Estimation of weighted instant case fatality rate of COVID-19 in Italy.

Massimo Pierini[1,*], Alessio Pamovio[1], and Sandra Mazzoli[1]

[1]EpiDataItalia group (epidata.it)
[*]Correspondence info@epidata.it

November 3, 2020

Abstract

Case Fatality Rate (CFR) is usually estimated as total death over total cases. This estimate though does not take into account likely dynamic variations of fatality along the time line. Is here presented a dynamic estimation of instant case fatality rate based upon onset-to-death interval: weighted instant Case Fatality Rate (wiCFR). Estimation of wiCFR in Italy shows values and trend different from classic cumulative CFR that can be useful to better understand current situation: the trend is indeed increasing from half September, differently from classic CFR apparently decreasing.

1 Introduction

The novel beta-coronavirus SARS-CoV-2 is causing a new pandemic; starting from December 2019, the disease COVID-19, started in Wuhan.

WHO announced that the ongoing COVID-19 outbreak is a Public Health Emergency of International Concern (PHEIC) and a pandemic [1].

Wuhan City and Hubei Province were the first affected and accounted for the majority of the cases. At present the outbreak in China is under control with the exception of sporadic clusters.

However, the cases of COVID-19 have rapidly increased in almost all the other countries around the world. Italy, from the western countries, was apparently the first to be infected, from February 22, showing the first case in Lombardy region. At moment the global pandemic outbreak is getting worse and worse with 44,500,017 cases and 1,174,314 deaths, according to the Coronavirus Resource Center of the John Hopkins University, October 28, 2020.

To identify the hazard indicators in an infectious disease is very important and pivotal in epidemiology. The case fatality rate, CFR [2], the percentage of deaths from a disease in the total number of infected-patients during the whole outbreak reflects the severity of the disease. The real CFR of an infectious disease could be obtained only after the epidemic is finished, on the basis of the numbers of confirmed cases and deaths and gives the better accurate estimate. At moment, given the severity of COVID-19 outbreak situation throughout the world, it is important to estimate and to predict COVID-19 CFR early and continuously in the epidemic and in each single countries; focusing on the severity evaluation might give the basis for the setting of public decisions to ensure a good and efficient security strategy on how to allocate public health resources and about the adjustment of medical treatment work. It is also of great help to the government's decision-makers and for citizen compliance and acceptance of the containment measures.

However, since the spreading of the virus worldwide, the numbers of confirmed cases, deaths and cured cases are constantly changing.10 Therefore, estimating the CFR needs to be continuously updated. This need was not in general accepted and considered, with the exception for the work of Cao et al. [3], written in March at the beginning of the epidemic in China, where they considered the cumulative daily data.

Case Fatality Rate (**CFR**) is usually estimated as total deaths over total cases [4] [5] [6]. However, this estimate does not take into account the possible differences between the observed cases during different periods, in fact other estimates of the instant CFR have been proposed [3]. In Italy, for example, in the so-called "first wave" the observed cases were mostly symptomatic, many of them required hospitalization and sufficient nasopharyngeal swabs were not available to carry out a population screening. From June onwards, the availability of tests increased and screening along with more effective contact tracking were also possible, so a growing percentage of new cases from June onwards were asymptomatic or paucisymptomatic and not hospitalized. This difference can lead to an incorrect estimate of the fatality rate by overestimating in the first period and underestimating later.

A more reliable estimate of the CFR can be obtained from the ratio of new deaths to new cases, taking into account the onset-to-death interval (time from onset of symptoms to eventual death) and estimating the median, interquartile and 95% confidence interval.

We will call this estimate **weighted instant Case-Fatality-Rate** (**wiCFR**).

The advantage of wiCFR, compared to other estimates proposed for the instant fatality rate, is that it can be calculated only from the data of new deaths and new cases, assuming a valid distribution of the onset-to-death interval. However, it remains a raw rate, not standardized for sex, age and other possible covariates and sensitive to an inappropriate onset-to-death distribution. Nonetheless, it can be a valuable tool for assessing instant and dynamic fatality along the timeline and can provide information on unexpected variations that could not be highlighted by observing the cumulative CFR alone.

---

2 Data

The official data published by the *Presidenza del Consiglio dei Ministri - Dipartimento di Protezione Civile* on GitHub will be used [7] `dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv` for Italy and `dati-regioni/dpc-covid19-ita-regioni.csv` for regions, grouped into three zones (North; Center; South and Islands).

The column for the new total cases is named `nuovi_positivi`. The daily change in deaths is calculated as the difference in the `deceduti` column.

3 Method

As estimated by Verity et Al. [8] the onset-to-death interval for COVID-19, i.e. from the onset of symptoms to eventual death, can be described as a Gamma distribution with mean $\mu = 18.8$ days and coefficient of variation $c_v = \sigma/\mu = 0.45$, from which we get $\sigma = 8.46$ days:

$$\text{onset-to-death} = \Gamma(\mu_{=18.8}, \sigma_{=8.46})$$

from which we can get shape and rate

$$\text{shape} = \mu^2/\sigma^2 = 18.8^2/8.46^2 = 4.9383$$

$$\text{rate} = \mu/\sigma^2 = 18.8/8.46^2 = 0.2627$$

for parameterization with `scipy.stats.gamma` [9], where `a`=shape and `scale`=1/rate (fig. 1).

We will choose a standard confidence interval (C.I.) of 95% ($\alpha = 0.05$).

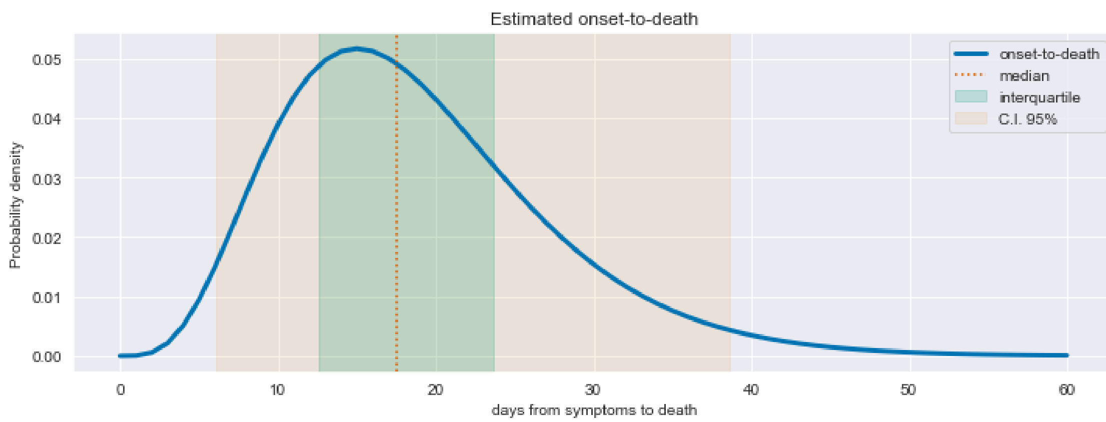


Figure 1: Onset-to-death distribution as estimated by Verity et al.

which can be discretized for each day $i = [0, 60]$ (fig. 2)

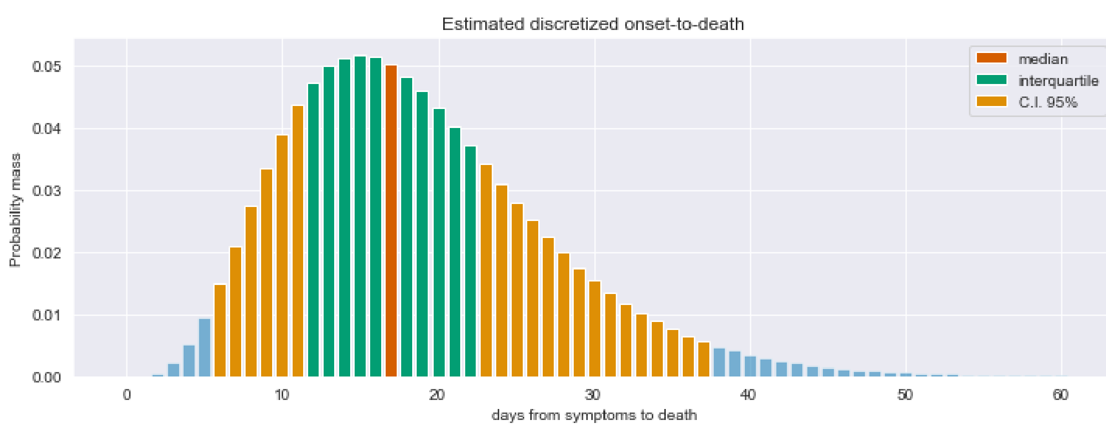


Figure 2: Discretized Onset-to-death distribution.

It is thus possible to calculate, for each day t in the time series of COVID-19 in Italy, the rate of new deceased in t over new cases observed in each $t-i$, assigning to each result the corresponding probability density of the discretized onset-to-death distribution

$$\text{CFR}_{t,i} = \frac{\Delta(\text{deaths})_t}{\Delta(\text{cases})_{t-i}}$$

where $i = [0...60]$ and

$$p(\text{CFR}_{t,i}) = p(\text{onset-to-death})_i$$

In the example plot fig. 3, the size of the points corresponds to the probability density:

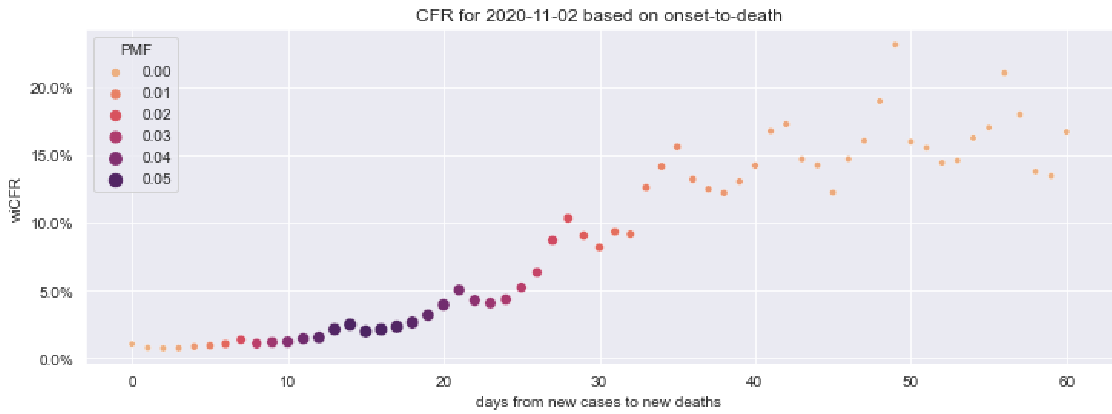


Figure 3: CFR for each day in onset-to-death distribution. The size of the points corresponds to probability density.

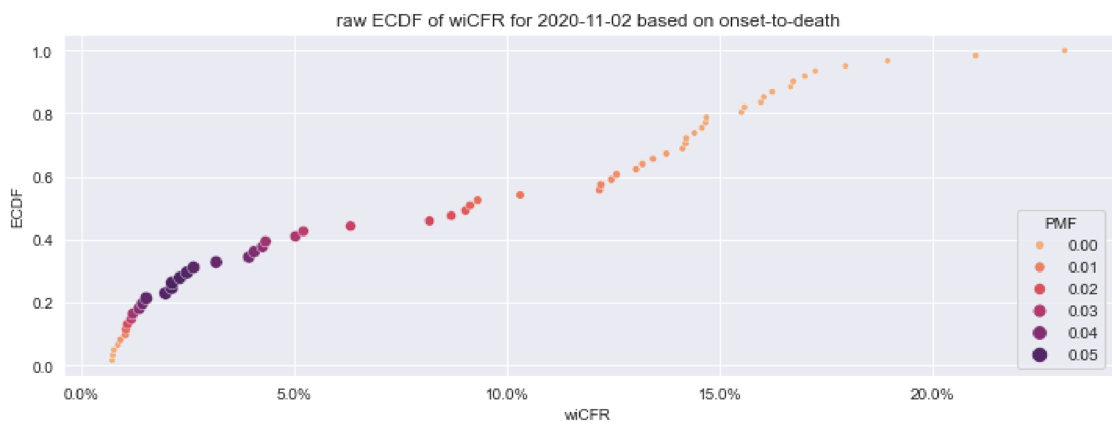


Figure 4: Estimated cumulative density function of CFR. Point size corresponds to its weight.

In order to estimate median, interquartile and confidence interval, we will proceed with resampling of one million samples with bootstrap method weighted on the discretized probability density of the onset-to-death interval (`numpy.random.choice` [10]). The estimate would not be correct on the unweighted data (fig. 4) because the probability of each single value is determined by the corresponding probability density of the discretized onset-to-death interval. Weighted resampling ensures that each value is chosen randomly with a probability equal to the relative probability density (fig. 5 and fig. 6).

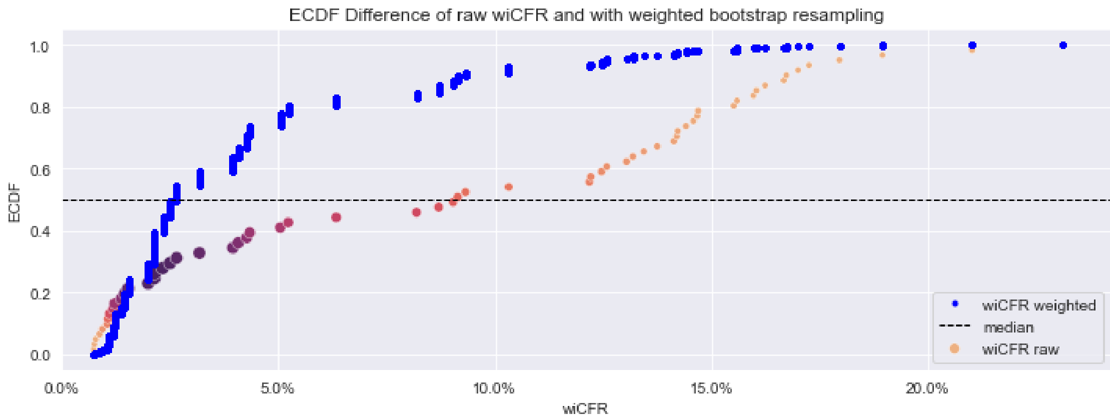


Figure 5: Weighted estimated cumulative density function of CFR.

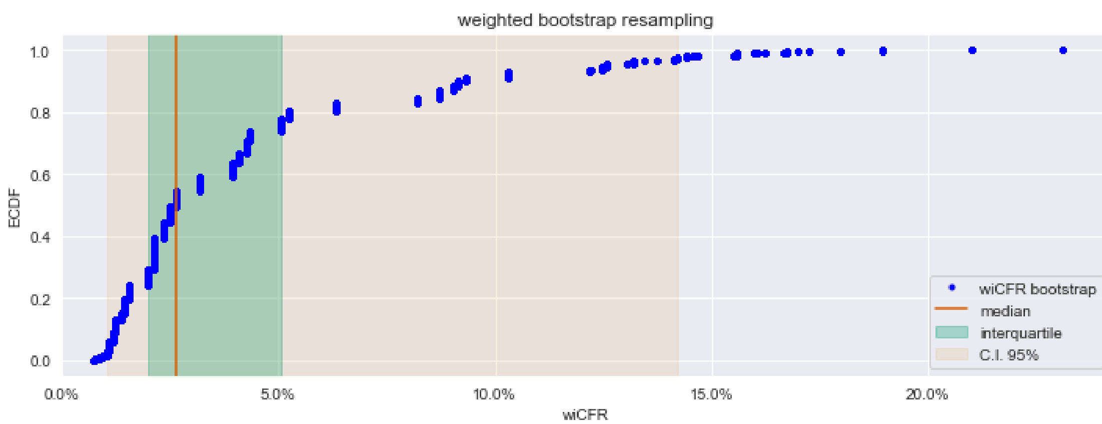


Figure 6: Weighted estimated cumulative density function of CFR. Median, interquartile and confidence interval are shown.

---

4 Results

Repeating the calculation backwards, for each day in the time series up to

`2020-04-23`

a time series of estimated weighted instant CFRs for each day t is obtained (fig. 7):

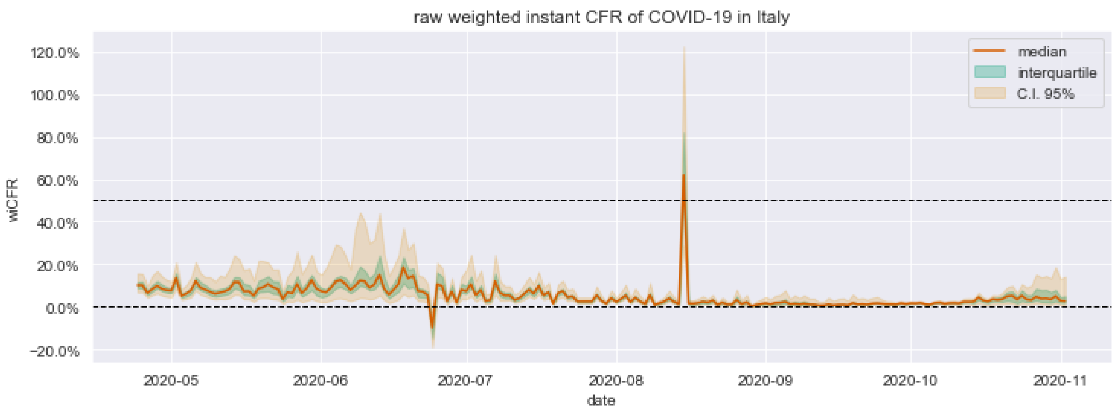


Figure 7: Raw weighted instant CFR of COVID-19 in Italy.

from which we can filter outliers less than 0 and greater than 0.5, imputing the average over a 7-day window (fig. 8):

```
OUTLIERS in col median (0< or >0.5)
========================================
              median        Q1        Q3        lo        hi
date
2020-06-24 -0.097484 -0.147619 -0.078086 -0.190184 -0.038130
2020-08-15  0.619608  0.516340  0.818653  0.286232  1.224806

Impute function `mean`, window 7
========================================
              median        Q1        Q3        lo        hi
date
2020-06-24  0.065129  0.035742  0.093421  0.004055  0.150556
2020-08-15  0.107225  0.088993  0.141138  0.051793  0.211517
```

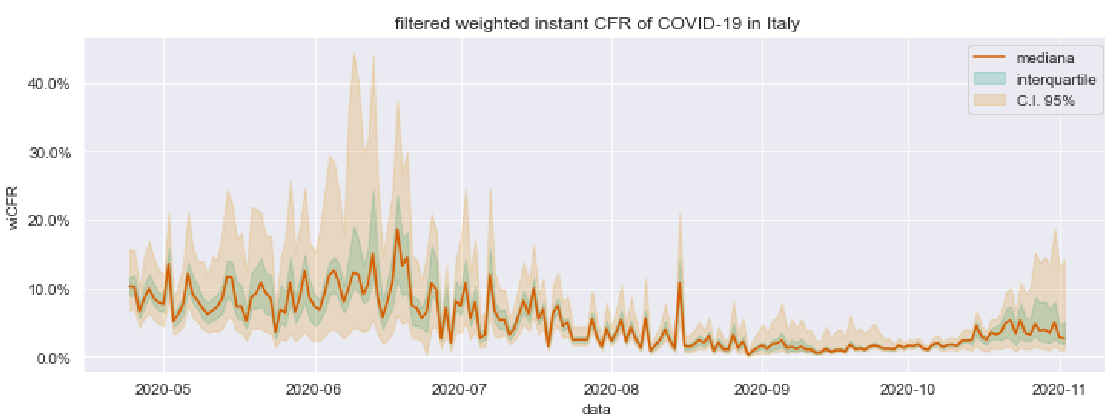


Figure 8: Filtered weighted instant CFR of COVID-19 in Italy.

and smoothing to reduce the natural volatility of the raw epidemiological data, with a fit on a 6th degree polynomial (`numpy.polyfit` e `numpy.poly1d` [10]), estimating interquartile (IQR) and confidence interval (C.I.) from the fit of IQR and C.I. of the filtered wiCFR (fig. 9)

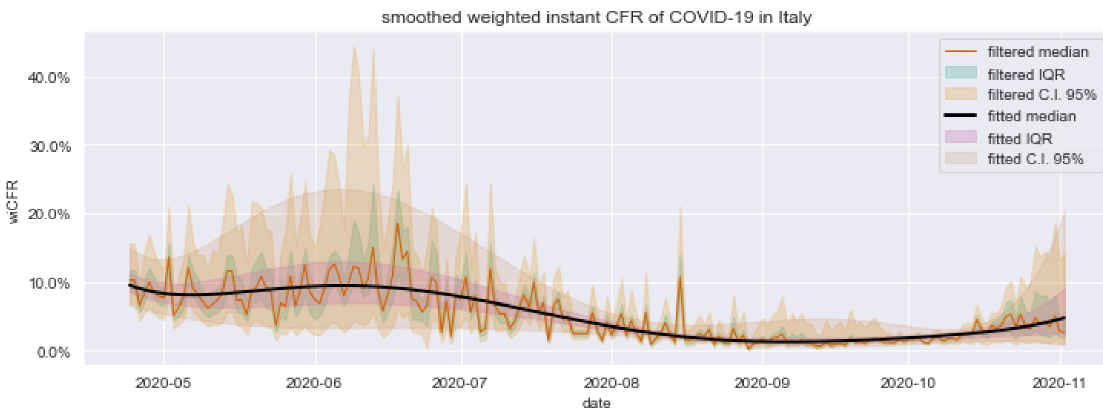


Figure 9: Smoothed filtered weighted instant CFR of COVID-19 in Italy.

comparing with classic CFR obtained by total deaths over total cases ratio (cumulative CFR) as in figure fig. 10:

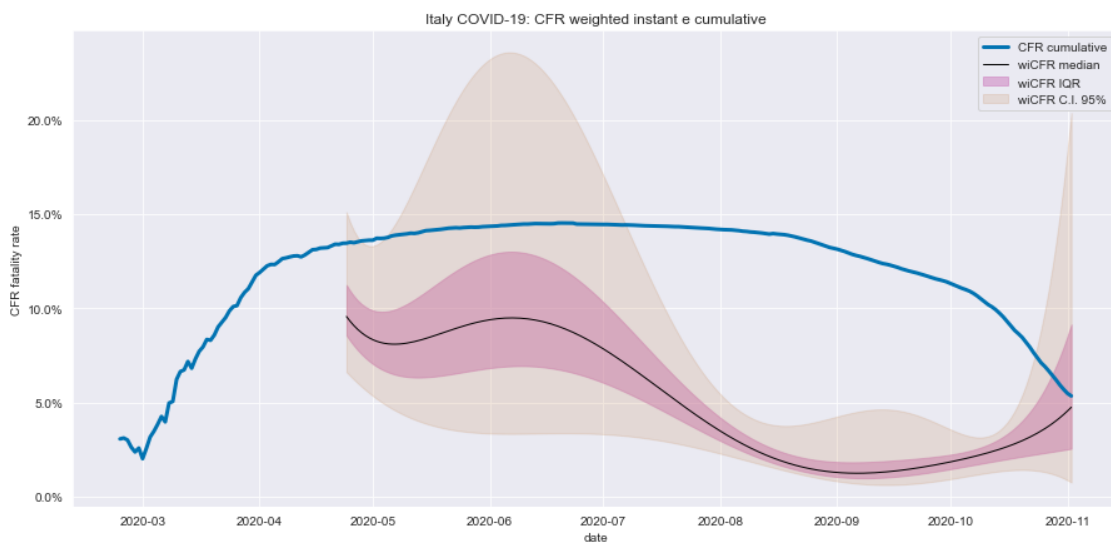


Figure 10: Weighted instant CFR and cumulative CFR of COVID-19 in Italy.

It can be noted that, despite being currently lower than the cumulative (4.74% compared to 5.34%), the cumulative CFR shows a decreasing trend (from 14.53% at 2020-06-20 to 5.34% at 2020-11-02) while the weighted instant CFR an uptrend from about mid September (from 1.23% at 2020-09-06 to 4.74% to 2020-11-02) strongly indicative of a rise in the true fatality rate during the last month.

	CFRcum %	wiCFR %
date		
2020-09-06	12.801386	1.234146
2020-09-07	12.752884	1.234657
2020-09-08	12.694135	1.237817

Continued on next page

date	CFRcum %	wiCFR %
2020-09-09	12.634641	1.243530
2020-09-10	12.566919	1.251701
2020-09-11	12.499122	1.262232
2020-09-12	12.435687	1.275027
2020-09-13	12.375197	1.289990
2020-09-14	12.336846	1.307025
2020-09-15	12.287665	1.326038
2020-09-16	12.230564	1.346939
2020-09-17	12.168928	1.369637
2020-09-18	12.093635	1.394047
2020-09-19	12.034973	1.420085
2020-09-20	11.975945	1.447673
2020-09-21	11.927641	1.476738
2020-09-22	11.877154	1.507212
2020-09-23	11.819381	1.539031
2020-09-24	11.757573	1.572141
2020-09-25	11.690695	1.606494
2020-09-26	11.625295	1.642049
2020-09-27	11.564527	1.678774
2020-09-28	11.514176	1.716648
2020-09-29	11.461259	1.755658
2020-09-30	11.399951	1.795803
2020-10-01	11.315999	1.837093
2020-10-02	11.234793	1.879550
2020-10-03	11.144195	1.923210
2020-10-04	11.061418	1.968123
2020-10-05	10.990091	2.014351
2020-10-06	10.909487	2.061975
2020-10-07	10.798646	2.111091
2020-10-08	10.662888	2.161810
2020-10-09	10.504407	2.214264
2020-10-10	10.340664	2.268603
2020-10-11	10.189041	2.324995
2020-10-12	10.068999	2.383630
2020-10-13	9.917722	2.444721
2020-10-14	9.734200	2.508501
2020-10-15	9.531397	2.575227
2020-10-16	9.301833	2.645181
2020-10-17	9.061053	2.718669
2020-10-18	8.821676	2.796027
2020-10-19	8.644453	2.877613
2020-10-20	8.448633	2.963818
2020-10-21	8.191296	3.055060

Continued on next page

date	CFRcum %	wiCFR %
2020-10-22	7.937714	3.151786
2020-10-23	7.643095	3.254479
2020-10-24	7.375488	3.363651
2020-10-25	7.101422	3.479848
2020-10-26	6.904893	3.603651
2020-10-27	6.675189	3.735679
2020-10-28	6.427125	3.876584
2020-10-29	6.182664	4.027060
2020-10-30	5.916711	4.187838
2020-10-31	5.683882	4.359691
2020-11-01	5.473577	4.543431
2020-11-02	5.338934	4.739917

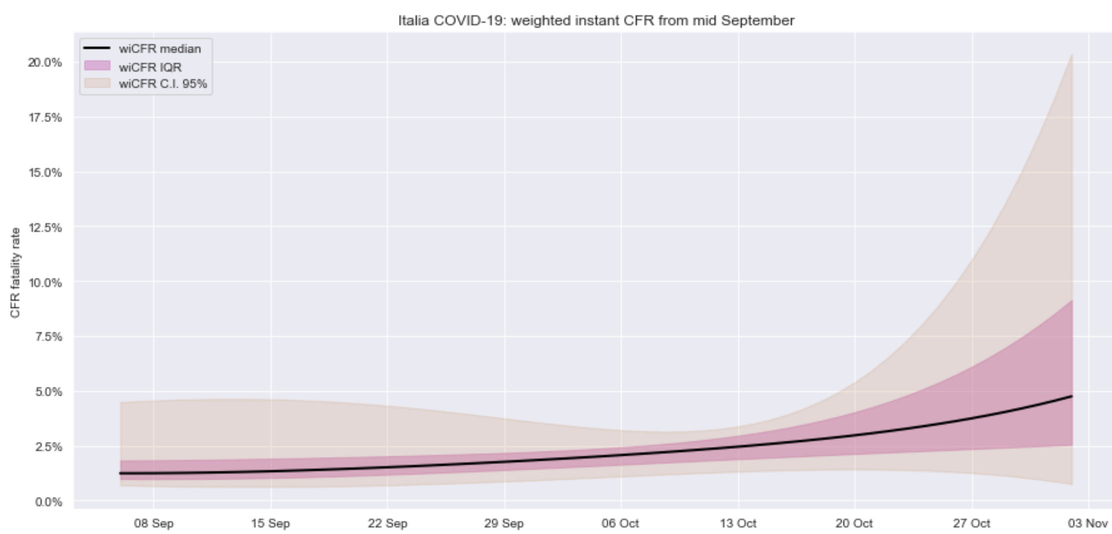


Figure 11: Last wiCFR of COVID-19 in Italy.

5 Groups of regions

- North: Piemonte
- North: Valle d'Aosta
- North: Liguria
- North: Lombardia
- North: P.A. Trento
- North: P.A. Bolzano
- North: Veneto
- North: Friuli Venezia Giulia
- North: Emilia-Romagna
- Center: Toscana
- Center: Umbria
- Center: Marche

- Center: Lazio
- South and Islands: Abruzzo
- South and Islands: Molise
- South and Islands: Campania
- South and Islands: Puglia
- South and Islands: Basilicata
- South and Islands: Calabria
- South and Islands: Sicilia
- South and Islands: Sardegna

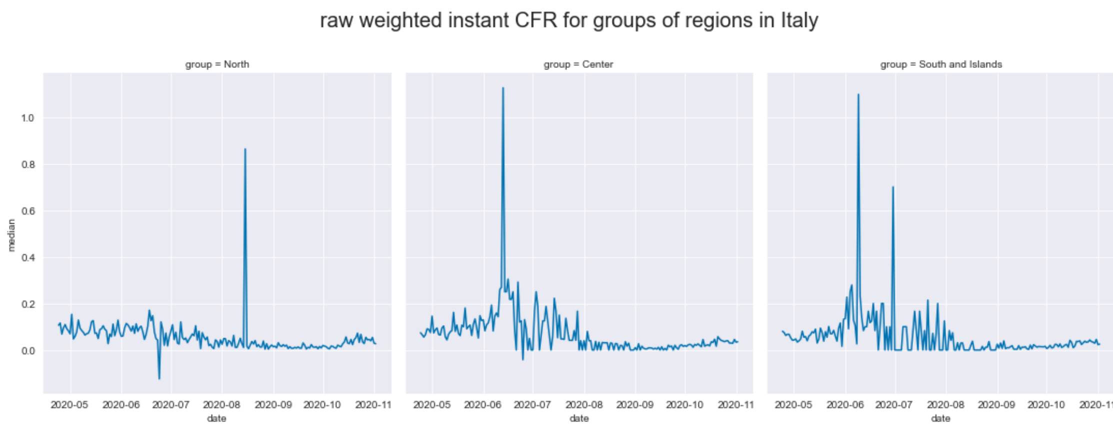


Figure 12: Raw weighted instant CFR of COVID-19 in Italian groups of regions.

- Outliers imputation North group:

```
OUTLIERS in col median (0< or >0.5)
========================================
               median        Q1        Q3        lo        hi
date
2020-06-24 -0.124138 -0.160000 -0.102857 -0.233766 -0.050350
2020-08-15  0.862637  0.723502  1.189394  0.426630  1.585859

Impute function `mean`, window 7
========================================
               median        Q1        Q3        lo        hi
date
2020-06-24  0.057600  0.032297  0.083185 -0.000577  0.139375
2020-08-15  0.143186  0.120703  0.196970  0.072852  0.262626
```

- Outliers imputation Center group:

```
OUTLIERS in col median (0< or >0.5)
========================================
               median        Q1        Q3        lo        hi
date
2020-06-13  1.125000  0.473684  1.350000  0.225000  2.076923
2020-06-25 -0.041667 -0.062500 -0.032258 -0.076923 -0.013514
```

```
Impute function `mean`, window 7
========================================
              median        Q1        Q3        lo        hi
date
2020-06-13  0.350777  0.156790  0.446270  0.072788  0.714286
2020-06-25  0.102335  0.075386  0.152138  0.020916  0.195872
```

- Outliers imputation South and Islands group:

```
OUTLIERS in col median (0< or >0.5)
========================================
              median        Q1        Q3        lo        hi
date
2020-06-09  1.096774  0.653846  1.888889  0.317757  3.400000
2020-06-30  0.700000  0.583333  1.166667 -0.031532  2.333333

Impute function `mean`, window 7
========================================
              median        Q1        Q3        lo        hi
date
2020-06-09  0.286699  0.179134  0.485701  0.085005  0.911472
2020-06-30  0.128571  0.104762  0.214286 -0.005792  0.428571
```

Filtered weighted instant case fatality rate for groups of regions (fig. 13):

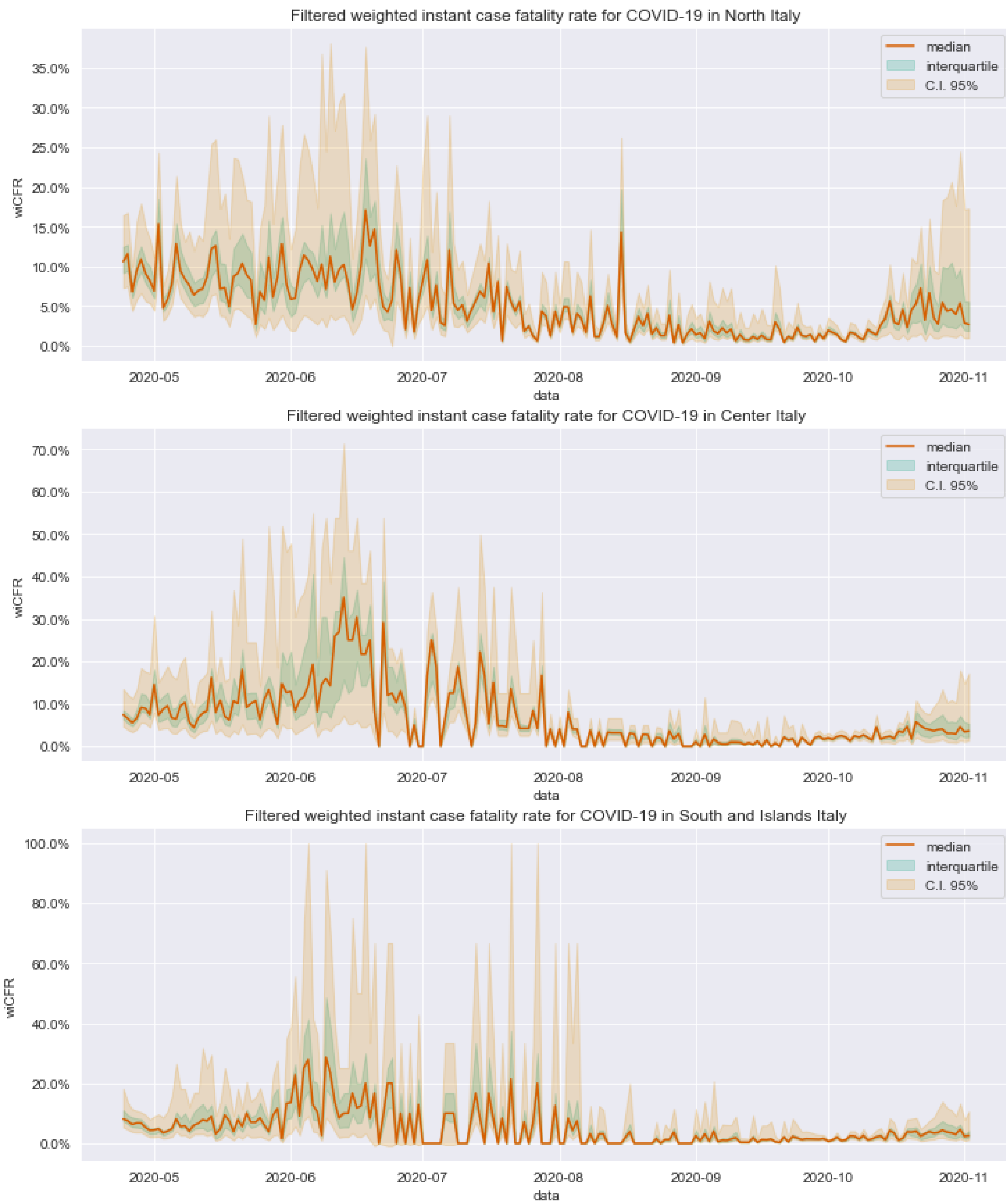


Figure 13: Filtered weighted instant CFR of COVID-19 in Italian groups of regions.

Smoothed weighted instant case fatality rate per groups of regions (fig. 14):

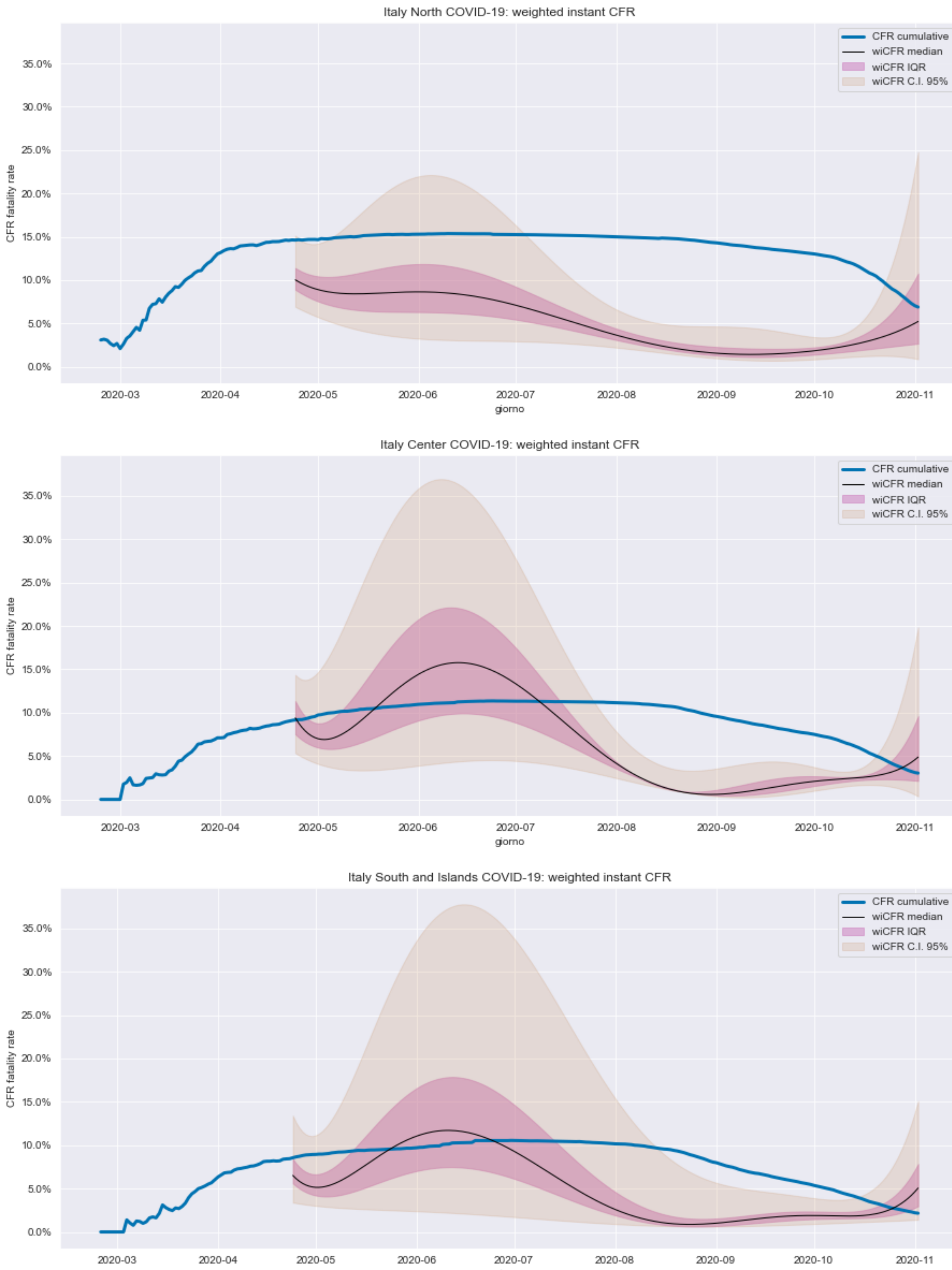


Figure 14: Smoothed filtered wiCFR and cumulative CFR of COVID-19 in Italian groups of regions.

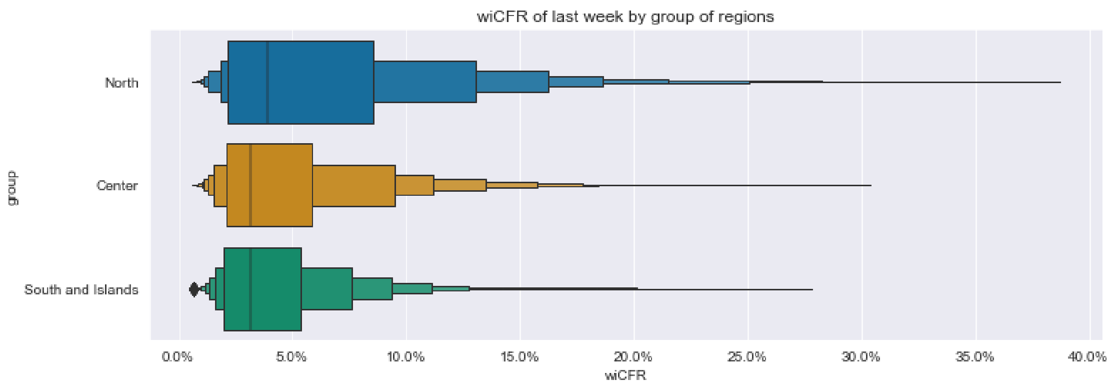


Figure 15: Last week wiCFR of COVID-19 in Italian groups of regions.

Last estimated medians of wiCFR by group of regions (fig. 15) are

	wiCFR				
	mean	std	25%	50%	75%
group					
Center	0.045884	0.035902	0.021045	0.031776	0.058520
North	0.060357	0.052859	0.021802	0.039401	0.085573
South and Islands	0.041123	0.029032	0.019754	0.031625	0.053318

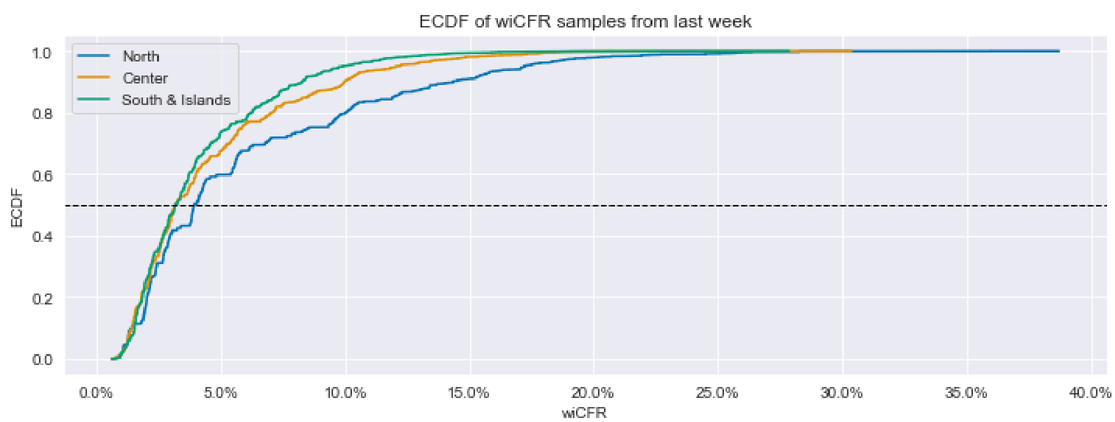


Figure 16: ECDF of last week wiCFR of COVID-19 in Italian groups of regions.

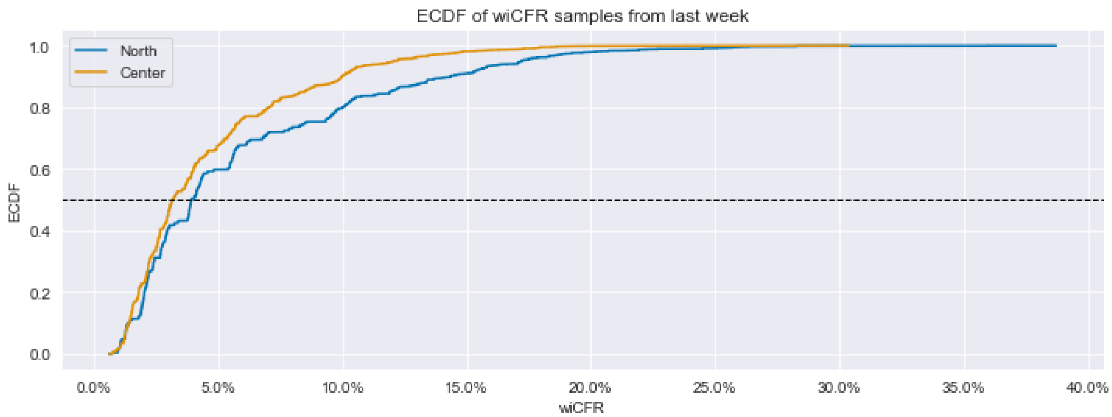
To verify if the estimated raw wiCFR (fig. 16) of last week in the three groups are significantly different from each other, we'll formulate three null hypothesis:

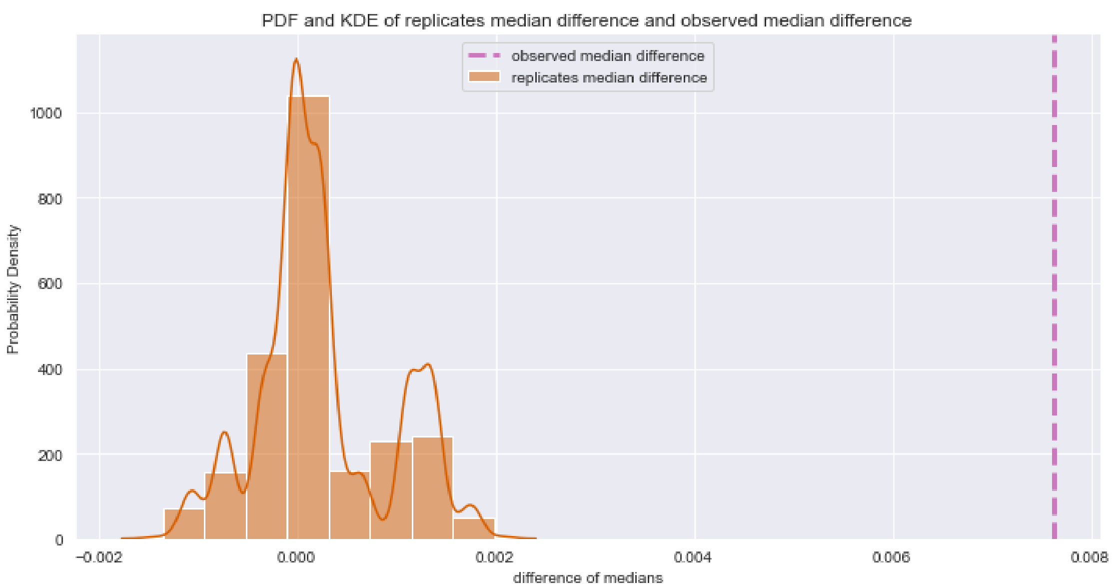
- H_0 that the last week wiCFR medians of "North" and "Center" group are equal
- H_0 that the last week wiCFR medians of "North" and "South and Islands" group are equal
- H_0 that the last week wiCFR medians of "Center" and "South and Islands" group are equal

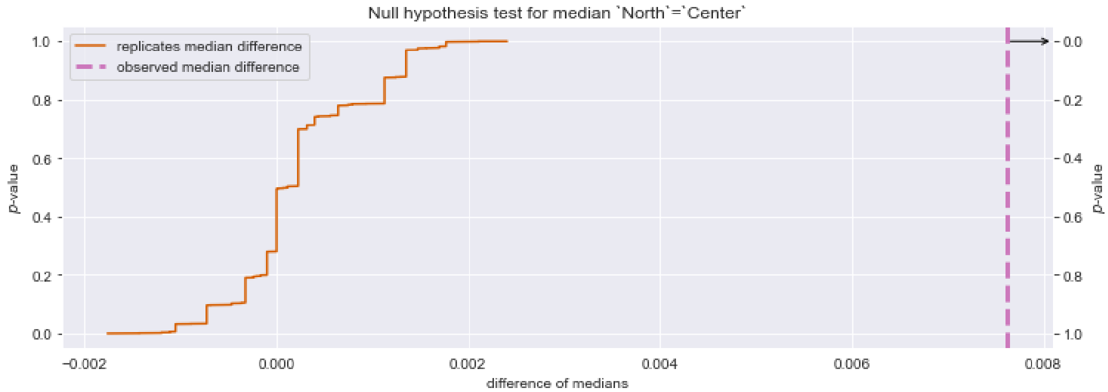
For each null hypothesis, we calculated the empirical observed difference between the medians of estimated wiCFR (subtracting smaller from larger), computed the median of pooled samples, then shifted the distributions so that they have the same median, obtained median replicates by resampling with bootstrap (size 10'000) the shifted samples and calculated the p-value to observe replicates difference of medians higher than the empirical observed difference.

If the p-value is lower than 5% ($\alpha = 0.05$) we can say that, under the null hypothesis that the samples medians are equal, the observed medians difference is statistically significant.

5.1 H0 North = Center



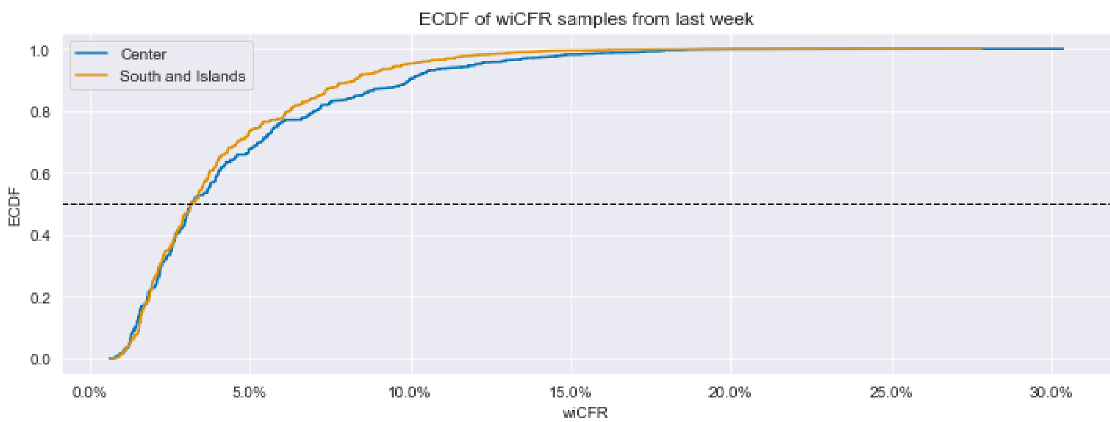


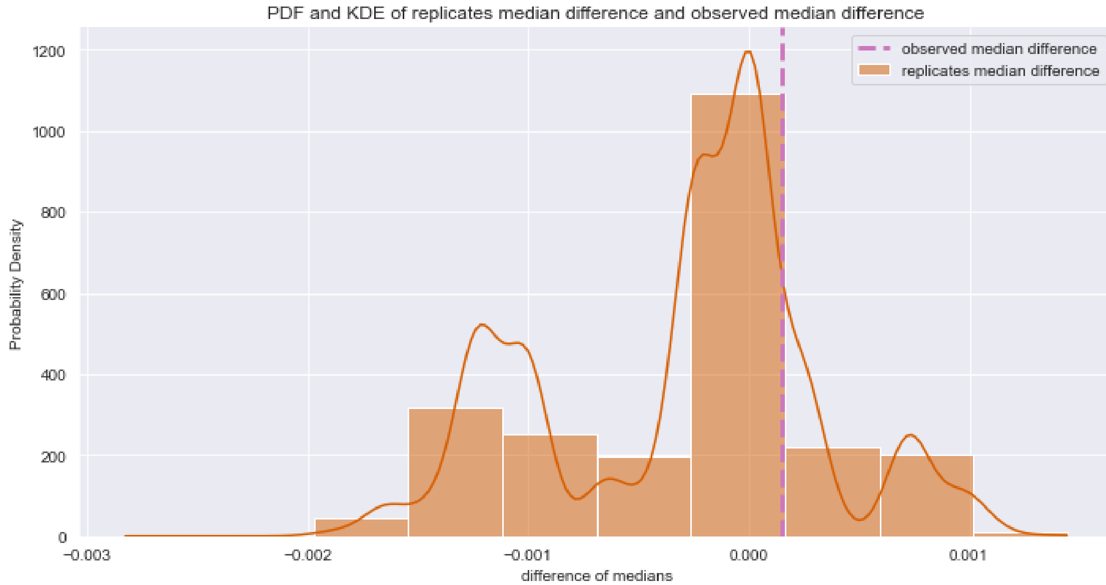


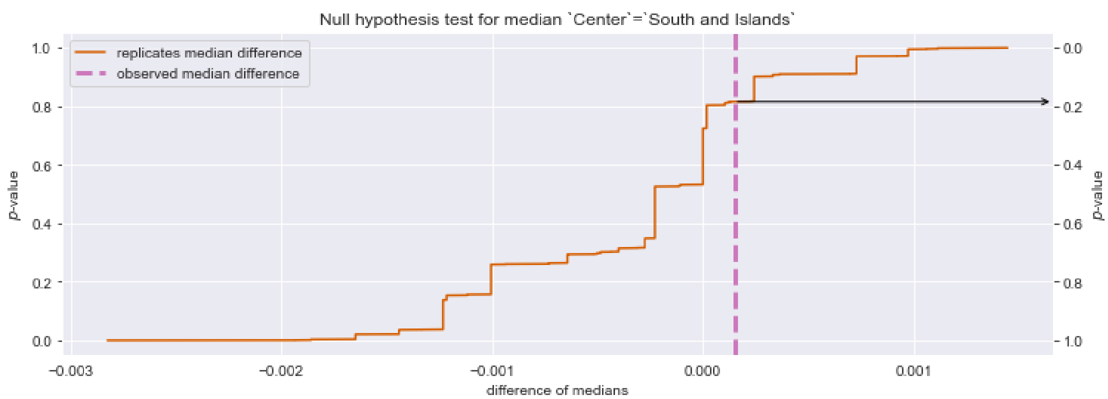
p-value for median `North`=`Center`: 0.0

Under the null hypothesis that the medians are equal, observed medians difference is statistically significant.

5.2 H0 Center = South and Islands



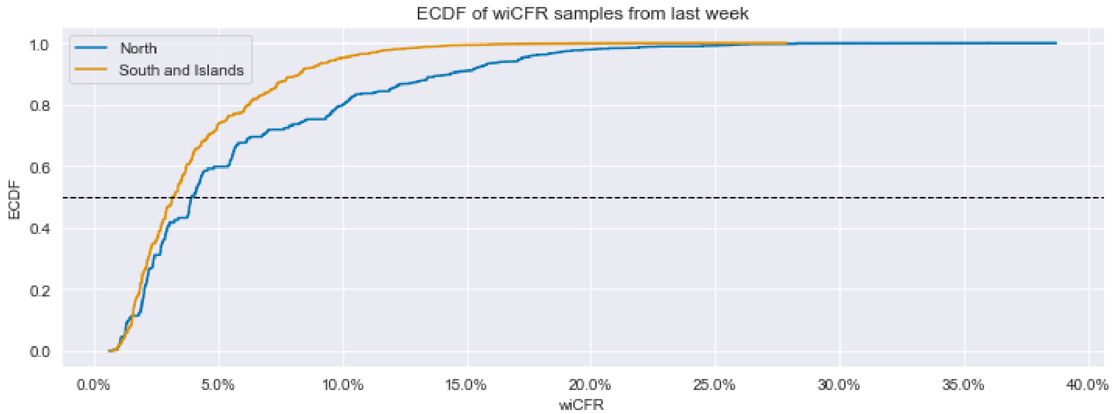


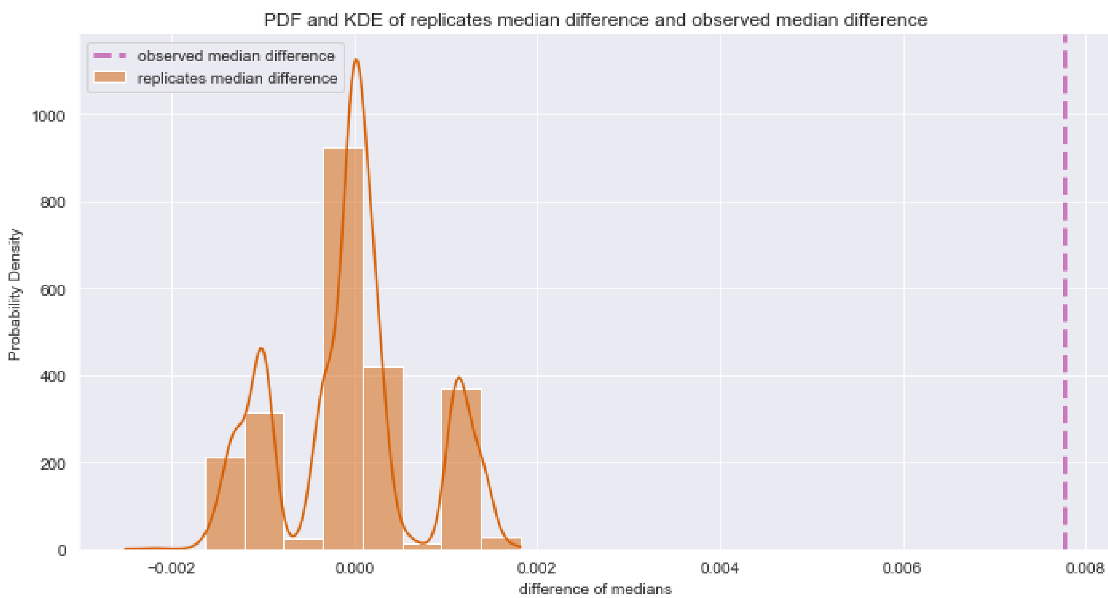


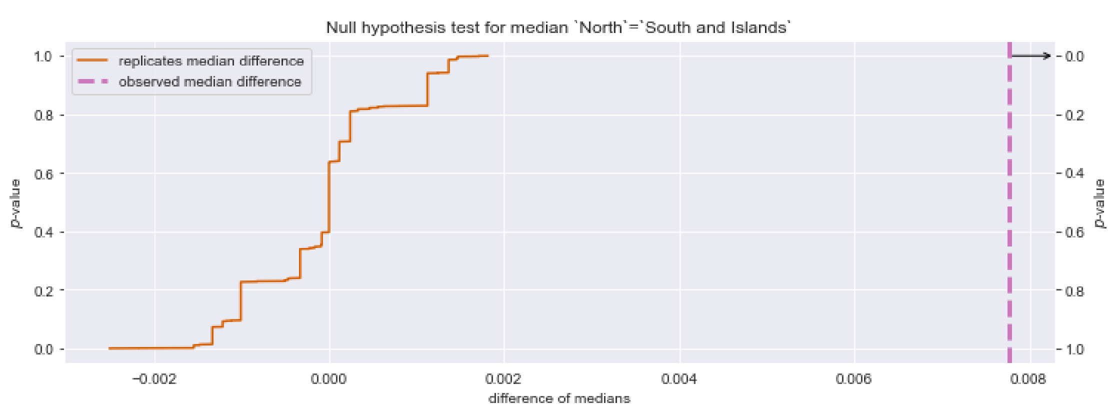
p-value for median `Center=South and Islands`: 0.1848

Under the null hypothesis that the medians are equal, observed medians difference is not statistically significant.

5.3 H0 North = South and Islands







p-value for median `North=South and Islands`: 0.0

Under the null hypothesis that the medians are equal, observed medians difference is statistically significant.

6 Discussion

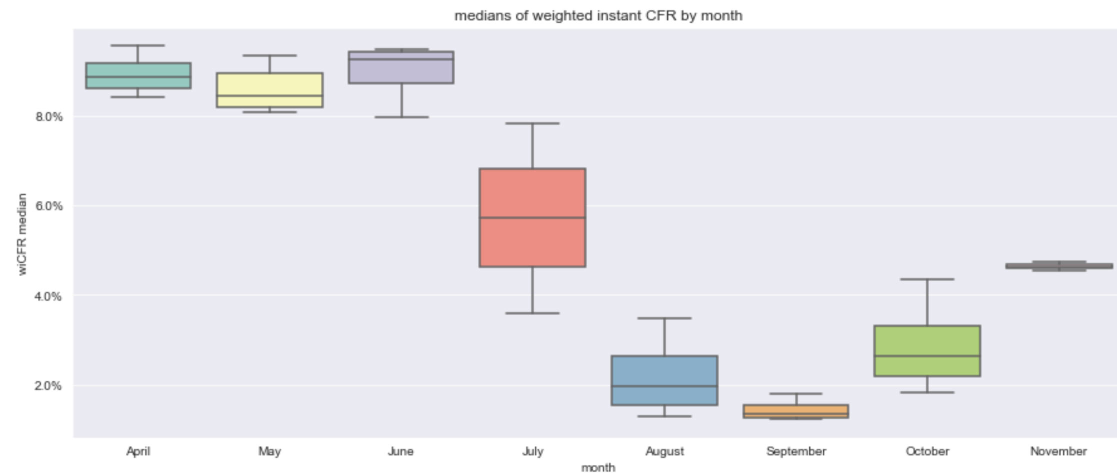


Figure 17: Weighted instant CFR of COVID-19 in Italy by month.

The estimate of weighted instant case fatality rate (wiCFR) is based on the assumption of a true estimate of the onset-to-death interval distribution. It is likely that the same onset-to-death interval could change along the timeline and that different regions can show different intervals (due to environmental conditions, median age of cases, viral strain, etc.). A more accurate estimate, based on Italian data and over several time periods, of the onset-to-death interval will therefore improve the estimate of the wiCFR.

The current estimate, with the available data and a 95% confidence interval, shows a downward trend until mid-September (fig. 17). The next trend is instead increasing, unlike the classic CFR from which trend it could instead be assumed a progressive decrease in fatality.

The weighted instant case fatality rate may be a more reliable estimate of current true fatality across the timeline than the cumulative CFR, assuming a valid distribution of the onset-to-death interval.

7 Conclusions

As the COVID-19 epidemic firstly appeared in Lombardy, the situation was very serious with the highest case fatality rate in Italy. Starting from the beginning of February, the number of patients largely increased and it was difficult for patients to receive medical care treatment. Data were lost. The admission and diagnosis of COVID-19 were delayed, the hospitals were crowded out, Intensive Care Units became insufficient. Many people were undiagnosed and died at home and these data were lost. As a result, the case fatality rate rose sharply to 20%. In March, a large number of domestic medical resources supported the North of Italy, and the medical conditions gradually improved. According to official data published by the Department of Civil Protection the cumulative case fatality rate varied over time from late February to May, arriving at a plateau, when the outbreaks became less serious and prevalent. The knowledge of the viral pathogenesis increased and treatment concept and methods were updated, the medical level was improved, and the treatment procedures became more and more standardized. Cumulative official CFR from May to middle August remained at plateau about 14,5%, but from the beginning of September started to have a decreasing trend till 7,94% at the end of October.

In Italy the epidemiological situation completely changed from the first period of the outbreak epidemic, the "first wave" and the actual situation from September up to now, the " second wave". At the beginning the tested individuals were symptomatic, tested for their symptoms according with the directives of the government scientific committee (CTS Comitato Tecnico Scientifico) and WHO; the high number of deaths and infected heavily symptomatic persons with the collapse of the health Structures, determined the government decision to try to interrupt the chain of infectivity with a total lock down, exactly as in Wuhan, China. After the end of this restriction period and with the availability of a number of swab tests by RT-PCR, to make evidence of viral RNA, and a wider panel of additional antigenic and rapid tests, including serology, a more wide testing was applied to the population, an intensive contact tracing was performed and the positive population switched from a symptomatic to an asymptomatic subjects population. From the half of September a new increase of the infected positives was detected, a potential "second wave": the CFR data by the Department of Civil Protection, cumulative CFR, showed a decrease in the later period from September up to now (fig. 10, fig. 11 fig. 17) in contrast with the other epidemiological indexes showing a marked increase of cases and incidence (Percent Positive and R_t).

We prefer to think that the whole epidemic in Italy could be considered as a fluent dynamic process of the same epidemic and, thus, assuming that case fatality rate of the whole outbreak process could be regarded as a collection of many sequential instant case fatality rates in a dynamic process, which could express the dynamic of the epidemic and the dynamic severity risk.

The weighted instant case fatality rate is directly related to various factors during time; it becames easier to analyze various factors and take possible actions to influence the disease progress by knowing this dynamic fatality rate in advance. Instant case fatality rate proposed up to now will gradually approach the case fatality rate as time goes on till the outbreak develops and tends to end, gradually. Our new approach provides a way to more accurately calculate the fatality rate, which consider the new deaths to new cases, taking into account the onset-to-death interval, the time from onset of symptoms to eventual death. Our weighted instant Case-Fatality-Rate is calculated from only deaths to new cases, assuming a valid distribution of the onset-to-death interval.

This rate varies from day to day with disease-related factors, viral and individual, from evidence of symptoms till death.

If we would have followed the cumulative CFR we would have skipped the new increase of case of infection and the impact to severity of the infection itself that well are demonstrated by our weighted instant CFR (fig. 10, fig. 11 fig. 17).

And, consequentially, immediate new health care, political and social interventions can be set up to decrease the severity and lethality of the infection providing the basis for the prevention and control of the worst events of SARS-CoV-2/COVID-19 disease.

References

[1] W. H. Organization *et al.*, "Coronavirus disease 2019 (covid-19): situation report, 51," 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf

[2] R. A. Harrington, "Case fatality rate," *Encyclopedia Britannica. URL: https://www. britannica. com/science/case-fatality-rate. Access Date: April*, vol. 4, 2020.

[3] Y.-n. Mi, T.-t. Huang, J.-x. Zhang, Q. Qin, Y.-x. Gong, S.-y. Liu, H.-m. Xue, C.-h. Ning, L. Cao, and Y.-x. Cao, "Estimating the instant case fatality rate of covid-19 in china," *International Journal of Infectious Diseases*, vol. 97, pp. 1–6, 2020.

[4] D. D. Rajgor, M. H. Lee, S. Archuleta, N. Bagdasarian, and S. C. Quek, "The many estimates of the covid-19 case fatality rate," *The Lancet Infectious Diseases*, vol. 20, no. 7, pp. 776–777, 2020.

[5] G. Onder, G. Rezza, and S. Brusaferro, "Case-fatality rate and characteristics of patients dying in relation to covid-19 in italy," *Jama*, vol. 323, no. 18, pp. 1775–1776, 2020.

[6] F. Öztoprak and A. Javed, "Case fatality rate estimation of covid-19 for european countries: Turkey's current scenario amidst a global pandemic; comparison of outbreaks with european countries," *world*, vol. 21, p. 24, 2020.

[7] Presidenza del Consiglio dei Ministri and Dipartimento di Protezione Civile, "Dati covid-19 italia," *GitHub repository*, 2020. [Online]. Available: https://github.com/pcm-dpc/COVID-19

[8] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu *et al.*, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet infectious diseases*, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1473309920302437

[9] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[10] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, p. 357–362, 2020.