

CovidStat: monitorare l'andamento di una pandemia con l'analisi dei dati.

G. Bonifazi^{1,2}, L. Lista^{3,4,8}, D. Menasce⁵, M. Mezzetto⁶, D. Pedrini⁵, R. Spighi², A. Zoccoli^{7,2}

¹ Università Politecnica delle Marche
² INFN Sezione di Bologna

³ Università degli Studi di Napoli Federico II
⁴ INFN Sezione di Napoli

⁵ INFN Sezione di Milano Bicocca
⁶ INFN Sezione di Padova

⁷ Alma Mater Studiorum Università di Bologna
⁸ Scuola Superiore Meridionale

SEQUENZA TEMPORALE DEGLI EVENTI

17
Novembre
2019

Viene **registrato il primo caso** di contagio accertato da COVID-19: si tratta di un cinquantacinquenne della **provincia dello Hubei**.

31
Dicembre
2019

L'**Organizzazione Mondiale della Sanità (OMS)** viene informata dalle autorità cinesi di una serie di casi simili alla polmonite nella città di **Wuhan**.

13
Gennaio
2020

In **Thailandia** viene registrato il **primo caso di positività fuori dalla Cina** noto in quel momento.

30
Gennaio
2020

Vengono accertati per la **prima volta in Italia due casi di persone infette**, si tratta di due turisti cinesi in vacanza a Roma

31
Marzo
2020

Nasce il sito **CovidStat** dell'INFN.

10
Marzo
2020

Nasce il **gruppo CovidStat** dell'INFN.

9
Marzo
2020

L'Italia impone il **lockdown a livello nazionale**, limitando i viaggi tranne che per necessità, lavoro e circostanze sanitarie.

21
Febbraio
2020

Vengono confermati **nuovi casi in Lombardia e in Veneto**. Vanno in lockdown 11 comuni (circa 50.000 abitanti)



IL GRUPPO COVIDSTAT



Nato ai primi di Marzo 2020 per aiutare l'Unità di Crisi INFN ad **interpretare** l'andamento della **pandemia**.



Lo **scopo** di questo gruppo è quello di **analizzare**, utilizzando le competenze di **statistica** proprie delle ricerche in fisica, i **dati** riguardanti la pandemia da COVID-19.

E' stato inoltre realizzato un **sito web** (covid19.infn.it) per fornire un **servizio utile** ad altre **comunità scientifiche** e al **pubblico** in generale.



IL SITO COVIDSTAT

- Aggiornamento **quotidiano** con i dati della **Protezione Civile**
- Aggiornamento **settimanale** con i dati dell'**Istituto Superiore di Sanità**
- **Dati scaricabili**
- **Contenuti interattivi**



ALCUNI NUMERI

3.000+

Grafici

20,000+

Righe di codice

5.000+

Media utenti giornalieri

150.000+

Visite al sito web mensili

154

Paesi di origine degli utenti

2.900.000+

Visite al sito web totali

PUBBLICAZIONI

A

“A simplified estimate of the Effective Reproduction Number R_t using its relation with the doubling time and application to Italian COVID-19 data”.

The European Physical Journal Plus, 136(4):386, 2021. Springer.

G. Bonifazi, L. Lista *, D. Menasce, M. Mezzetto, D Pedrini, R. Spighi, A. Zoccoli.

B

“A statistical analysis of death rates in Italy for the years 2015-2020 and a comparison with the casualties reported for the COVID-19 pandemic”.

Infectious Disease Reports, 13(2):285–301, 2021. MDPI.

G. Bonifazi, L. Lista, D. Menasce *, M. Mezzetto, A. Oliva, D Pedrini, R. Spighi, A. Zoccoli.

C

“A study on the possible merits of using symptomatic cases to trace the development of the COVID-19 pandemic”.

The European Physical Journal Plus, 136(5):481, 2021. Springer.

G. Bonifazi, L. Lista, D. Menasce, M. Mezzetto *, D Pedrini, R. Spighi, A. Zoccoli.

D

“Study on the effects of the restrictive measures for containment of the COVID-19 pandemic on the reproduction number R_t in Italian regions”.

Sottomesso a The European Physical Journal Plus [arXiv: 2106.02603]

G. Bonifazi, L. Lista, D. Menasce, M. Mezzetto *, D Pedrini, R. Spighi, A. Zoccoli.



A

**“A simplified estimate of the
Effective Reproduction Number R_t
using its relation with the doubling
time and application to Italian
COVID-19 data.”**

The European Physical Journal Plus

<https://doi.org/10.1140/epjp/s13360-021-01339-6>



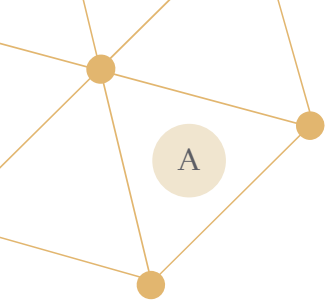
A

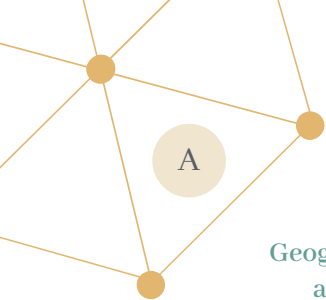
Il numero effettivo di riproduzione del virus, meglio conosciuto come R_t , è uno dei principali parametri che controlla l'evoluzione di un'epidemia.

Durante la pandemia da COVID-19 ha acquisito un ruolo centrale perché è stato usato come indicatore per determinare misure restrittive.

In letteratura esistono diversi algoritmi in grado di calcolare il valore di R_t a partire dai dati sui contagi ma alcuni di questi sono molto onerosi dal punto di vista computazionale.

Abbiamo proposto un approccio semplificato, comunque in ottimo accordo con gli altri metodi per la stima di R_t dimostrando la relazione di questo valore con il tempo di raddoppio dei casi.





Geographic area	Abitanti (mln)	CovidStat	Wallinga Teunis	Bettencourt Ribeiro	Cori et al.	RKI
Italia	60.36	<u>0.138</u>	510.1	80.3	11.5	0.587
Lombardia	10.06	<u>0.120</u>	80.8	78.3	11.5	0.585
Lazio	5.88	<u>0.105</u>	30.0	76.1	11.2	0.586
Campania	5.80	<u>0.103</u>	31.9	33.9	11.2	0.584
Emilia-Romagna	4.46	<u>0.101</u>	33.5	75.2	11.4	0.585
Basilicata	0.56	<u>0.098</u>	20.1	28.9	11.3	0.584



Tempo di utilizzo della CPU
in secondi



B

**“A statistical analysis of death rates
in Italy for the years 2015–2020 and
a comparison with the casualties
reported from the COVID-19
pandemic.”**

Infectious Disease Reports

<https://doi.org/10.3390/idr13020030>



B

Abbiamo analizzato i **dati sui decessi** avvenuti **in Italia** nel periodo 01/01/2015 - 30/09/2020 rilasciati dall'Istituto Nazionale di Statistica (**ISTAT**).

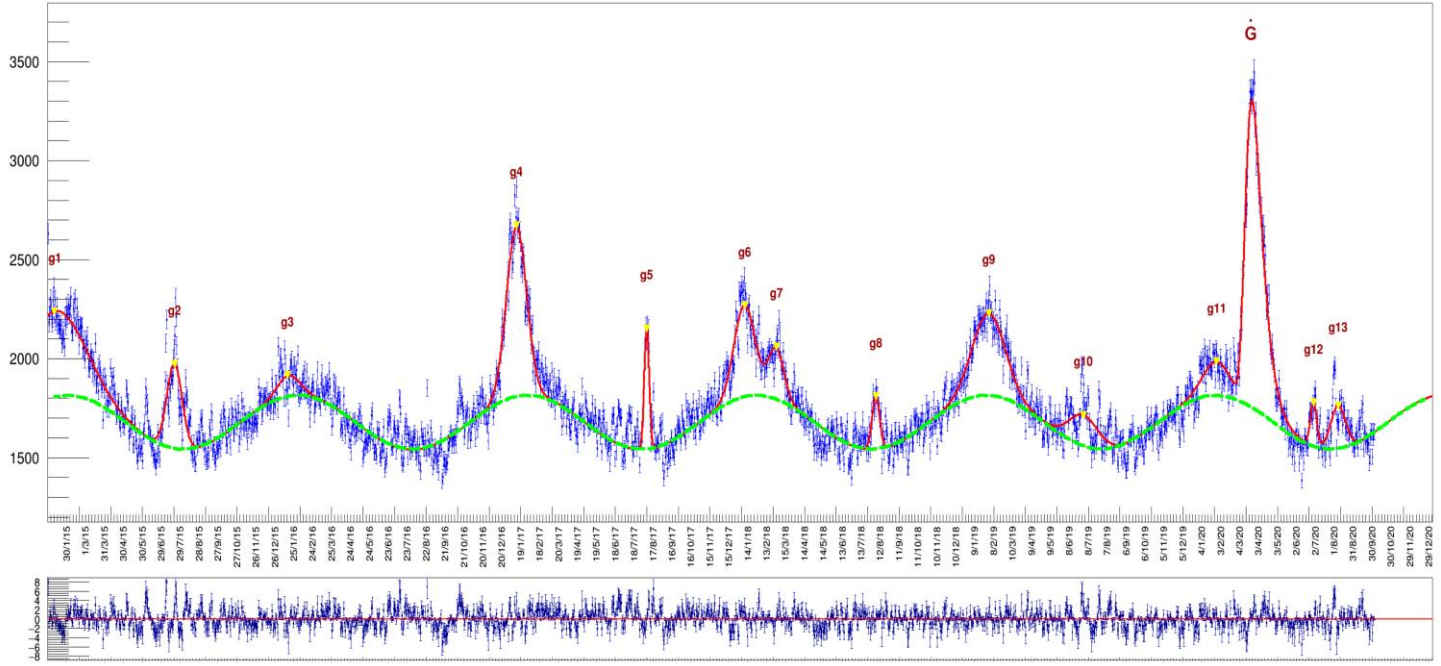
Lo scopo di questa analisi è quello di **estrarre valori quantitativi** per gli **eccessi stagionali dei decessi**.

I dati mostrano un **chiaro comportamento sinusoidale** di base, il cui fit permette una robusta sottrazione rendendo possibile il calcolo del **surplus di mortalità** in corrispondenza delle epidemie influenzali in inverno e dei periodi più caldi in estate.

Infine, abbiamo **confrontato i risultati** ottenuti con i **decessi ufficiali COVID-19** pubblicati dalla **Protezione Civile** evidenziando differenze nei due campioni.

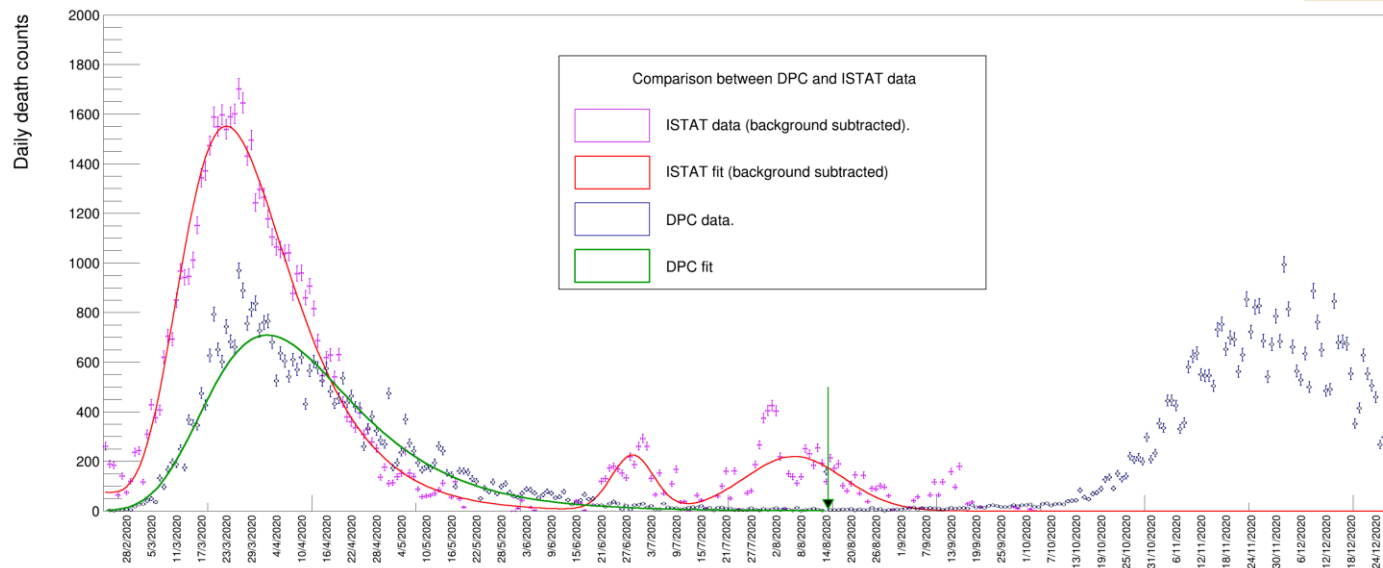
B

Number of daily deaths



B

Periodo: 24/02/2020 - 15/08/2020 (1° ondata)
Decessi COVID-19 ufficiali (da dati Protezione Civile): **35.468**
Decessi totali (integrale del fit sui dati ISTAT): **54,387 (± 557)**



Confrontando i dati ISTAT e DPC si ha una differenza di **18.919 (± 557)** decessi.

Questa differenza così marcata potrebbe essere dovuta a diverse ragioni ma non abbiamo elementi che possano permetterci di discernere i diversi contributi a questa discrepanza.

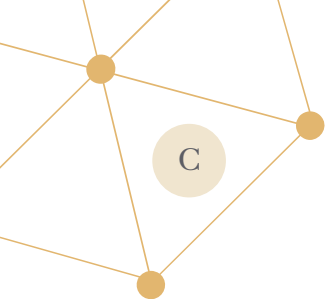


C

“A study on the possible merits of using symptomatic cases to trace the development of the COVID-19 pandemic.”

The European Physical Journal Plus

<https://doi.org/10.1140/epjp/s13360-021-01448-2>



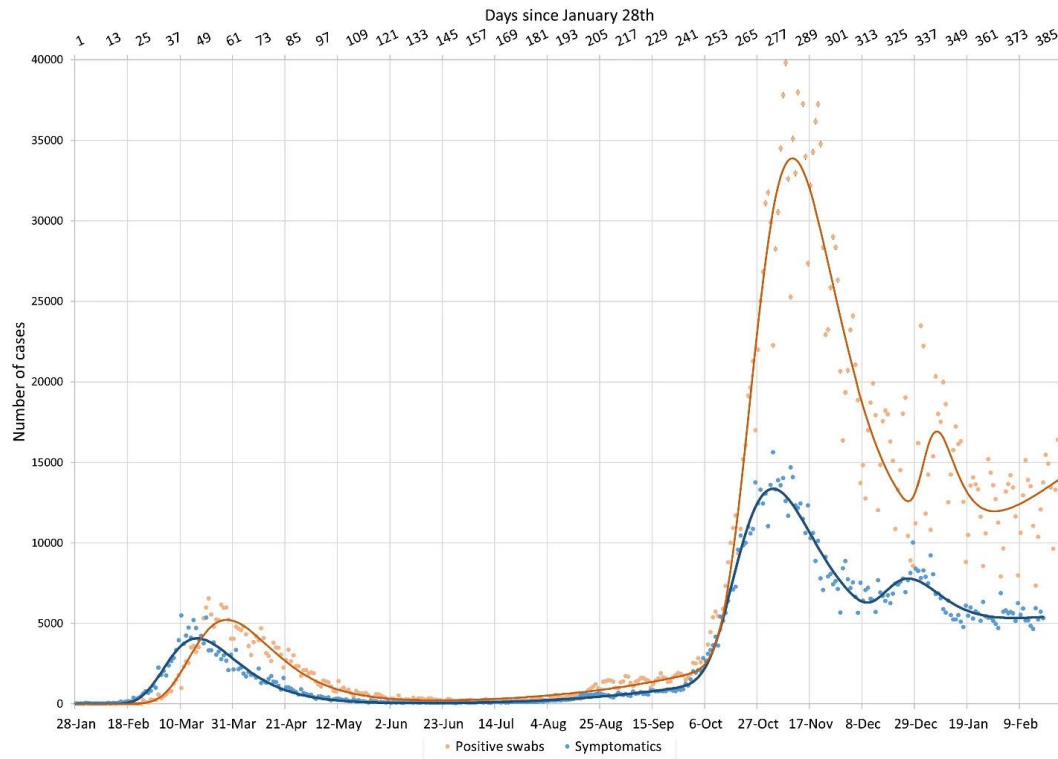
La stima ufficiale di R_t da parte dell'Istituto Superiore di Sanità viene calcolata a partire dalla data di comparsa sintomi dei soli casi sintomatici.

Quindi, utilizzando informazioni sui casi sintomatici si dovrebbe avere un quadro più preciso della situazione ma la raccolta dati comporta comunque un ritardo.

Questa scelta deriva dal fatto che, i dati relativi ai «casi positivi» pubblicati quotidianamente dalla Protezione Civile soffrono di un fisiologico ritardo tra la data del contagio e la data della segnalazione.

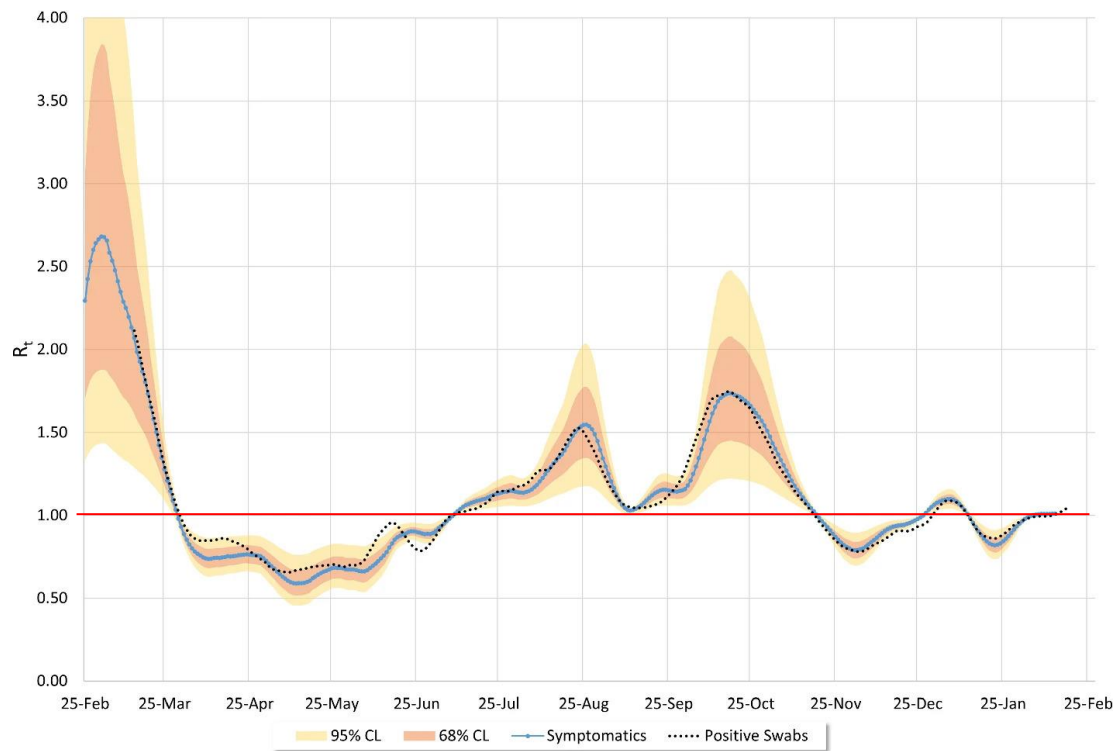
D'altra parte, il campione di casi sintomatici è un sottoinsieme del numero totale di casi e ciò diventa un problema per popolazioni statistiche relativamente piccole.

Abbiamo quindi elaborato diversi indicatori per confrontare i vantaggi e gli svantaggi per monitorare l'andamento di un'epidemia con questi due set di dati.



Confrontando le date del secondo picco abbiamo calcolato che il dataset «casi positivi» è in ritardo di **7,9±0,2 giorni** rispetto a quello dei «casi sintomatici».

Questo valore può quindi essere considerato il **ritardo medio** tra la **comparsa dei sintomi** e la **segnalazione di un tampone positivo**.



Utilizzando l'algoritmo CovidStat abbiamo calcolato R_t per entrambi i dataset spostando di **8 giorni** indietro il valore di R_t per il dataset «casi positivi».

I dati raccolti per data di comparsa sintomi **impiegano 14 giorni per consolidarsi**.
 Quindi utilizzare questa **correzione temporale** garantirebbe una stima di R_t **6 giorni in anticipo**.



D

“Study on the effects of the restrictive measures for containment of the COVID-19 pandemic on the reproduction number R_t in Italian regions.”

In revisione su EPJ+

<https://arxiv.org/abs/2106.02603>



D

Dal 6 Novembre 2020, le regioni e le province autonome italiane sono state classificate secondo quattro differenti scenari di rischio: zona bianca, zona gialla, zona arancione e zona rossa.

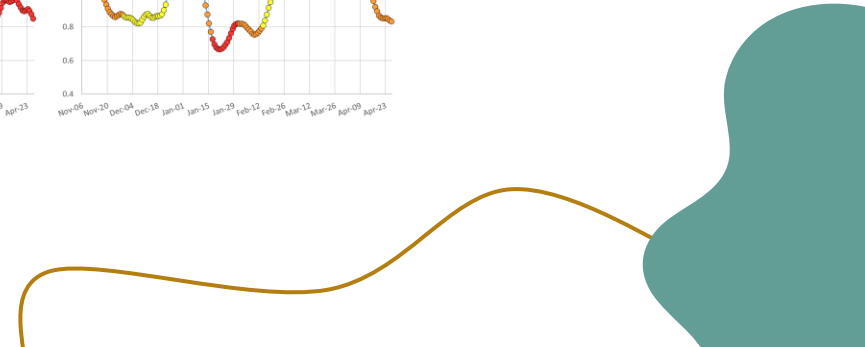
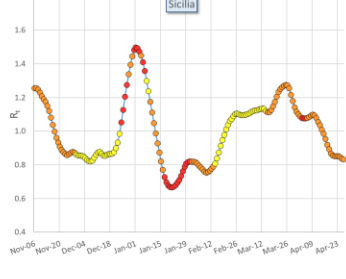
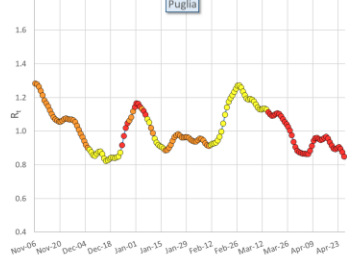
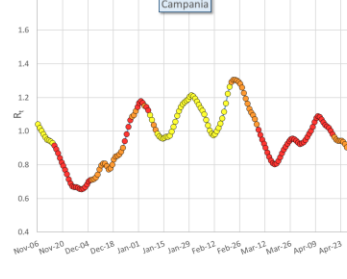
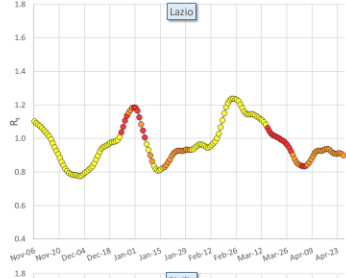
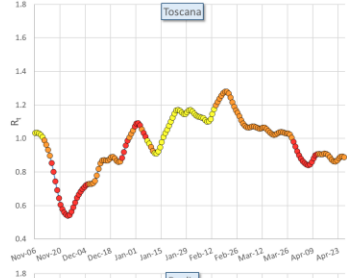
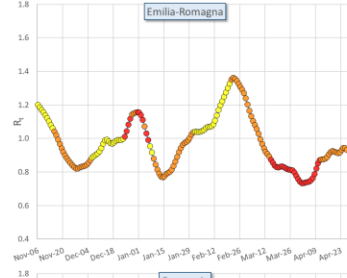
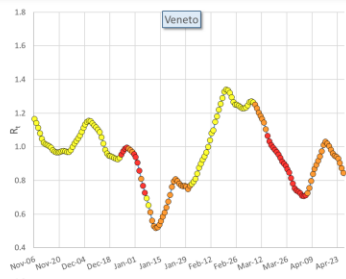
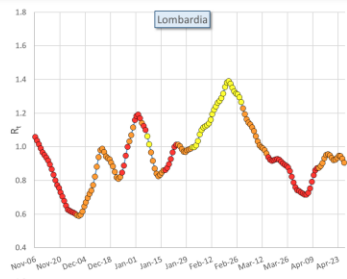
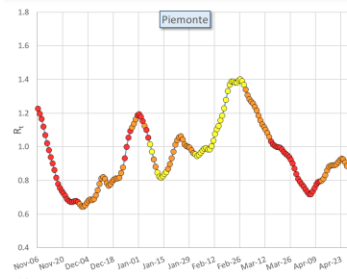
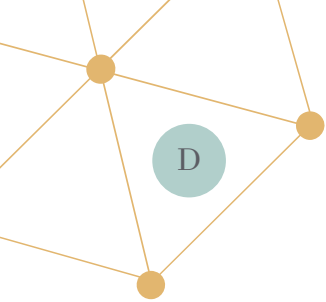
Analizzando l'evoluzione temporale dell'effettivo numero di riproduzione del virus, R_t , abbiamo stimato quanto queste misure restrittive abbiano influenzato la diffusione del virus.

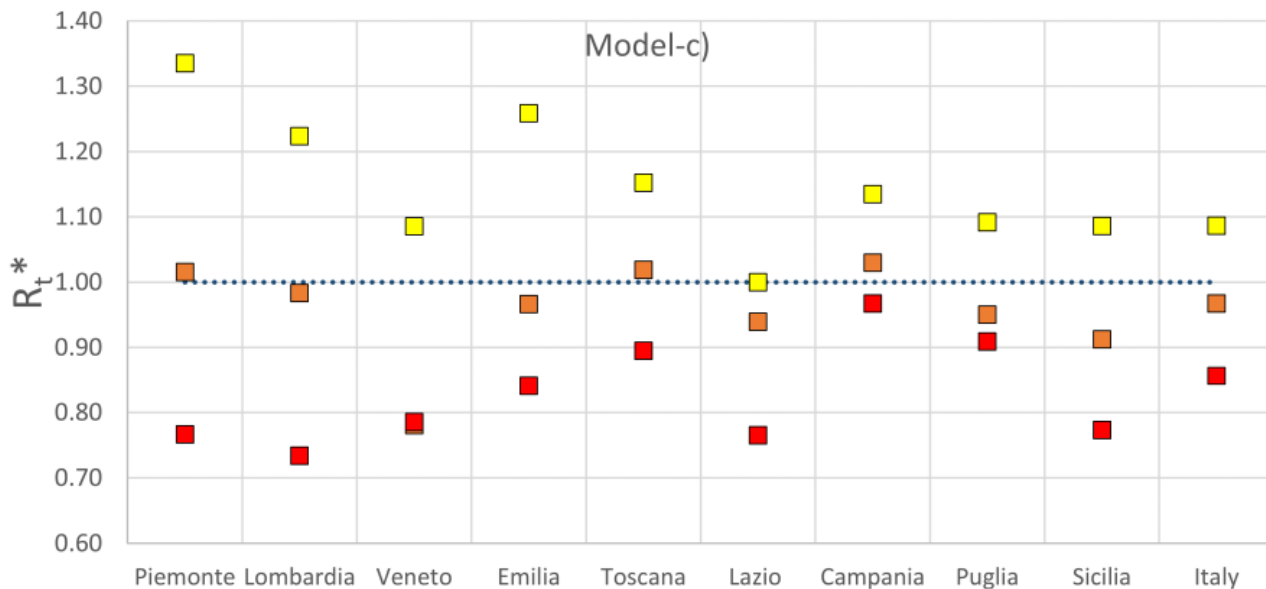
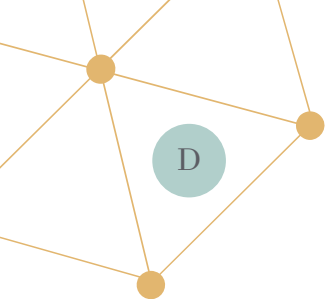


Ognuna di queste «zone» prevedeva specifiche misure restrittive, via via crescenti.



Inoltre abbiamo anche calcolato il ritardo temporale tra l'attuazione delle misure restrittive e gli effetti che esse portavano.





I risultati indicano che le restrizioni associate alla zona di rischio «gialla» sono **inadeguate** a garantire il contenimento della pandemia, poiché il valore di R_t è sempre superiore ad 1.

Il livello di rischio «arancione» tende ad avvicinarsi quasi sempre al **valore** di $R_t=1$ mentre la zona «rossa» sembra **garantire un buon contenimento**.



107° CONGRESSO NAZIONALE
SOCIETÀ ITALIANA DI FISICA

THANKS!

Do you have any questions?

gianluca.bonifazi@bo.infn.it | g.bonifazi@univpm.it

<https://biuni.it>



Istituto Nazionale di Fisica Nucleare

« I have no special talents.
I am only passionately
curious. »

– Albert Einstein