic and asymptomatic. If we have zero delay in testing $P(t) = S(t)$.

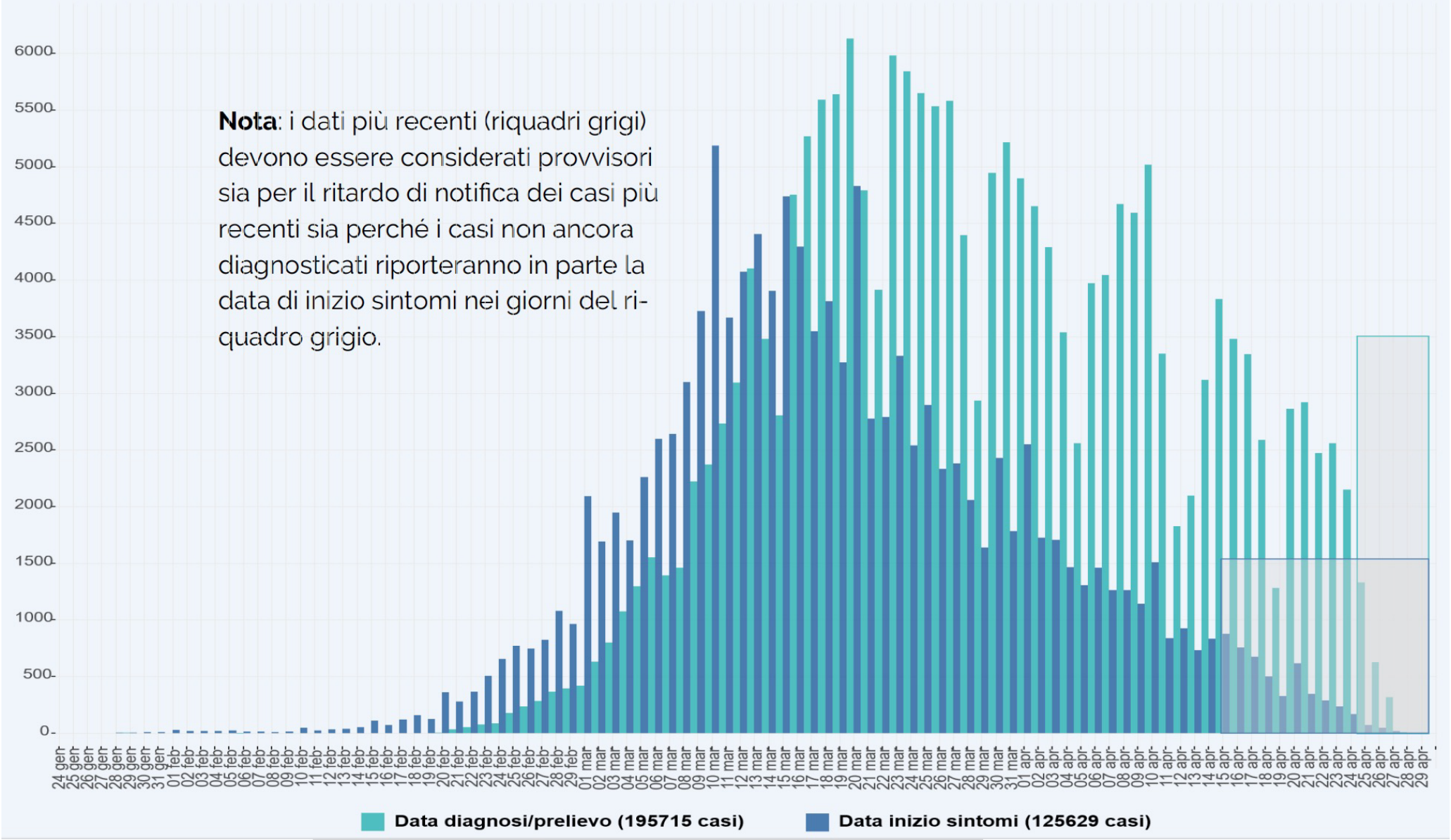
The interesting *intrinsic* feature is $S(t)$. It becomes known with some situation-dependent delay.

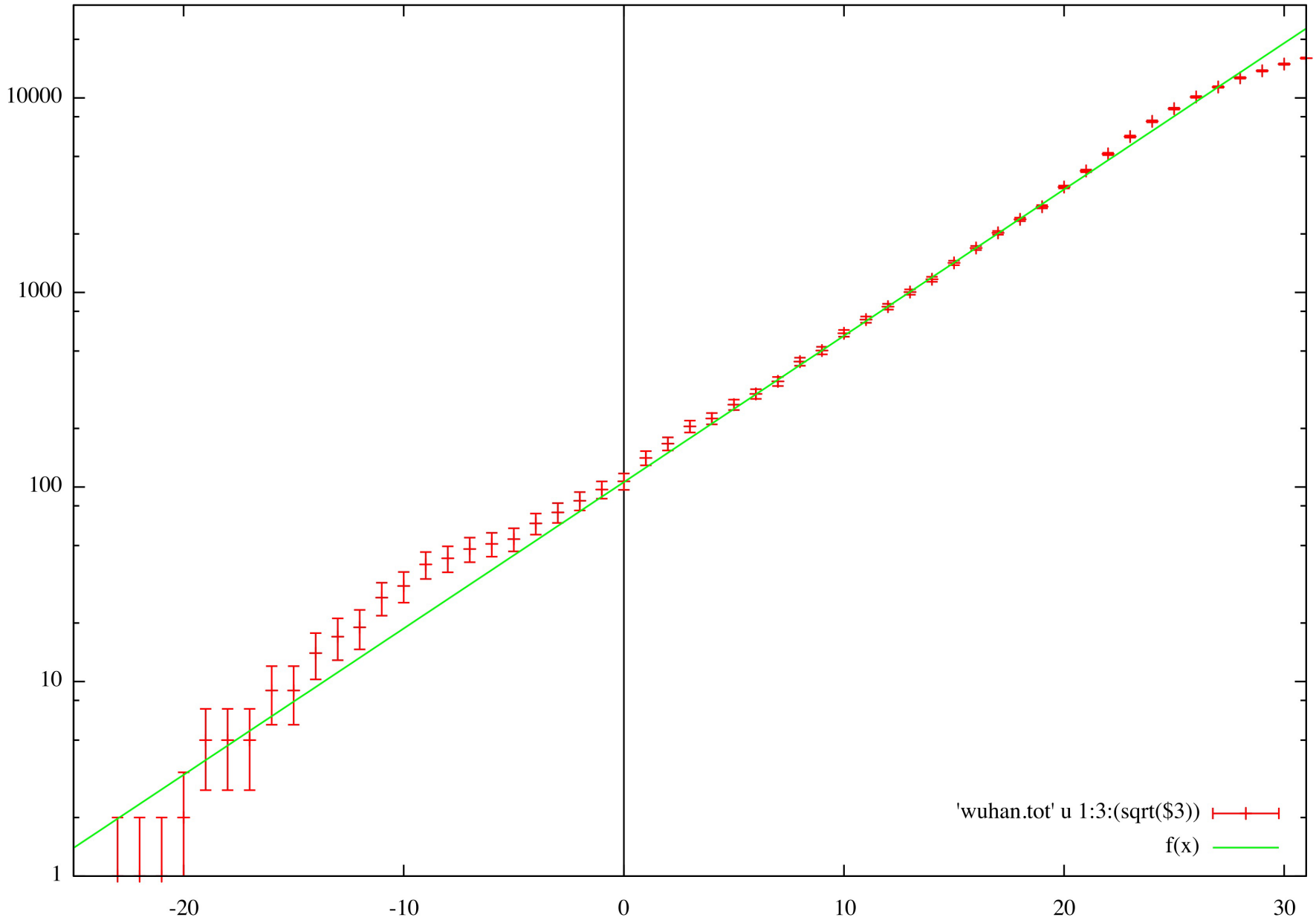
Let us see how it looks like:


Confirmed (34537)
date of onset
date of report
4000
3500
3000
2500
2000
1500
1000
500
0
12/8
12/10
12/12
12/14
12/16
12/18
12/20
12/22
12/24
12/26
12/28
12/30
1/1
1/3
1/5
1/7
1/9
1/11
1/13
1/15
1/17
1/19
1/21
1/23
1/25
1/27
1/29
1/31
2/2
2/4
2/6
2/8
2/10
2/12
2/14
2/16
2/18
2/20

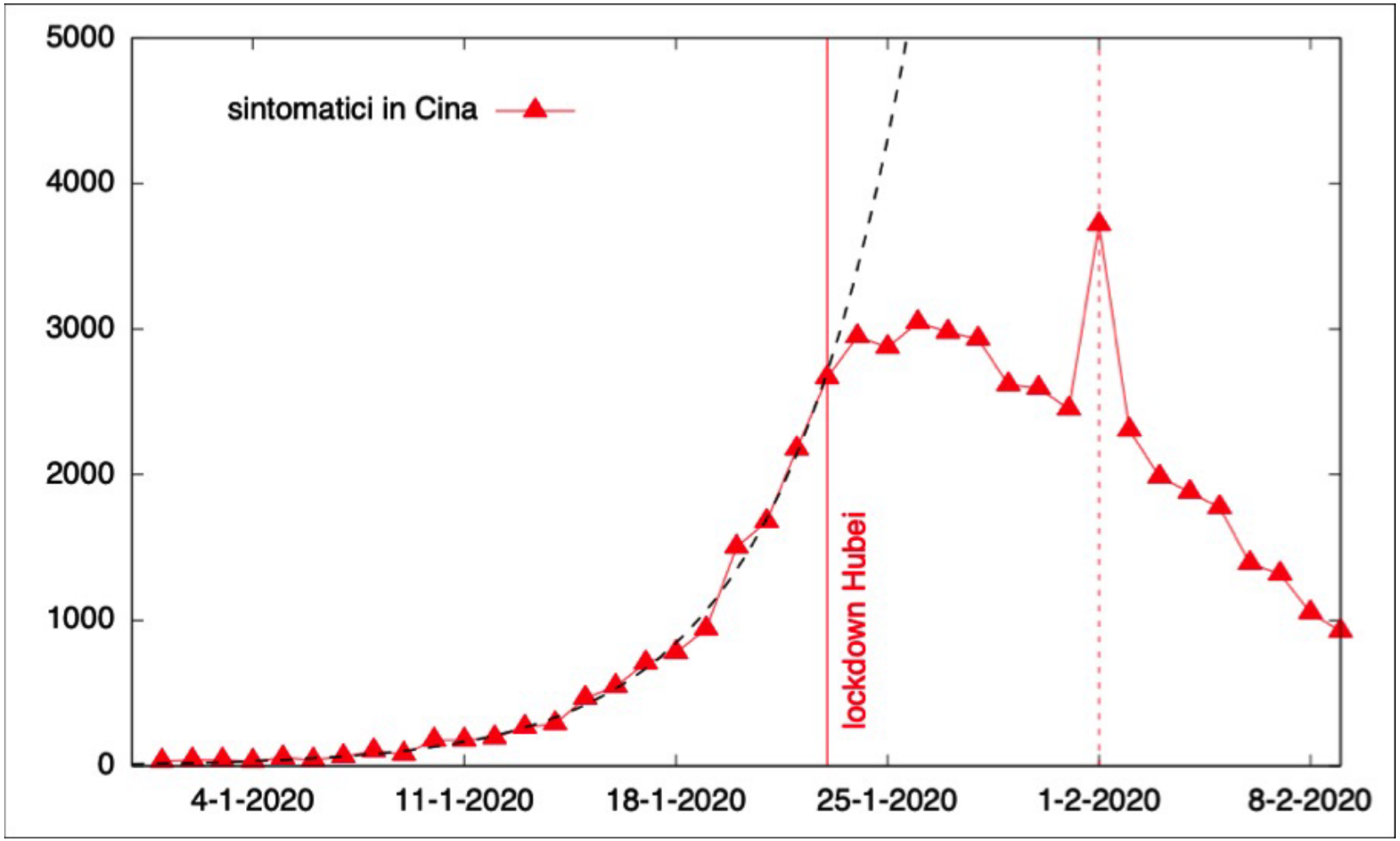

Confirmed (n=56848)
date of onset
date of report
4000
3500
3000
2500
2000
1500
1000
500
0
12/8
12/10
12/12
12/14
12/16
12/18
12/20
12/22
12/24
12/26
12/28
12/30
1/1
1/3
1/5
1/7
1/9
1/11
1/13
1/15
1/17
1/19
1/21
1/23
1/25
1/27
1/29
1/31
2/2
2/4
2/6
2/8
2/10
2/12
2/14
2/16
2/18
2/20

AGGIORNAMENTO 28 aprile 2020

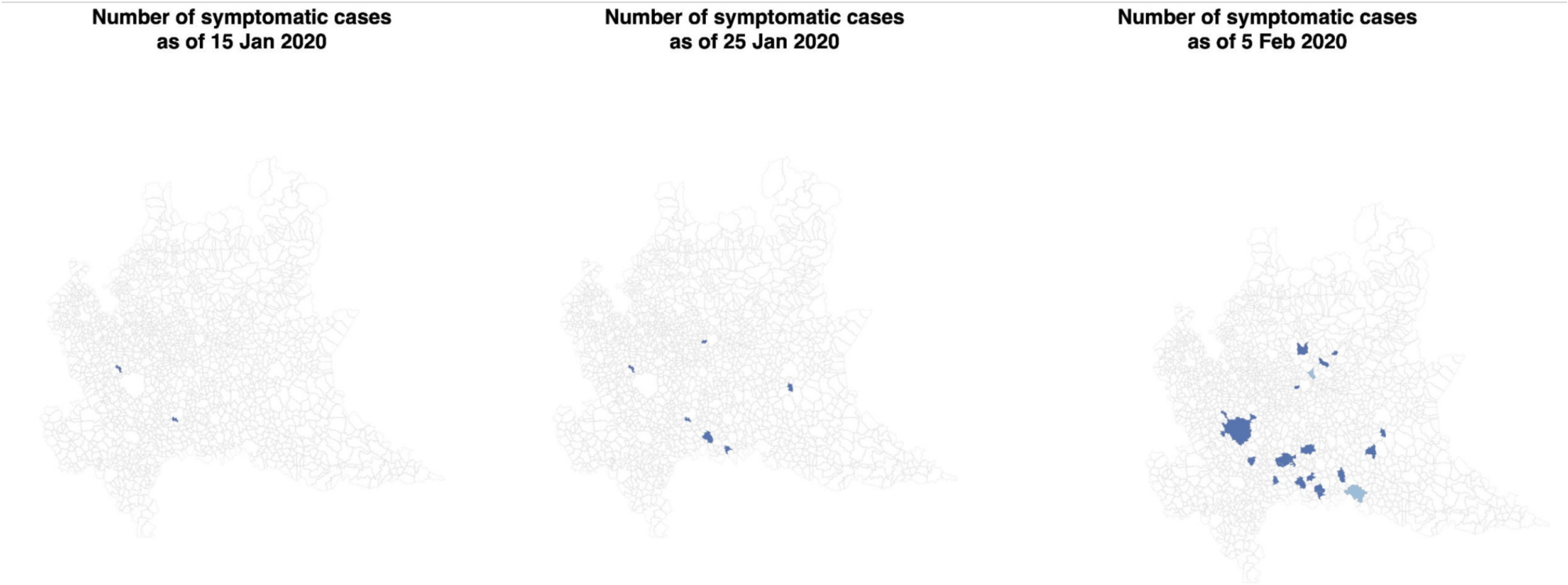




$s_0 = 0.173 \pm 0.003$ Doubling time 4 days

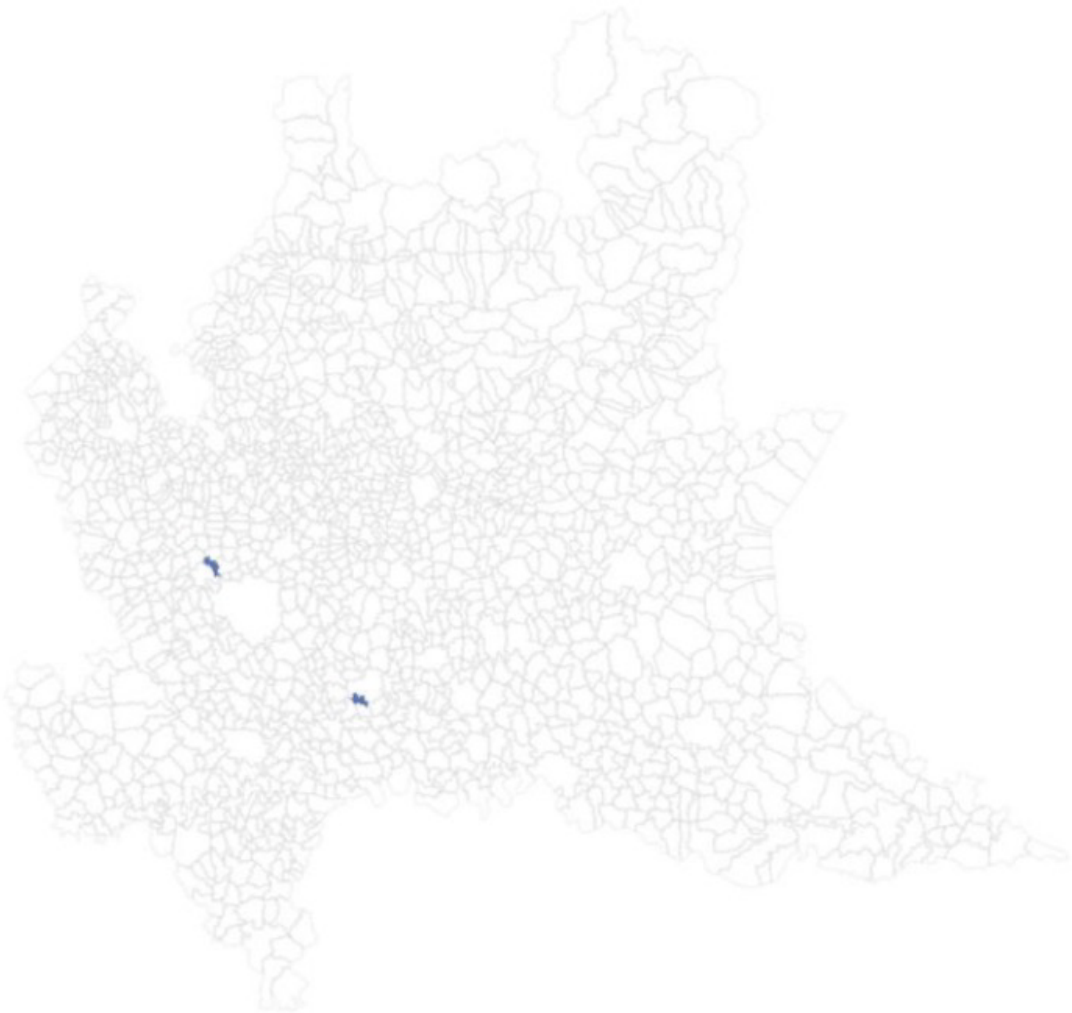


Huge underreporting! Lack of tests, fear to go to hospitals, disorganization!

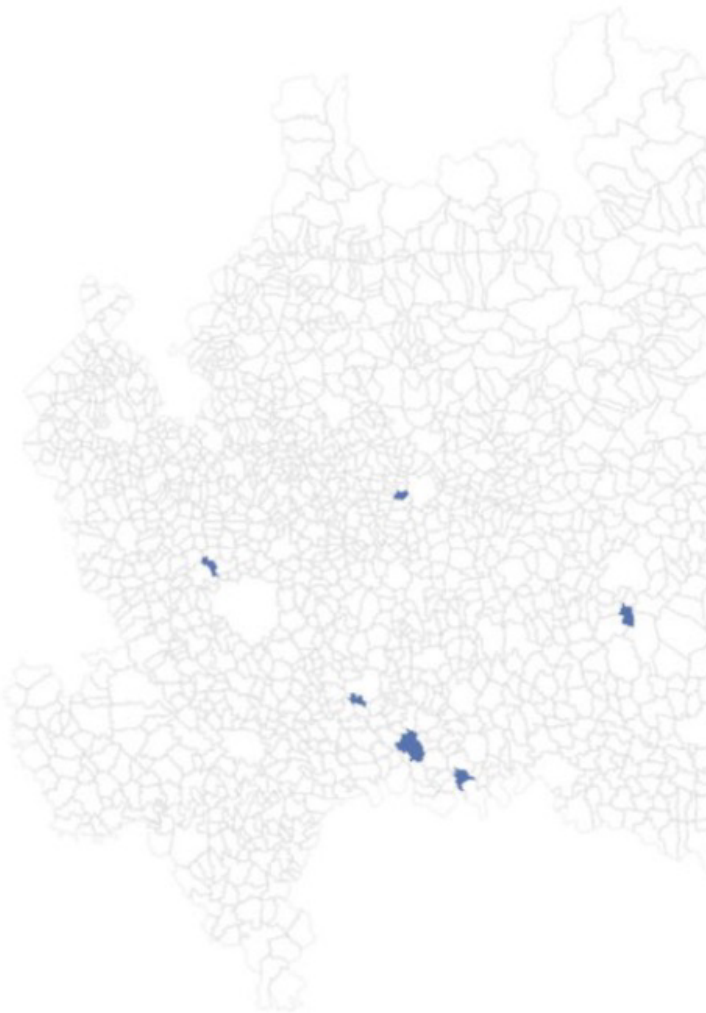


https://arxiv.org/abs/2003.09320

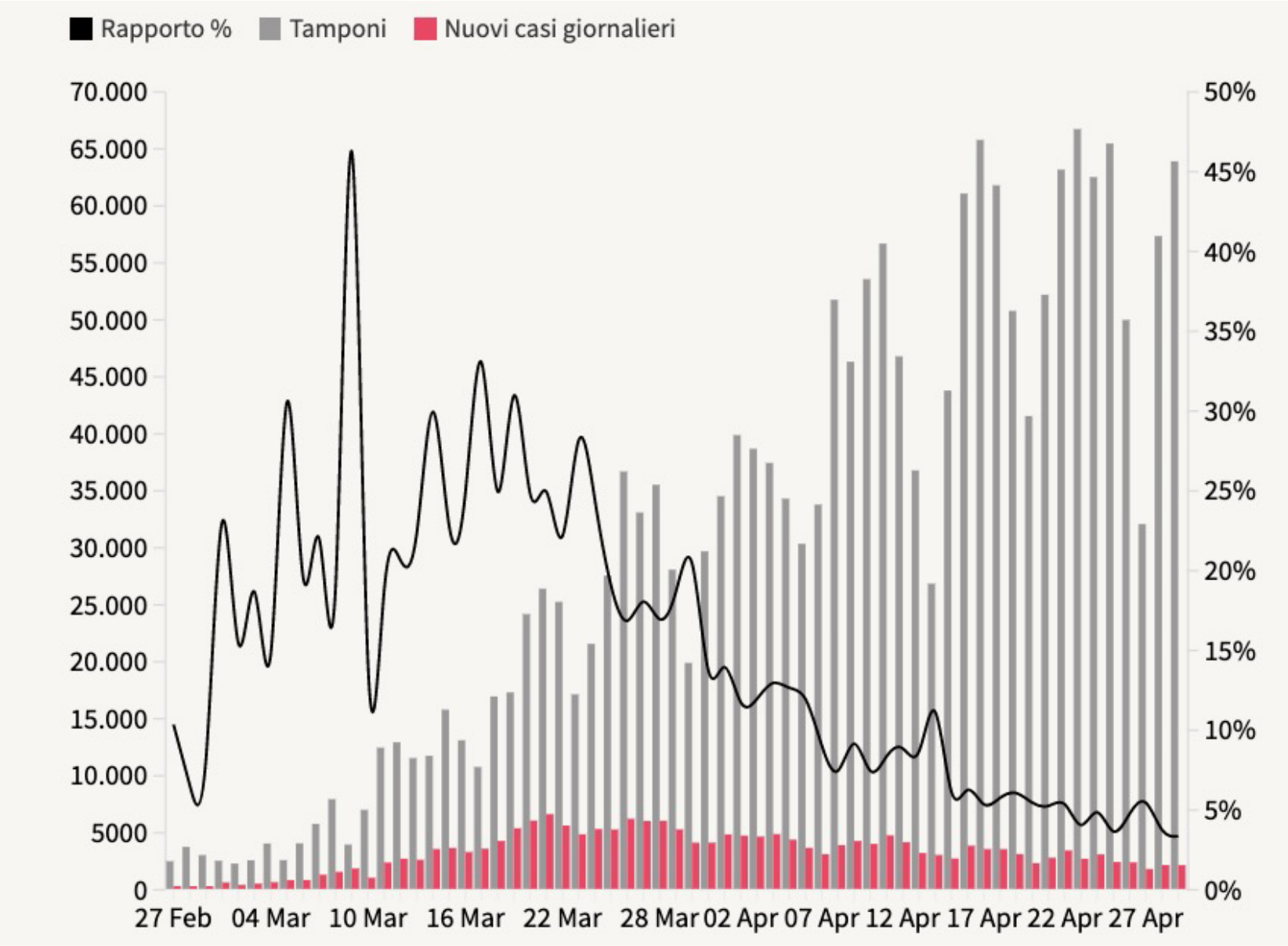
Number of symptomatic cases as of 15 Jan 2020

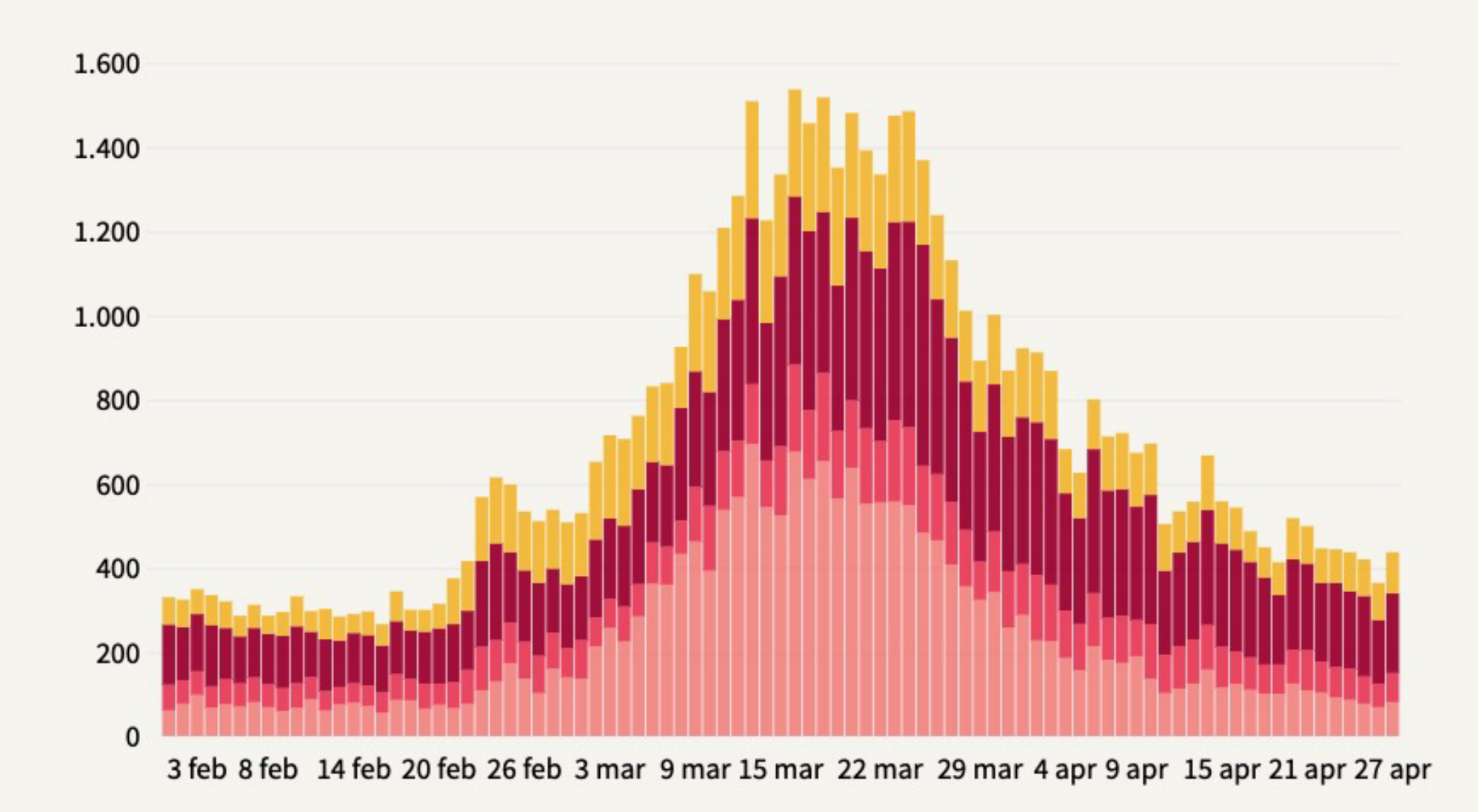


Number of symptomat as of 25 Jan 202

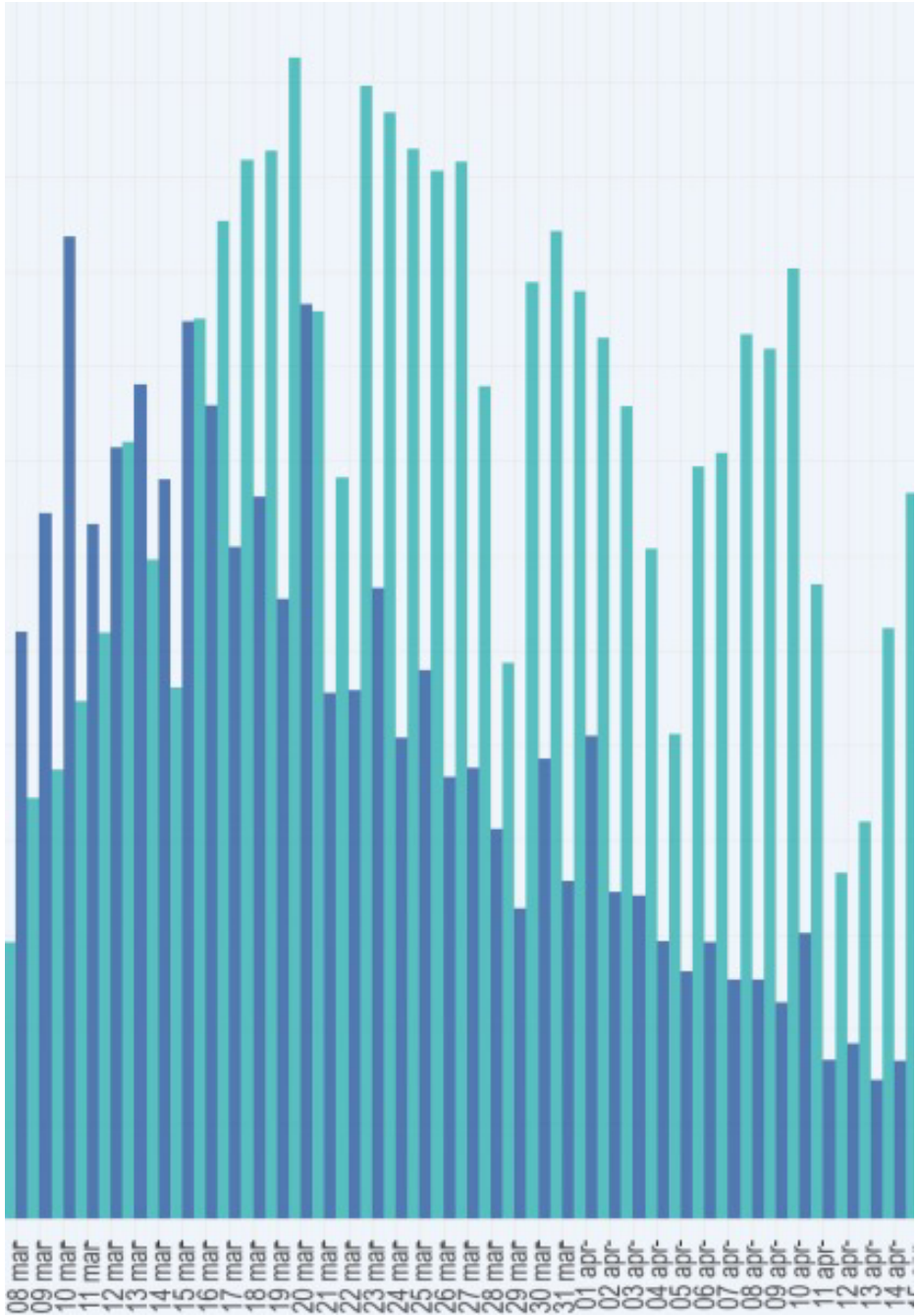


Underreporting in Italy? Let us see some data from Sole24Ore





Emergency calls in Lombardy per day for respiratory problems



Data from March 8 to April 14

Which is the time relation among symptoms and other events? Let us look at the median!

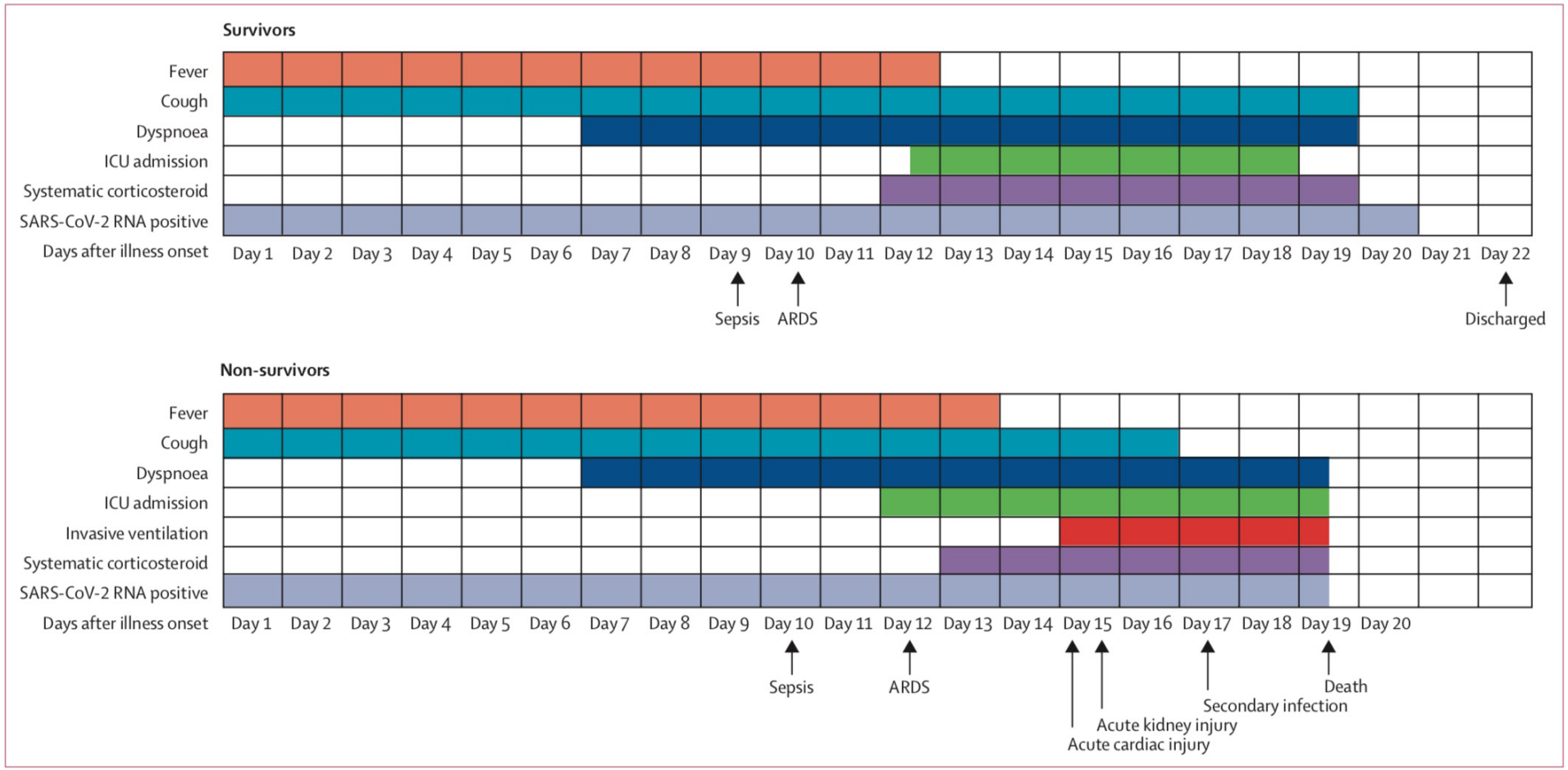


Figure 1: **Clinical courses of major symptoms and outcomes and duration of viral shedding from illness onset in patients hospitalised with COVID-19**
Figure shows median duration of symptoms and onset of complications and outcomes. ICU=intensive care unit. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2. ARDS=acute respiratory distress syndrome. COVID-19=coronavirus disease 2019.

We have seen the median. Which is the distribution: not so many data. The typical fits are Lognormal functions:

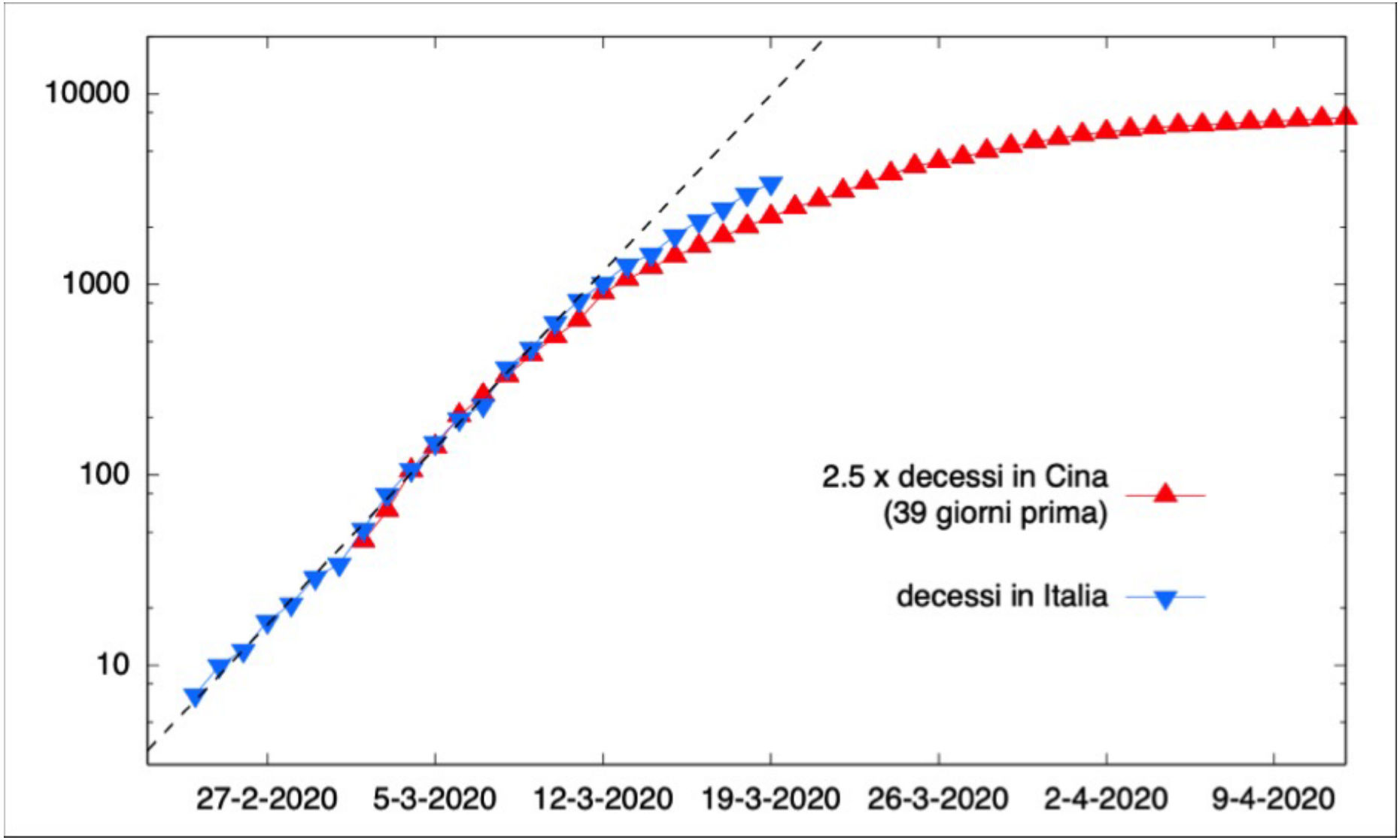
$$P(t) \propto t^{-1} \exp\left(-A(\log(t) - \log(t^*))^2\right) \tag{4}$$

Table 2. Right-truncated incubation period and other time-delay distributions for COVID-19 outbreak cases reported in Janu lognormal distribution.

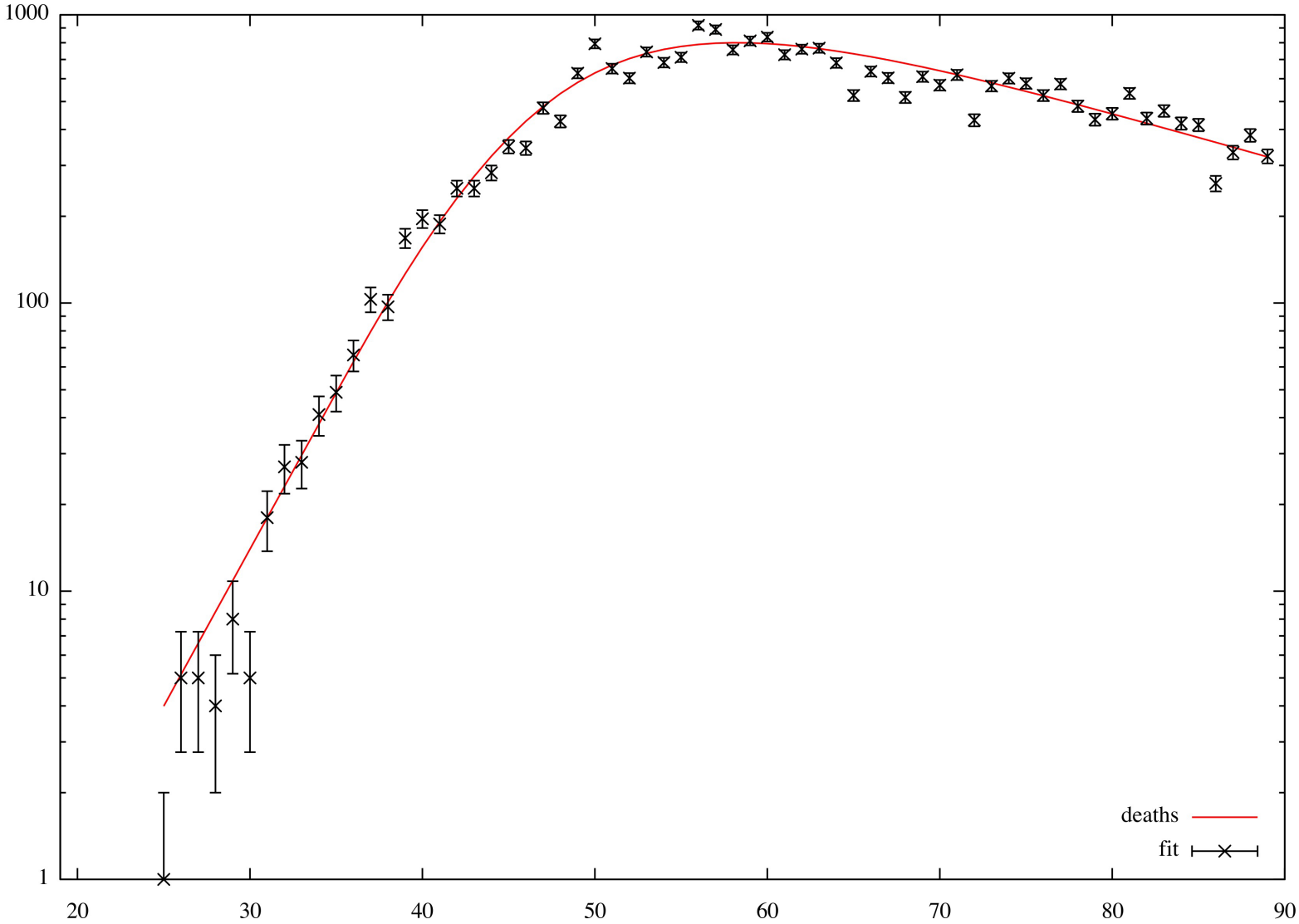
Lognormal	Incubation Period Excluding WR (Days)	Onset to Hospital Admission, Living (Days)	Onset to Hospital Admission, Deceased (Days)	Onset to Death (Days)
Mean	5.6 (4.4, 7.4)	9.7 (5.4, 17.0)	6.6 (5.2, 8.8)	20.2 (15.1, 29.5)
SD	3.9 (2.4, 6.9)	35.2 (12.9, 84.5)	4.9 (3.0, 8.6)	11.6 (6.6, 21.8)
5%	1.7 (1.1, 2.3)	0.2 (0.1, 0.3)	1.9 (1.2, 2.5)	7.4 (5.6, 9.5)
Median	4.6 (3.7, 5.7)	2.6 (1.9, 3.8)	5.3 (4.2, 6.8)	17.1 (13.5, 24.1)
95%	12.3 (9.1, 19.8)	35.1 (20.5, 65.0)	15.0 (11.2, 23.8)	39.9 (27.9, 69.6)
99%	18.6 (12.7, 34.2)	102.8 (52.8, 218.4)	23.2 (16.0, 41.6)	56.7 (37.0, 109.2)

Median incubation time 4.6 ± 1
Median time from infection to death 22 ± 4

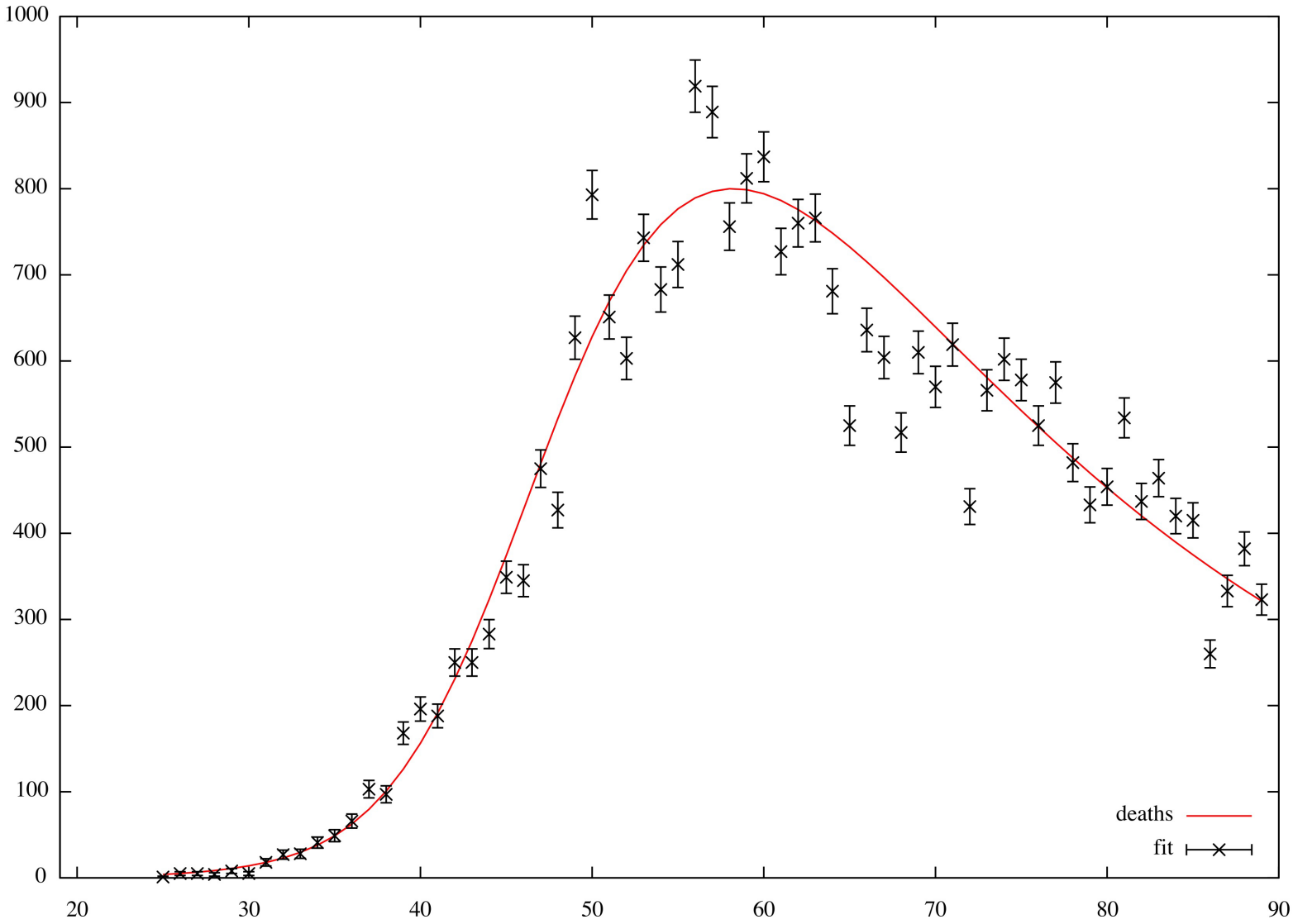
Let us look to the deaths!



(Enzo Marinari, G.P., Federico Ricci Tersenghi)



Day count starts on January 1



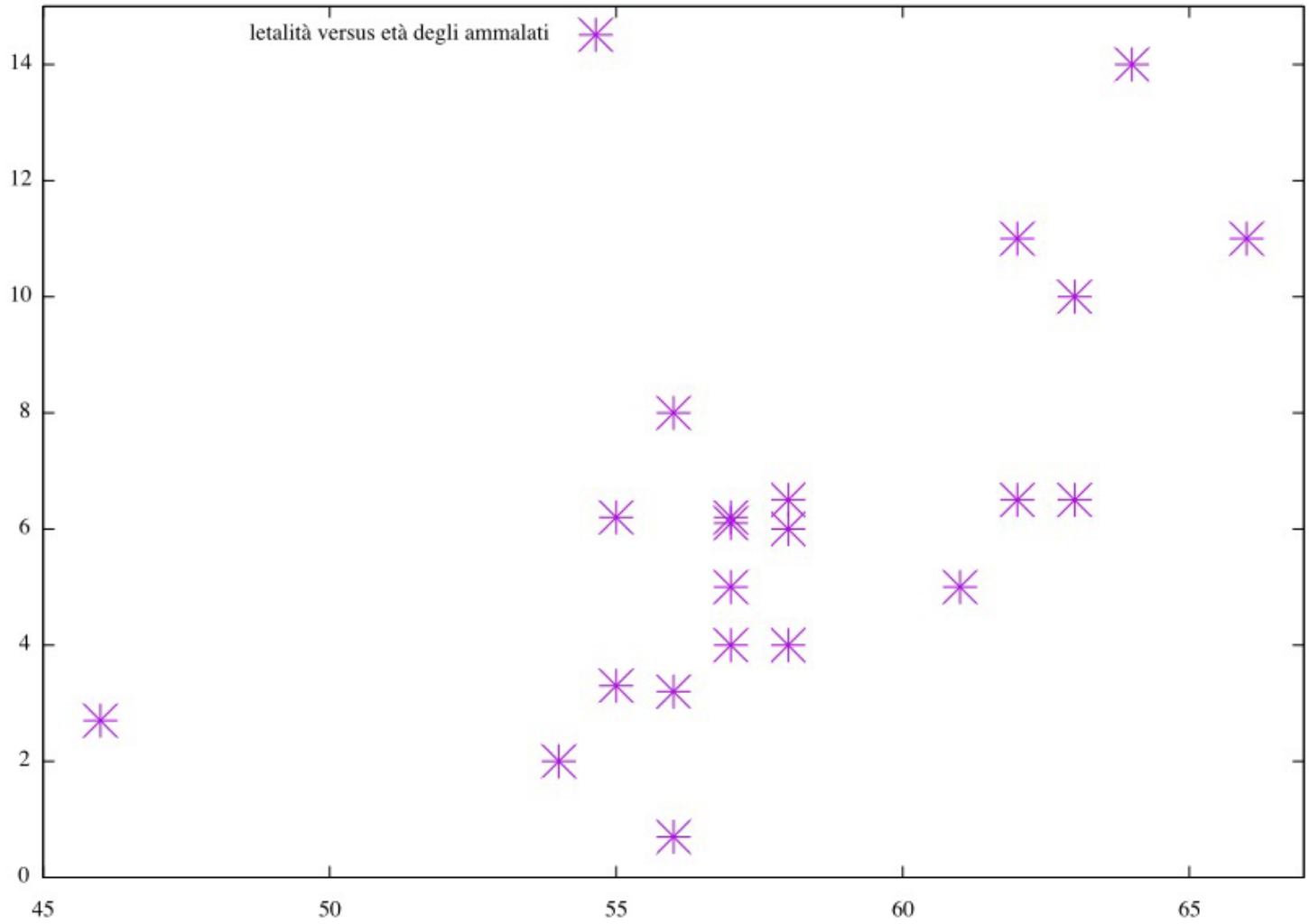
Day count starts on January 1

Assuming an asymptotic case fatality rate of the illness of 2.5, the same analysis estimates that there has been an underreporting of a factor 10: the total number of cases could be a factor 10 bigger of the officially reported.

The case fatality rate is the ratio between dead people and infected people. It is a function of time that starts from zero.

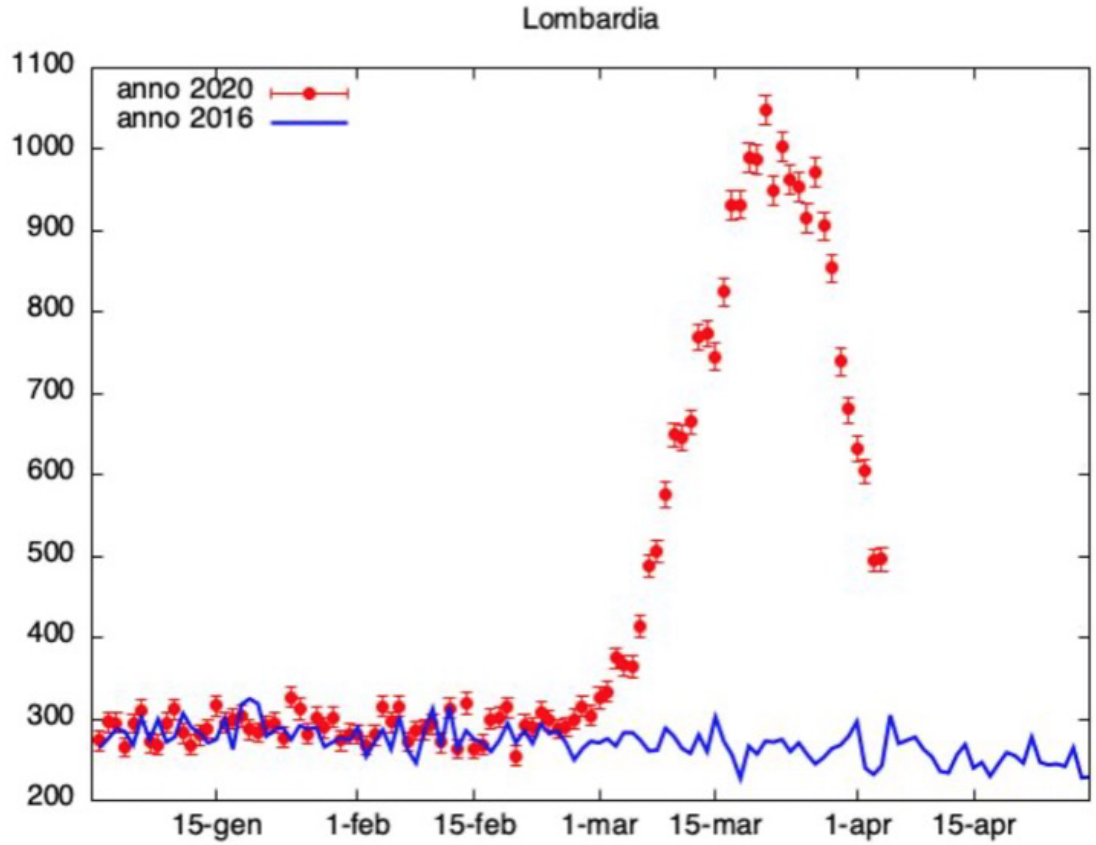
If there is only one wave it is an increasing function of time: e.g. in the last stage there are no new cases but people still die.

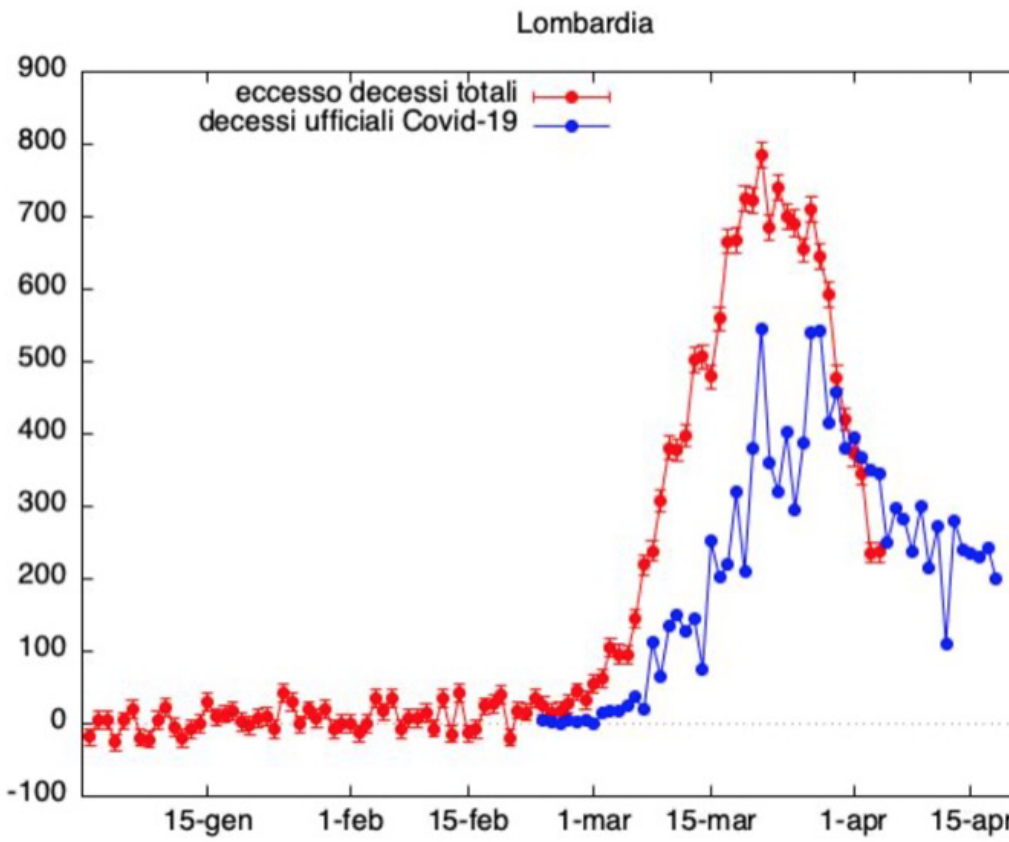
The smoking gun for underreporting is the average age of infected people. Underreported cases are usually mild.

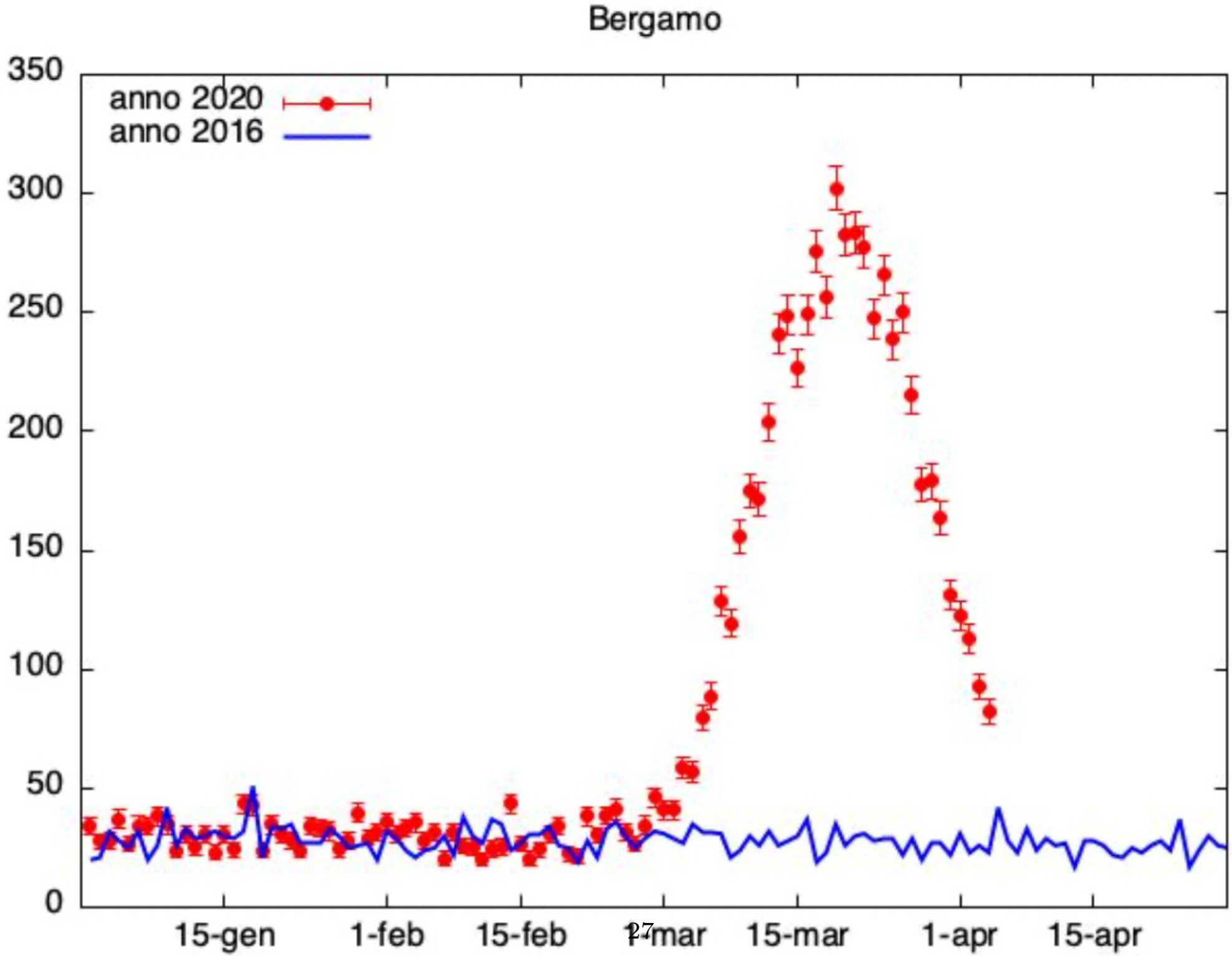


Fatality rate versus average age in different regions of Italy and Germany.

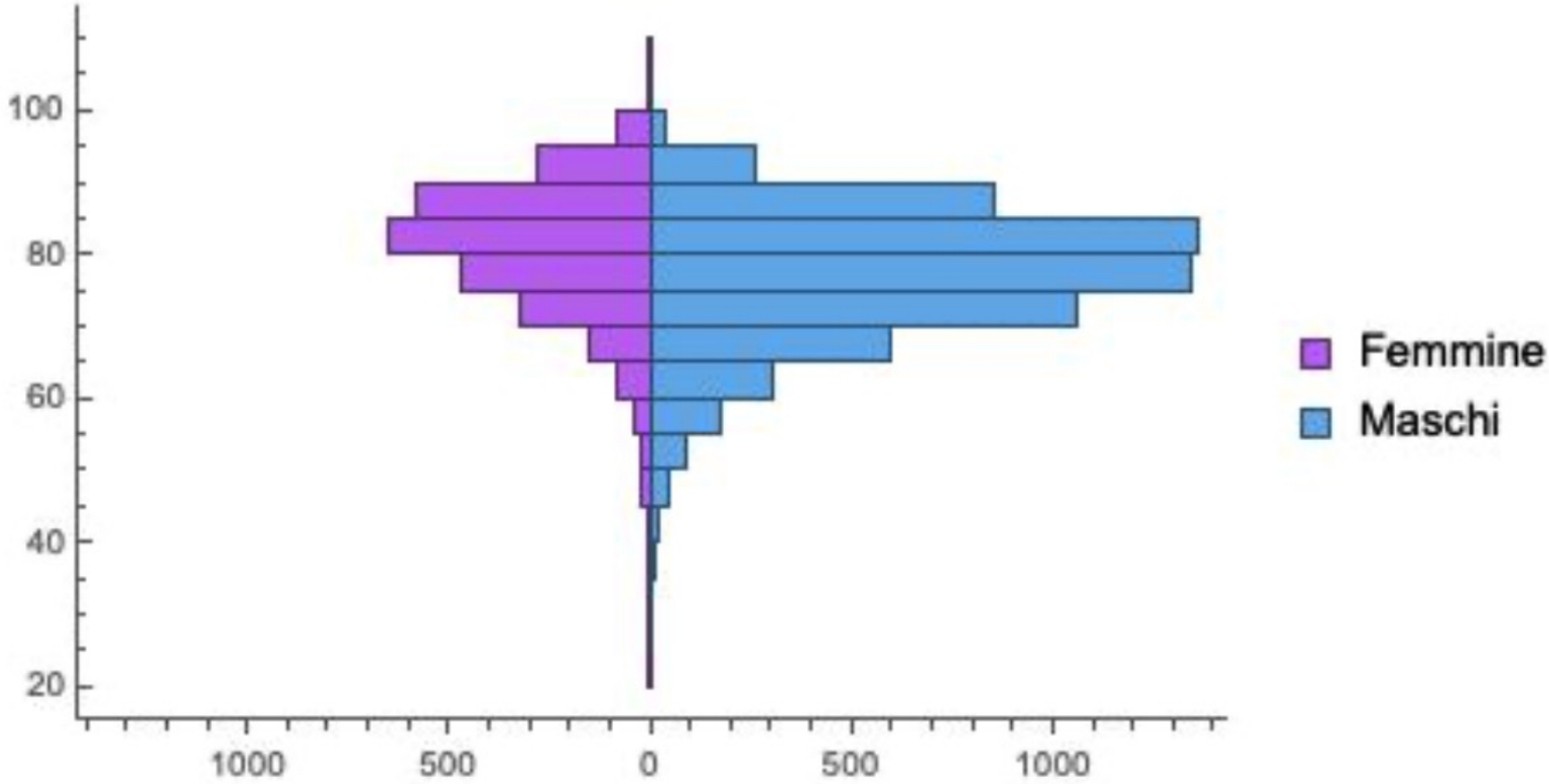
Underreporting of deaths? (E. Bucci, E. Marinari, L. Leuzzi, G.P., F. Ricci Tersenghi)






Bergamo
anno 2020
anno 2016
350
300
250
200
150
100
50
0
15-gen
1-feb
15-feb
1-mar
15-mar
1-apr
15-apr

Sex ratio of deaths



Let us call f_c the percentage of women that officially die from COVID. This number changes with the country

Cina 36.2, Spain 36.6, Italy 35.8, Lombardy 30.6, France 40, Germany 42, Belgium, Canada and South Corea 48-49.

We can assume that in the excess of death respect to the previous there are some COVID contributions and some non-COVID contributions.

In Lombardy, we have an excess of 15000-16000 deaths against 8905 COVID certified deaths.

Just doing linear algebra, assuming that the non-COVID deaths are sex-independent, we get in Lombardy

If $f_c = .36$, unofficial COVID deaths 1300-2500, non-COVID deaths 4200-5400

If $f_c = .4$, unofficial COVID deaths 1900-3700, non-COVID deaths 3000-4800

Very tentative conclusions

Underreporting is a common phenomenon

Underreporting of clinical cases may be very large, maybe a factor 10.

The underreporting of COVID deaths is not very strong. The true value of COVID deaths in Lombardy may be higher of some tens of percent.

A significant increase in non-COVID deaths is very likely.

It is difficult to reach firms conclusions with public available data.